

MEDICAL GEOGRAPHY

THIRD EDITION



Melinda S. Meade
Michael Emch

MEDICAL GEOGRAPHY

THIRD EDITION

Melinda S. Meade
Michael Emch



THE GUILFORD PRESS

New York London

MEDICAL GEOGRAPHY

©2010 The Guilford Press
A Division of Guilford Publications, Inc.
72 Spring Street, New York, NY 10012
www.guilford.com

All rights reserved

No part of this book may be reproduced, translated, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the Publisher.

Printed in the United States of America

This book is printed on acid-free paper.

Last digit is print number: 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Meade, Melinda S.

Medical geography / Melinda S. Meade, Michael Emch.—3rd ed.
p. cm.

Includes bibliographical references and index.

ISBN 978-1-60623-016-9 (hardcover: alk. paper)

1. Medical geography. I. Emch, Michael. II. Title.

RA792.M42 2010

614.4'2—dc22

2009046573

Preface

Much has changed in the world and in the discipline of geography since the second edition of this text was published in 2000. For the most part, progressing globalization has extended and intensified the processes of ecological simplification, demographic transition, and cultural/economic/technological connectivity that result in world patterns of disease. These and other processes and their impacts are integrated in place, in region—in *geography*, which spatially analyzes and socially explains patterns in cause of death and etiology of health or disease. This text is about both geography as integration in place, and geography as spatial analysis.

Specific post-2000 changes worthy of attention here include the following:

- World population has grown to about 6.7 billion; its fertility rate has dropped to 2.6 children; and its growth rate has decreased to only 1.2% a year. The richest European countries (and Japan) have raised life expectancy to about 80 years, but their birth rates have dropped so low that they have negative population growth today. The youthful rural population that resulted from the high fertility rates of 20 years ago has picked up and headed for the cities, sometimes internationally.

- The poorest countries still have the highest mortality: More than 10% of infants in these countries never reach their first birthday, and life expectancy remains less than 50 years. A million of the poorest still die every year from tuberculosis, and another million from malaria.

- Whereas AIDS had just emerged before 2000, it now causes 2 million deaths annually. About 30 million people are HIV-positive, with the highest prevalence rates of about 25% in southern Africa. AIDS has become a modern global plague of enormous demographic and social impact; however, we are little closer to a cure for it than we were two decades ago. Although there has been great progress in developing treatments for this disease—enough that some people in the richest countries

are starting to consider AIDS a chronic disease—there are huge disparities in who gets the drugs.

- China has developed faster than any other society in history. It has completed one massive project of environmental engineering (the Three Gorges Dam) and has plans for others, all concerned with redistributing the flow of water from south to arid north. There has been devastating environmental degradation associated with this alteration, as well as industrial growth and insatiable coal burning.
- Other new diseases have emerged; old diseases have been redistributed; and the basis for horrific epidemics of cancer and chronic diseases has been laid in a world population no longer as young as it once was.
- What may be the first impacts of climate warming on human health have been expressed in increasing drought conditions in arid lands and increasing flooding and cyclone severity, which have already destroyed crops and villages and have displaced tens of millions in sub-Saharan Africa and South Asia.

Health promotion has become the world theme. Biomedicine has been forced off its pedestal. A genetic component has been discovered for many diseases; however, the environmental context for the expressions of those genes is more than ever something that must be explored in order to understand the diseases' complex etiologies. Social science has developed an interdisciplinary theoretical perspective and a new vocabulary for addressing old processes. Geographic information systems (GIS), with their ability to manage and portray spatial data, have become the dominant tools in geography and have transformed the structuring of public data and health analyses. Medical geography as a subdiscipline has become less concerned with the optimization of health service delivery or a dichotomy between health service and disease ecology (etiology). Instead, it has become increasingly concerned with health geography as a behavioral and social construction, and with disease ecology as an interface between the natural (physical world) and cultural dimensions of existence.

This text still endeavors to provide a broad-based, comprehensive survey of the rich diversity of medical geography for upper-division undergraduates and graduate students, while also serving as a sound reference for the complexities of classifications, processes, and systems. Our perspective remains holistic and international. We still hope to provide the necessary biological background for geographers to understand disease processes, as well as the necessary geographic background for health researchers to understand spatial processes. Students who have used the text in the past decade have included medical doctors pursuing doctorates in epidemiology; graduate students in geography working on their doctorates; graduate students from such public health disciplines as epidemiology, biostatistics, health behavior, nutrition, health administration, and public policy; undergraduate geography majors; and premedical undergraduates with majors in chemistry or biology, but little background in the social sciences or geography. It is difficult to meet the needs of such varied students, with their different competencies in statistics, mapping, and geography. The vignettes (fewer than in past editions) offer focused discussions, usually of techniques or other topics that not all teachers will want to go

into (in chapters). Some may want to teach the vignettes with chapters other than the ones in which they are presented. We have added a chapter on GIS in health and disease research, and have expanded the chapter on spatial analysis. We have also added a chapter on neighborhoods and health, which is an area that has grown exponentially during the past decade.

Something also needs to be said about references. We have attempted to write a general textbook that can be read without the constant interruption of strings of citations, most of which are publications themselves citing the common source of an idea/term or general knowledge in a field (“It is hot in the tropics [x, 1958]” or “The malaria plasmodia (schistosome, tick, etc.) has several life stages [y, 1999]”). Where possible, we have referenced the seminal ideas and influential discussions, not all of which represent the most recent applications of those ideas. Our references are not meant to be comprehensive bibliographies of works in medical geography. The approaches, information, or examples we have consulted (i.e., actually used in writing the chapter) are listed under “References.” Students can usually identify relevant material simply by its title. “Further Reading” contains relevant geographic studies or foundations that have not actually been used in writing the chapter. Some of the “Further Reading” entries were cited in past editions or are important, somewhat parallel, works. Many of them are suitable for term papers and further study. Finally, specific original ideas, quotes, or individual studies used for data are specifically cited in the body of the text. Thus the chapter on disease diffusion has more references than the others, because so many of the ideas, methodologies, and research findings have been contributed by a few individuals, and so few of these have yet become general knowledge.

Our hope is that this text will be a sound foundation for the future development and practice of medical geography, and that it will inspire geographers and others to bring their own special interdisciplinary knowledge and theoretical approaches to enrich and advance this growing course of study. Finally, we wish to acknowledge and appreciate the examples and even words in this text that belong to the three coauthors of the first and second editions, John W. Florin, Wilbert M. Gesler, and Robert J. Earickson, whose work has become an indistinguishable part of the text. We also acknowledge Cameron Taylor, a University of North Carolina at Chapel Hill medical geography undergraduate student, who helped with graphics, cartography, and picture taking in Malawi.

MELINDA S. MEADE
MICHAEL EMCH

Contents

1. Questions of Medical Geography	1
<i>What's in a Name?</i>	2
<i>A Brief History of Medical Geography</i>	9
<i>Definitions and Terminology</i>	17
<i>The Challenge of Medical Geography</i>	21
<i>References</i>	22
<i>Further Reading</i>	24
2. The Human Ecology of Disease	26
<i>Health</i>	27
<i>The Triangle of Human Ecology</i>	30
<i>Transmission and Creation of Infectious Disease</i>	44
<i>Nutrition and Health</i>	55
<i>Conclusion</i>	65
<i>References</i>	65
<i>Further Reading</i>	67
VIGNETTE 2.1. BIOLOGICAL CLASSIFICATIONS OF IMPORTANCE TO HEALTH	68
3. Maps and Geographic Information Systems in Medical Geography	73
<i>Cartography of Disease</i>	74
<i>Types of Maps: Some Issues</i>	77
<i>Geographic Information Systems</i>	83
<i>Disease Maps on the Web</i>	91
<i>Conclusion</i>	93
<i>References</i>	94
<i>Further Reading</i>	95

4. Landscape Epidemiology	98
<i>Regions</i>	99
<i>Transmissible Disease Systems</i>	100
<i>The Landscape Epidemiology Approach</i>	110
<i>The Cultural Dimension of Water-Based Disease Transmission</i>	116
<i>The Cultural Ecology of Tick-Borne and Other Transmissible Diseases</i>	131
<i>Regionalization</i>	136
<i>Ecological Complications</i>	138
<i>Conclusion</i>	142
<i>References</i>	142
<i>Further Reading</i>	143
VIGNETTE 4.1. FIELD MAPPING FOR LANDSCAPE EPIDEMIOLOGY	145
5. Transitions and Development	149
<i>Ecologies of Population Change: Multiple Transitions</i>	150
<i>Major Impacts of Population Change</i>	162
<i>Nutrition in Transition</i>	165
<i>The Mobility Transition and Time–Space Geography</i>	168
<i>Disease Ecologies of the Agricultural Frontier</i>	173
<i>Other Development Impacts on Rural Ecologies</i>	180
<i>World Urbanization and Changing Disease Ecologies</i>	184
<i>Globalization of Movements</i>	192
<i>Conclusion: Emerging Diseases in Your Future</i>	195
<i>References</i>	196
<i>Further Reading</i>	199
VIGNETTE 5.1. AGE STANDARDIZATION	201
VIGNETTE 5.2. MICROSPATIAL EXPOSURE ANALYSIS	204
6. Climate and Weather: Influences on Health	207
<i>Direct Biometeorological Influences</i>	209
<i>The Influences of the Weather</i>	215
<i>Seasonality of Death and Birth</i>	221
<i>How Climate Change Is Likely to Affect Health and Disease</i>	227
<i>Conclusion</i>	232
<i>References</i>	234
<i>Further Reading</i>	236
VIGNETTE 6.1. MONTHLY INDEXES	236
VIGNETTE 6.2. SEASONALITY OF BIRTH	238
VIGNETTE 6.3. PHYSICAL ZONATION OF CLIMATES AND BIOMES	244
7. The Pollution Syndrome	247
<i>Toxic Hazards of Natural and Economic Origins</i>	248
<i>Outdoor Air Pollution</i>	249
<i>Indoor Air Pollution</i>	259
<i>Water Pollution</i>	260
<i>Radioactive Pollution</i>	265
<i>Risk Assessment and Prevention</i>	271
<i>Globalization and the Perception of Health Hazards</i>	272
<i>The Geometry of Hazards, Power, and Policy</i>	274
<i>Conclusion</i>	276

References 276
Further Reading 279

8. Political Ecology of Noncommunicable Diseases 281

The Dimension of Mortality 283
The Poverty Syndrome 288
Race in the Study of Health Risks 292
Gender: Women's Health 294
Causal Reasoning and Epidemiological Design 300
Disease Ecology: Cancer 304
Disease Ecology: Cardiovascular Disease 314
Unknown Etiology and Other Questions 325
The Precautionary Principle and Some Political Ecology of Research 329
Conclusion 333
References 334
Further Reading 337

9. Neighborhoods and Health 339

The Concept of Neighborhood Health 339
Neighborhood Definition and Units 341
Analyzing Neighborhood Effects on Health 345
Effects of the Built Environment on Health 346
The Challenges of Neighborhood Health Studies 347
Conclusion 348
References 348
Further Reading 350

10. Disease Diffusion in Space 351

Terminology 352
Disease Diffusion 358
Modeling Disease Diffusion 359
Influenzas 375
Geographic Approaches to the Pandemic of AIDS 380
Other Epidemics 398
References 398
Further Reading 401
VIGNETTE 10.1. DIFFUSION WAVES OR STOCHASTIC SIMULATION OF INDIVIDUAL CONTACT? 403

11. Health Care and Promotion 407

Asclepius: Provision of Medical Care 409
Asclepius Unrobed: Cultural Alternatives and Perceptions 431
Hygeia: Health Promotion 442
References 444
Further Reading 447
VIGNETTE 11.1. APPLICATION OF SPATIAL STATISTICS TO HEALTH CARE DELIVERY 448

12. Scale, Spatial Analysis, and Geographic Visualization 453

Some Issues of Scale 454
The Ecological Fallacy 454

<i>Scale of Analysis and Units of Observation</i>	455
<i>Sources for Spatial Disease Data</i>	458
<i>Visualizing and Summarizing Disease Distributions</i>	459
<i>Spatial Analysis</i>	463
<i>Spatial Statistics</i>	466
<i>Conclusion</i>	468
<i>References</i>	468
<i>Further Reading</i>	469
VIGNETTE 12.1. CHI-SQUARE STATISTICAL TEST	471
VIGNETTE 12.2. REGRESSION ANALYSIS	474
VIGNETTE 12.3. SPATIAL AUTOCORRELATION	479
13. Concluding Words	483
 Index	 486
 About the Authors	 498

ON AIRS, WATERS, AND PLACES

Whoever wishes to investigate medicine properly, should proceed thus: in the first place to consider the seasons of the year, and what effects each of them produces (for they are not at all alike, but differ much from themselves in regard to their changes). Then the winds, the hot and the cold, especially such as are common to all countries, and then such as are peculiar to each locality. We must also consider the qualities of the waters, for as they differ from one another in taste and weight, so also do they differ much in their qualities. In the same manner, when one comes into a city to which he is a stranger, he ought to consider its situation, how it lies as to the winds and the rising of the sun; for its influence is not the same whether it lies to the north or the south, to the rising or setting sun. These things one ought to consider most attentively, and concerning the waters which the inhabitants use, whether they be marshy and soft, or hard and running from elevated and rocky situations, and then if saltish and unfit for cooking; and the ground, whether it be naked and deficient in water, or wooded and well watered, and whether it lies in a hollow, confined situation, or is elevated and cold; and the mode in which the inhabitants live, and what are their pursuits, whether they are fond of drinking and eating to excess and given to indolence, or are fond of exercise and labor, and not given to excess in eating and drinking.

From these things he must proceed to investigate everything else.

—HIPPOCRATES (c. 400 B.C.)

CHAPTER 1

Questions of Medical Geography

Medical geography uses concepts and methodologies from the discipline of geography to investigate health-related topics. The viewpoint is holistic, dealing within a variety of cultural systems and a diverse biosphere. Drawing freely from the facts, concepts, theories, and techniques of other social, physical, and biological sciences, it approaches health and disease through its own core questions and perspectives, and uses its own techniques of spatial analysis. Health/medical geography is an integrative, multistranded subdiscipline that has room within its broad scope for a wide range of specialist contributions.

Medical geography is both an ancient perspective and a new specialization. As illustrated by the quotation preceding this chapter, Hippocrates (c. 460–377 B.C.) was familiar with the importance of cultural–environmental interactions more than 2,000 years ago. The study of these interactions, which are important to disease etiology, health promotion, and health service provision alike, continues to this day as medical geography. As old diseases, almost forgotten, are reemerging amidst new risks; as the majority of the world’s population becomes urban and moves toward stabilization of numbers; as biotechnology transforms medicine, agriculture, and our understanding of the nature of life; and as climate change, air and water pollution, metastasizing consumption, and parasitic inequalities transform the ecology of disease—so the ancient study about how people, their cultures and societies in different environments, create and spread disease, promote health, and provide care for the ill has never been more relevant. It is important to understand the interactive processes of cultural and environmental change, as well as the importance of distance and location. By doing this, medical geography can not only advance knowledge, but be useful for analysis and have applied outcomes.

This first chapter undertakes three things. First, it introduces medical geography and health geography. To do so, it presents a framework for the complexity of this specialization and explains something of the divergence of “ecological” para-

digms between the social sciences and biological sciences, which medical geography straddles. It also takes a look at the overall questions and nature of the discipline of geography. Second, this chapter addresses the history of medical geography, especially in the United States, and its evolution in a changing world. Third, it briefly introduces some of the most basic terminology for talking about the subject. It closes by revisiting “the challenge of medical geography” to geographers.

WHAT’S IN A NAME?

The term *medical geography* has more than half a century of use in the United States. As it has developed in that time, it has, in our opinion, become inclusive of concepts, methodologies, and subject matter related to the geography of health and disease. It arches over biological, environmental, and social sciences; it uses both the quantitative and qualitative methodologies of cartography and statistical analysis, as well as participatory fieldwork, interview techniques, and even perceptions best addressed through the humanities. *Health geography* has become a popular term as the social dimension of health has become the central concern in economically advanced countries. The need for “health promotion” and investment in “healthy cities” and for understanding of the experience of the disabled, elderly, and otherwise vulnerable individuals has become obvious as our population has become older, more obese, more sedentary, more drug-addicted, and more promiscuous. Health geographers usually consider medical geography too concerned with disease etiology; with accessibility to treatment; and with models that can be used to generalize, predict, and intervene—in short, with “medicine.” Medical geographers usually consider health geography too limited to behavioralism, ignoring environmental and biological contexts of disease, and denying cultural interactions with environmental influences lest these be considered “determinants.” Studies often involve few interviews, sometimes just a single insight. Concerned with social-theoretical constructs, health geography basically seems to expand upon just one dimension of the triangle of human ecology (presented in Chapter 2).

The accelerating rush of globalization is spreading and connecting political/social/economic structures and behaviors around the world in ways both deleterious and advantageous to health. Individual social behaviors and priorities of society that increase well-being act to promote health and prevent disease. Those that expose people to more risk or increase hazards in their environment promote disease—of body, mind, or society. Medical geography and health geography both view health as more than the absence of disease. Both—as geography—work across a variety of scales, deal in space, and are concerned about place. By whatever name, the subdiscipline is really *public health*^{*} *geography*—a rather awkward but more

^{*}The School of Public Health at the University of North Carolina at Chapel Hill, for example, is composed of departments of epidemiology, biostatistics, environmental science and engineering (developed out of sanitation), maternal and child health, nutrition, health behavior and health education, and health policy and administration—all nonspatial and increasingly aware of the deficiency! Geographers are involved with all of them.

descriptive name. In the spring of 2009 the Medical Geography Specialty Group of the Association of American Geographers (AAG) recognized the contributions and research perspectives of all its members by renaming itself the Health and Medical Geography Specialty Group.

Some Complexities of Geography and of Ecology

Geography is a complex subject, as the above-described arguments about a name for our subdiscipline suggest. Some of the points of contention among geographers and certainly among epidemiologists arise out of the layers of complexity that the term *ecology* has developed. This section explains more about the nature of geographic study. It looks at some theoretical differences between medical and health geographers, as well as between these and epidemiologists and sociologists, in more depth than may be suitable for every level of student or every use of this text.

One way of viewing the complexities of the discipline of geography is portrayed in the matrix of Figure 1.1. The matrix shows what are known as the *topical* or *systematic* specializations of study within the discipline (i.e., the subdisciplines), each of which usually has a *cognate* discipline of study, which is aspatial. Medical geography's aspatial cognate is public health. Studying one or more topical aspects over space, across multiple scales, addresses the core questions of geography: "Where?" and "Why there?" Why are things—rainfall, carcinogens, primary care

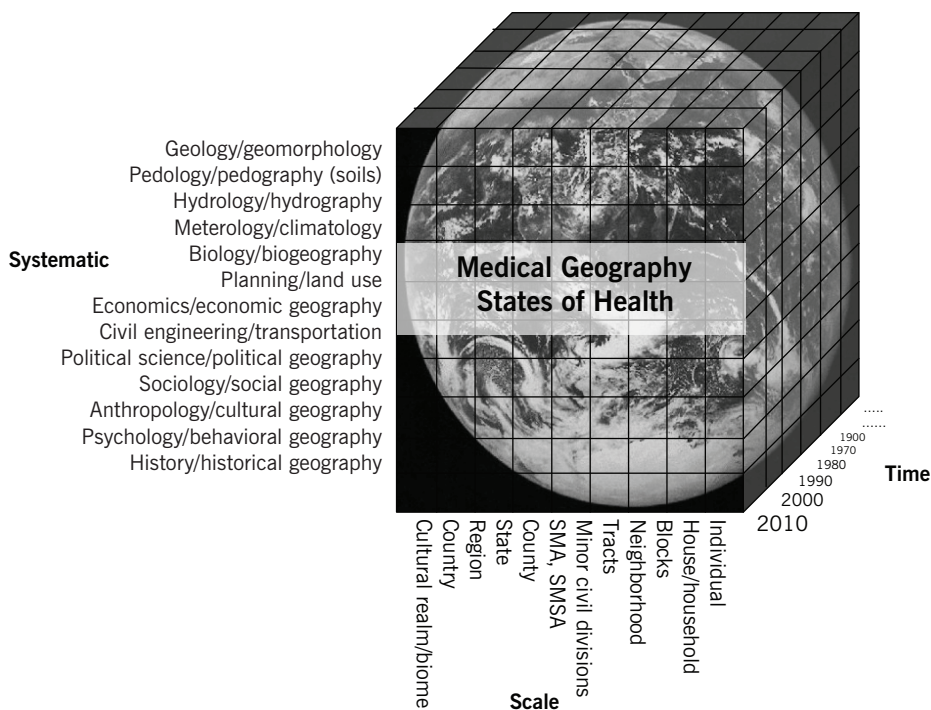


FIGURE 1.1. The matrix of geography.

physicians—distributed unevenly over space? Why are they located where they are? The question “Therefore, what?” (i.e., what the interactions and consequences of that distribution in space are) is addressed through the holistic integrations of region. The scales in Figure 1.1 for regionalization may be those of analysis, or they may be those of integration of all relevant topical processes and variables, which occur together in space and create its reality—and its state of health. The distinctiveness—the uniqueness of *place* that can be so emotively powerful—is thus a microregionalization, not different in kind from those considered at different scales: as country (the United States), as region of country (the South), as city (home town), or as census tract (the neighborhood).

Integrating the interactions in a place and a time is a difficult endeavor approached through a variety of methodologies. These include qualitative ones, such as fieldwork, participant observation, interviews, and focus groups; they also include more quantitative ones, such as multivariate statistical analysis of pattern, mathematical analysis of diffusion, or remote sensing (satellite images) of land use change. Many kinds of study and realms of knowledge can be bridged and connected in geography because they all occur in space and can be integrated in place or region. As the cube of Figure 1.1 shows, the study of air, water, soil, and their processes and expression in the physical environment; the study of biological systems, from molecular to organism to ecosystem; the study of society in varied culture systems, with different religions and governmental forms; the study of individual people and their perceptions and beliefs, as well as genetic constitution—all these and more are connected by their overlay in space. Not surprisingly, academic divisions between the “hard” sciences and social ones, and of both types of sciences with meaning and perception in the humanities, are all mirrored in arguments and divisions within geography. There are so many ways to classify and to interrelate things. True integration even in place is so difficult!

Various efforts have recently been made to conceptualize the complexity discussed above in terms relevant to the developing theories and structures of the broader science community. Geographers on the National Research Council (1997, p. 29) have emphasized synthesis to group the topical interests into three sets: environmental systems (alias physical geography)—topics in Figure 1.1 from geomorphology to biogeography), environmental/societal dynamics (interactions), and human/societal dynamics. Sections of major journals (such as the *Annals of the Association of American Geographers*), departmental groupings of faculty interests, and even the courses required of undergraduate majors increasingly follow suit. Subjects like urban land use planning, cultural beliefs about the causation of disease and nature of its treatment, globalization of diabetes and lung cancer, or purposes and scales of population mobility, however, are hard to separate in this manner. Previous editions of this text have emphasized Pattison’s (1964) “four traditions” of geography: earth science; spatial (and locational) analysis; man–land (today, society–environment, human–environment, nature–society, or [as in this text] cultural ecology); and regionalization (integration). They, and all geography, address questions about why things are located and distributed in space the way they are

(where); why they are moving here and not there (migration, circulation, diffusion); what importance distance has; how and why people vary in perception of these things; and how objects, landscapes, ideas, and people interact and overlap in space to create *place*. There is today a greater emphasis on the larger picture in the context of society. Whether this is addressed through concepts of structure, agency, power, and dependency (in social theory); through identifying the overarching economic and political processes in globalization (in political economy), or through questions about decision, purpose, gain, and loss at different scales (in political ecology), often the individual, local, micro-scale, “where the rubber meets the road” interactions and effects are neglected.

What’s in a name? There are so many disciplines and perspectives involved in the study of health and disease that scholars and professionals can often talk past each other by addressing totally different paradigms even while using the same words. Probably the most fundamental of these is the term—the process, the system, the fallacy, the methodology—*ecology*. Consider the definitions and disciplinary descriptions in Figure 1.2.

The concept of ecology changed within biology as it matured from an emphasis on organisms to an emphasis on relations and functions of the system. The foundational biological concepts and processes, however, entered in the 1960s into cultural ecology in both anthropology and geography, and thus into the emerging subdisciplines of medical anthropology and medical geography. The concepts of cultural–environmental interactions in layers (scales) of systems seemed especially useful for understanding the basis of infectious and parasitic diseases in the newly independent tropical countries. As the ideas of systems, multiple layers, interactions, and complexity entered the social sciences as social ecology, Hawley’s *ecology* as a “descriptive term applied to complex relationships” was increasingly used to focus on relationships between individual people and their social groups. The connection with the natural environment and processes was progressively lost. The environment became mostly social, although some built environment of urban blight or noxious facility was occasionally included. What were measured and modeled were the socioeconomic and demographic variables. They were studied with new statistical methodologies, factor analysis and its relatives, which were able to model complex relationships. The dimensions of socioecological differentiation thus found could be used to develop theories of socioecological structure, beginning with *macro*-level units based on sets of *variables*. Through sociology into political science and other social sciences, including social and urban geography, *ecological* came to mean complex, multivariate, quantitative (statistical, factorial) analysis of characteristics of people, their groups, and social/economic/political conditions (perhaps housing, crowding, or other “environmental” effects), conditions identified and measured as variables. The macro-scale approach was not about individuals, families, streets, and neighborhoods, but about populations, such as teenagers or households below the poverty line in income.

Ecological studies thus became almost anathema to epidemiology, since they seemed to be full of confounding variables and spoke to population-level data

Ecology is the study of the interactions between living organisms and their biotic and abiotic environments. (Odum, 1959—biology)

Ecology is a descriptive term applied to complex relationships between organisms and their environment. (Hawley, 1950—sociology [human ecology])

Social ecology is the study of the relationships between individuals, social groups, and their environments. (Sociology [human ecology])

Let us define **social ecology** as the part of **human ecology** that deals with social and societal aspects of that environment and thus fall within the realm of society. **Factorial social ecology** is the orderly factor analytical approach and model. [It] tries to establish sets of basic dimensions in socioecological differentiation to reach empirical generalizations about these dimensions and later to develop a theory of socioecological structure and change in terms of the dimensions. . . . Typical study begins by describing a set of socioecological macro-units by means of a set of variables which then are factor analyzed. (Janson, 1980—social ecology)

Types of epidemiological studies include cross-sectional, proportionate mortality studies, and **ecological studies**, in which the unit of observation is a group of people (instead of individuals). Such “ecologic studies” have data averaged over population. Therefore the degree of association between exposure and disease is usually more tenuous. . . . [The] greatest difficulty is [that] the unavailability of data necessary for control of confounding in the analysis . . . produces results of questionable validity because of “inappropriate inferences”: the problem of the “ecologic fallacy.” (Rothman, 1986—epidemiology)

The **social ecological** perspective incorporates concepts from systems theory. . . . The field of social ecology gives greater attention to the social, institutional, and cultural contexts of people—environment relations than did earlier versions of human ecology. . . . In the area of health promotion, the multi-level aspect of the assumption that the effectiveness of an intervention can be enhanced significantly through the coordination of individuals and groups acting at different levels . . . is illustrated by family members who make efforts to improve their health practices, the efforts of managers to shape organizational health policies, and the activities of public health officials who direct community health services. (Stokols, 1992—social ecology)

A core principle of social ecology is that the environmental contexts of human activity function as dynamic systems. This systemic quality of settings is reflected in the interdependencies between physical and social conditions within particular environments and in the nested structure of multiple settings and life domains. (Stokols, 1996—social ecology)

The term **social ecology** refers to the nested arrangement of family, school, neighborhood, and community contexts in which children grow up. (Earls & Carlson, 2001—social ecology)

The development of authentic selfhood means the simultaneous unfolding of both individuality and social being. The replacement of the voracious yet fragile and underdeveloped ego of consumer society with such a richly-developed selfhood is one of the preeminent goals of **social ecology**. (Clark, 1998—social ecology)

Political ecology is concerned with how higher-level socio-economic and political structures and processes create the context in which local cultural and individual behavior interacts with the environment.

FIGURE 1.2. Definitions and disciplinary descriptions of ecology.

and not individual-level causation or intervention. *Ecological* came to mean population studies: multivariate, multiple-scaled perhaps, complex, uncontrollable, and fuzzy—interesting, perhaps, but of limited scientific use. Data “averaged” for a group made causal connections between an exposure and disease results very murky. A group average could neither connect individuals of that group with exposure or results, nor connect larger populations or other groups with those average exposures. This is the *ecological fallacy*—the idea that associations statistically identified at one scale of analysis are valid at or can be generalized to either larger or smaller scales. (The ecological fallacy is discussed at several points later in this text, most notably in Chapter 12.) The real environment and its ecology largely disappeared from perspective. Epidemiology (and the U.S. Surgeon General) could say by the 1980s that more than 80% of cancer was “environmentally caused,” meaning caused not by pollution or radiation, but by behavioral inputs like smoking tobacco. This “environment” was that of a cell, not the biosphere of a human population. Health promotion became concerned primarily with the institutional and structural impediments to changing behavior and accessing health care. For a couple of decades, epidemiological questions were mostly limited to small controlled populations, and any generalizations (such as those about cigarette smoking and cancer risk) were sought through methodology from them. One of the most exciting things happening in public health today is that the power of geographic information systems (GIS) is allowing epidemiology to ask questions and consider relationships at different scales again. Social ecology, as evidenced in Stokols’s influential public health model, held firm to some context of people–environment relations. It was concerned with coordinating actions at multiple levels. These have become “nested” relationships of schools, neighborhoods, and communities within cities, states, and nations. It is still rare, however, to find anything about the natural environment in these studies.

Medical geographers throughout the 1970s and 1980s could differentiate between two “schools” of ecology: disease ecology and health service delivery. Social/urban/economic geographers drew on their models to study health service delivery. They used *ecology* to mean multivariate complexity as the sociologists did, although as geographers they recognized the importance of study at different scales. They sought to understand the roles of distance and accessibility in determining the optimal locations to provide health services and to locate physicians and other caregivers, facilities, and equipment. They also sought then to determine the obstacles and social/political/economic creation of need and obstacles to its provision. (The main body of their work is explained in Chapter 11.) In contrast, geographers who were especially concerned with human–environment interactions identified primarily with cultural ecology. Their study of disease ecology became inclusive of the influences in the biological and physical as well as the social environment as they modeled disease etiology. They sought to integrate in place in order to understand the creation and maintenance of health and disease systems. This would enable the vulnerable points and protective interventions to be identified, and so to permit illness to be reduced and health promoted. Over time, disease ecology became

concerned not just with infectious diseases, but with the cultural ecology of degenerative diseases and other health conditions of people in economically and demographically advanced countries as well. Such studies have also needed to address higher social/economic/political structures, processes, and questions. In recent years, developing a theoretical framework for how social and economic power create and structure the processes of risk and the expressed conditions of health and disease have become important. As Harold Brookfield (2005) wrote in his review of Paul Robbins's (2004) introduction to political ecology, Robbins explains how political ecology developed separately from cultural ecology because of the "inability of apolitical forms of explanation to answer critical questions concerning why landscapes are degraded, constructed, or conserved" (p. 286). The development of ideas in political ecology connected with peasant studies, feminist development studies, and others, but not the field-generated cultural ecology studies in anthropology and geography. Neo-Marxist influences predominated in early writing of political ecology, Brookfield notes, but a pluralism of thought has developed and class analysis need not be its central issue. Intellectual struggles abound between social scientists, for example, arguing over whether there is no theory in political ecology or whether it is so overwhelmed with theory that outcomes are predetermined. Political ecology is itself a large subject about many aspects of geography. Perhaps that is the cause of its amorphous nature and vague definitions. Its relevance to health is more clearly presented by some studies in cultural ecology/medical anthropology, which focus on such topics as household ecologies of disease transmission and others focusing on the differential distribution of risk of disease (Inhorn & Brown, 1997). Although Jonathan Mayer identified the political ecology of disease as a new focus for medical geography, it is not one that has actually in the last decade produced research to be presented or discussed. Perhaps this text can encourage some. (Most of this work is presented in Chapter 8.)

Many geographers called everything that was not about health service *disease ecology*. This included studies of disease diffusion, methodologies of mapping disease, and social-factorial analyses of the pattern of disease. For example, a journal editor looking for a reviewer for a study that mapped cancer, modeled influenza diffusion, or factor-analyzed housing and alcoholism would describe the study as "disease ecology." This dichotomous approach was and still is a source of confusion, however, as these studies in medical geography seldom involved environmental interactions or ecological processes. It continues to confuse things between health geography, the successor to the social school, and medical geography, which is more than disease ecology.

Since most research specialists tend to read research literature relevant to their own paradigms and approaches, the cross-definitions and evolved usages explained above became and have remained major sources of misunderstanding. The reader may have noticed that no references are given for the discussion above; it is our own understanding of what has happened over the past 40 years. With the help of GIS (see Chapters 3, 9, and 12), most of these ways of studying health and disease today seem to be converging—but researchers are still speaking the language of different paradigms, as the following brief history of medical geography makes clear.

A BRIEF HISTORY OF MEDICAL GEOGRAPHY

Place was important to medicine until the middle of the 19th century. For 2,000 years medicine was concerned with geographic variations in air, water, soil, and vegetation; insects and animals; diet, habit, and custom; clothing and house type; government; and economy. The late 19th century saw a paradigm change—that is, a change in the great overarching idea of how disease occurred and what questions were worth asking. In the 20th century *germ theory*, otherwise known as the “doctrine of specific etiology,” resulted in revolutionary advances. *Specific etiology*—one cause (germ) that is both necessary and sufficient for each disease—is less relevant to a society in which people die from heart disease, cancer, kidney failure, alcoholism, and violence. These are diseases of multiple, complex causes based as much in culture and society as in biology. Even for infectious diseases, germs are no longer considered to be the “sufficient” causes. The tubercle bacillus is necessary to cause tuberculosis, but the disease and its course depend on many factors: nutrition; genetics (resistance); treatment; bacillus Calmette-Guérin (BCG) vaccination (see Chapter 2); the presence of disease-promoting conditions (especially HIV today); crowding, ventilation, and housing quality; the custom of spitting; and mental attitudes. The etiology and control of strokes or diabetes are even more complex.

The progress of specific etiology was paralleled by the progress of specialization in medical study and treatment, and by the increasing divorce of body from mind and environment (and even organ from body, and cell from organ). As the contradictions between the dominant biomedical orientation and the health needs of people increased, the social sciences became more involved. Besides medical geography, the last few decades have seen the development of flourishing concentrations in medical anthropology, medical sociology, medical economics, and medical psychology. Historians also have reconsidered the significance of disease for major social and economic change and the consequences of the connection of empires.

Geographic variation in health and disease has long been studied under such interdisciplinary rubrics as geographic pathology, medical ecology, medical topography, geographical epidemiology, geomedicine, and so forth. The perspective that cultural–environmental interactions were important for disease etiology and for health continued to be philosophically important, even dominant, until the emergence of germ theory. Thus the 18th- to 19th-century physicians (e.g., Finke, 1792–1795) who first used the term *medical geography*, and who struggled in dozens of works to describe and organize the avalanche of new information about human diseases, cultures, and environments that came with world exploration, were continuing the holistic Hippocratic tradition. Their descriptions of where diseases were endemic and where they didn’t occur, and their explanations for this and for epidemic spread, are being rediscovered and reevaluated by geographers and other scientists once again concerned with disease ecology (Barrett, 1980). Barrett (2000) established that “medical geographical and geographical medical” research reached its zenith in the decade of the 1850s with the publications of Drake (1854), Fuchs (1853), and others and was being increasingly challenged. It had been thought that understanding the geographic patterns might lead to understanding why there

was disease, but the rise of bacteriology brought other answers. By the time Hirsch (1883–1886) published his masterful handbook, the “doctrine of specific etiology” of disease was finding one bacterium after another. The discovery that microbes invade human bodies and cause alterations that result in disease led to asepsis and sterilization, vaccination, antibiotics, chlorination of water, treatment of sewage, and over 30 more years of life for the average person in a developed country. The enormously successful paradigm of germ theory is probably responsible for the existence of half the people reading this book, as without it they would never have been born.

After World War II, with the development of widely available antibiotics and powerful insecticides such as DDT, people dreamed that infectious disease might become a thing of the past. Some scientists, especially bacteriologist René Dubos, thought that hope a chimera (see Chapter 2). At the same time, however, degenerative diseases such as heart disease, stroke, cancer, liver and kidney failure, respiratory obstruction, and such became the major killers and disablers. It was obvious that these diseases did not have a specific etiology; they had complex etiologies involving many factors—cultural, social, and environmental, as well as biological. New theories about society and cultural behavior were needed to explain their causes and to guide research and interventions.

The perspective and methodology of geography have now been applied to the study of health, disease, and health care for over 50 years. The emergence of a systematic (topical; see Figure 1.1) interest in medical geography can be dated from the first “Report of the Commission on Medical Geography (Ecology) of Health and Disease to the International Geographic Union” in 1952 (Geddes, 1978). The next 15 years saw pioneering researchers and teachers in a dozen countries develop a substantive, international focus on medical geography. This is described in the first few chapters of a Festschrift to Andrew Learmonth (McGlashan & Blunden, 1983). The following sections follow the developments in the United States.

Jacques May

Jacques May was the “father” of medical geography in the United States: He initiated the subdiscipline (May, 1950, 1954) with *Medical Geography: Its Methods and Objectives*, developed the first maps of global distributions (May, 1950–1954), and wrote *The Ecology of Human Disease* (May, 1958).^{*} This represented the culmination of a long intellectual evolution on May’s part. He had begun as a French surgeon in Siam (now Thailand) and spent more than a decade as head of surgery at the major French hospital in Hanoi, Vietnam. He started his intellectual journey by questioning why his patients in the tropics experienced and responded differently to disease and surgery than did the European patients described in his textbooks.

^{*}To this day, the annual award of the Medical Geography Specialty Group of the Association of American Geographers for the best thesis has been the Jacques May Award. Even some of those who receive it do not know why!

He progressed to an interest in his patients' multiple, simultaneous infections and then to the conditions of their lives. Then, from his own travel and experience, he reached an understanding of the cultural and environmental conditions that produced and limited their health and disease. May was influenced in his development of disease ecology by the French approach to geography, which emphasized distinctive local regional expressions of interaction between human and physical domains, expressed as *genres de vie*. He referenced Sorre (1947) in his own seminal article (1950), including M. Sorre's "pathological complexes" (Barrett, 2000). World War II found him in New York City, where he conducted research at Rockefeller University; he then taught briefly at several major northeastern universities.

Jacques May became the founding director of the American Geographical Society's Medical Geography Department. By 1951 he was publishing the first maps of distribution of species of helminthes and microbes, and their tick and mosquito vectors; of the distribution of diseases such as leprosy; and of key variations in diet and nutritional deficiencies. Plate 1 ("Distribution of Malaria Vectors") and Plate 2 ("Study in Human Starvation") reproduce two of these ground-breaking maps, which were published as inserts in the *Geographical Review*, the journal of the American Geographical Society. The Plate 2 map of diet and nutritional deficiencies is especially notable because May spent the next decades publishing almost annually, from 1961 to 1974, books on the ecology of malnutrition in a world cultural region (listed in "Further Reading" for Chapter 2). One of us (Melinda S. Meade) remembers reading in a long 1975 obituary in *The New York Times* (from which she learned most of the personal information above) that Jacques May died in a jeep accident in Tunisia while doing fieldwork at the age of 79.

The Early Years

John Hunter (1966, 1974) taught medical geography at Michigan State University from the 1960s until the 1990s, and from that large department seeded American colleges with geographers who had at least that initial exposure to the subject. In the 1960s the curriculum of human ecology at the University of Illinois strongly affected the founding of a medical anthropology curriculum at the University of California at Berkeley; it also influenced J. Ralph Audy (1971), the chair of international health at San Francisco Medical School. From the Illinois curriculum also came Warwick Armstrong (1965, 1973), a New Zealander and the first medical geographer who was also a professor of epidemiology (at the University of Hawaii). Meade became the first American geographer to do both a master's degree (with Hunter) and a doctorate (with Armstrong) in medical geography (as cultural ecology). (Your other author, Michael E. Emch, was Hunter's last doctoral student!)

Also in the 1960s, the first studies on locational analysis for health services—the ground-breaking regionalization of the Swedish health system by geographers at the University of Lund (Godlund, 1961)—occurred, and Richard Morrill brought both Gerald Pyle (1971) and Robert Earickson (Morrill & Earickson, 1968) into the geography of health services through the Chicago Regional Hospital Studies. Soon

afterward, Gary Shannon and Alan Dever (1974) wrote the widely read book that first introduced a generation of graduate students in geography to the subject: *Health Care Delivery*. Settlement forms, systems of cities (urban hierarchy), transportation networks, and population mobility had become objects of modeling in urban geography. As a doctoral student in urban geography at the University of Chicago, besides writing a dissertation on future heart disease and its spatial need for hospital care in that city, Pyle (1969) studied how the changing process of diffusion of cholera in the 19th-century United States reflected the maturing infrastructure and urban system of the country. In doing so, he wrote the seminal paper in spatial analysis of disease diffusion (see Chapter 10).

As can be discerned from the summary above, medical geography was from the beginning concerned with disease ecology, mapping of disease, spatial analysis, and disease diffusion, as well as locational analysis and regionalization of health services. In workshops, proposals (for funding) for atlases, and dozens of other ways, efforts were made to integrate these approaches despite root differences. For example, it was thought that disease ecology and mapping, spatial analysis of demographic and socioeconomic data, could not only raise robust hypotheses for the unknown etiology of diseases, but would identify spatial patterns of need for the health services to meet, thus optimizing facility and specialist distribution by locational analysis. The efforts to connect the two research dimensions never really succeeded, however.

At its November 1974 meeting, the Council of the AAG established a Committee on Medical Geography and Health Care (the source for this paragraph is Meade's personal archive). Pyle was appointed chair, with a committee composed of Armstrong, Dever, J. L. Girt (a Canadian), and Hunter. By January 1975 this committee had proposed an atlas with chapters on these topics: the "continental ecospatial framework," "biogeochemical environment and health," "man-made environmental hazards," "organizations and use of health services," "illness and health behavior," and "policy planning and implementation." A similar framework, though with much more emphasis on health services, could have been used for paper sessions at meetings 20 years later. In 1975, also, the Medical Geography "Special Interest Group" (as the AAG then considered it) elected its first steering committee, with Meade emerging with the most votes as chair. By organizing in 1975 two sections of paper presentations (one in "general" medical geography and one in disease perception) and offering a sold-out workshop ("The mini-course organized by Gary Shannon is designed to provide an organization scheme for health/disease research, a working knowledge of appropriate data collection and analytic techniques, and health planning program developments pertinent to geographers"), the Medical Geography Special Interest Group became a model for the reorganization of the AAG and its national meetings by specialty group. In January 1979, having gathered an abundance of signatures following the protocol for founding the new AAG organizations, Meade became the first chair of the new Medical Geography Specialty Group—the organizing body today, now known as the Health and Medical Geography Specialty Group.

The Middle Modeling Years

The numbers of medical geographers and of their college course offerings increased in the 1980s, but government funding was almost exclusively allotted to health service research. The geographic research agenda followed the funding. Annual meetings presented hours of papers on health service location and utilization; scores of medical geographers found jobs in planning and marketing. The U.S. health care system was an “un-system” that had grown rapidly without planning or organization at any level of government. Cities and states did not know where hospitals were located, let alone doctors or specialists. Unschooled in geography, those who began to grapple with provision of services, construction of needed beds, location of expensive new scanners and other technology, and racial desegregation, among other things, had no idea about the friction (and its costs) of distance or how to optimize the locations of or accessibility to new facilities. Techniques and computer programs were developed to regionalize emergency rooms and route ambulances, as well as to estimate the accessibility consequences of hospital closings or retirements of physicians in rural areas. The classic “journey to work” studies and their locational analysis models of economic geography became health service demand and market area analyses (see Chapter 11).

As geographers began to focus on the spatial distribution and inequalities of health care provision, it became clear that the way countries organized their health care systems had a major effect on the dimensions of health inequality. Concepts of political economy became increasingly important. Such topics as public policy and the structure of society moved out of sociology into geographic analysis (Eyles & Woods, 1983) and across the sea from Britain to geographers involved with the national Canadian health promotion endeavor. Studies increasingly focused on topics such as territorial injustice, environmental racism, and the structural context of why people get sick, rather than simplicities of distance. Jones and Moon's (1987) influential book moved the social context of not only health care but health itself into the mainstream of geographic study. The deinstitutionalization of persons with mental illnesses in the United States, and the failure to develop the community service provision and shelter to care for those affected by mental illness, substance abuse, or domestic violence, influenced the focus and methodologies of geographic research and exposition (Dear & Wolch, 1987; Wolch & Dear, 1993).

During this time there were also developments in spatial analysis, which involved fewer geographers but had a large impact upon public health. Spatial analysis is so intrinsic to geographic thinking that many geographers would find it strange to discuss it separately. In medical geography, however, it has received special attention in two areas: the study of disease diffusion and the analysis of diseases of unknown etiology. When the epidemiological and mathematical study of the spread of contagious disease within a population and its subgroups is considered spatially, some of the most authoritative models in geography come to bear on population movement, settlement hierarchy, and transportation network. Analytical models of the diffusion of influenza, measles, cholera, and hepatitis during this time became elaborate and powerful, as reviewed by Cliff and Haggett (1988) and

Thomas (1992) (see Chapter 10). A different aspect of spatial analysis has involved study of the covariation in space of disease occurrence and possibly related factors of environmental and social conditions. As noted above, these studies have often been classified with disease ecology, but the “ecology” in them has been of the social-factorial kind and during this time involved little or no “real” environment (see Chapters 8 and 9).

The special and basic tool of geographers, in the service of all its traditions, is the map. *Cartography*, the construction and interpretation of maps, has held a central place in geography for more than 2,000 years. Most geographers have a profound love affair with maps. An old saying in the field is “If it can’t be mapped, it’s not geography.” Although a wide variety of statistical techniques are used in geography, as in other social sciences, maps are unique, powerful, and flexible tools for analysis of geographic phenomena. A map is a model of the world that, through the use of point, line, and area symbols, can integrate many dimensions of reality. Following on the great success of a world atlas produced at Heidelberg University in Germany (Rodenwaldt & Juszatz, 1952–1961), the studies in geomedicine at Heidelberg (Juszatz, 1968–1984) attempted to pull the multiple dimensions of the geography of health and disease together in place and region by mapping. In the 1980s the ancient dimension of cartography as an art as well as a science gave way to the accessible but mindless precision of computer cartography. In the 1990s this gave way in turn to the larger application of the computer to the methodologies of spatial data management GISs (see Chapter 3). GISs have not so much changed the questions that geographers ask, as revolutionized their ability to answer them. The long history of mapping in geographic and public health analyses of prevalence, etiology, diffusion, and health care distribution is exhaustively and beautifully related in Koch’s (2005) *Cartographies of Disease*. Cliff, Haggett, and Smallman-Raynor (2004) have shown multiple forms of mapping as they have taken on global, regional, and local maps and interpretations of 50 of the major epidemic diseases in their beautiful *World Atlas of Epidemic Diseases*.

Recent Times

Over the last decade or two, developments in availability of digital data; in theory in social science; in population mobility and world urbanization; and in globalization of pollution, trade, and other processes that could be listed to fill this page have changed everything—and changed nothing. In this section we look quickly at developments in social theory, health geography, place and therapeutic landscapes, and political ecology.

Social science has seen fundamental questioning of and turbid debates over epistemology. One result is a social-theoretical framework that is much more interdisciplinary than before, and that values qualitative methods and is more open to questions about meaning (previously the domain of the humanities). Geographic concepts of place, landscape, and global interconnectedness have become familiar and commonly expressed in other disciplines. They have also become more domi-

nant concerns in geography. Social theories variously known as *social-interactionist*, *structuralist*, *structurationist*, *poststructural*, and *critical* have progressed from being positivist—empirically trying to verify data and test observations and models (disease etiology)—to concerns about interpreting the “meaning of illness” to an individual and, with or without human agency, describing the deep structures of commercial and political domination and the inequalities they produce (see Gattrell, 2002).

Kearns and Moon (2002, p. 606) have identified three main themes in a new health geography that distinguish its concern for health and well-being (and distance it from disease concerns and all things medical): emergence of place as a framework for understanding health; application of social- and cultural-theoretical positions to health; and a quest to develop “critical” geographies of health. Their approach to health geography brings the concerns underlying the drive of social theories over the past decades to bear. Individual people are persons, not just “observations”; place is a “living construct” that has meaning, not a container in which to put observations (as when counties are used as units in mapping cases of influenza). Qualitative methods of interview, observation, focus groups, and text analysis are more appropriate for analysis. Well-being and justice are the objects of interest. These perspectives and concerns have been directed at physical disabilities, mental health problems, and old age—vulnerabilities that were seldom addressed before.

Gesler (2003) developed the constructs of landscape and place (Kearns & Gesler, 1998) in an original study of therapeutic places, such as the ancient healing grounds of the sanctuary of Asclepius at Epidauros or the healing waters of Lourdes. Gesler has noted (2005, p. 295) that following Williams (1999), others have taken the concept and used it to look at places that support well-being (such as children’s camps), as well as landscapes of risk.

The development of political ecology discussed above is another response to the need to consider the political, social, and economic structures and processes that construct policy and form the context of local and individual exposure to health risks (conditions good and bad that people must cope with and react to) and medical treatment. Political ecology is quite analogous to political economy, but is concerned with more than the consequences of capitalist production or with power, domination, and blame. Political ecology is explicitly about causal *interactions* with local ecology. (Yes, it can be a two-way flow! Drought can change political and social institutions and structures just as well as result from them.). Political/economic/social policies, structural institutions, and processes are not ecologically neutral. The construction of a large dam and water impoundment to provide electricity for a major city, industrialized jobs for the urban migrants, and profits for the corporations can create a habitat for *Anopheles* mosquitoes and malaria where before there was none. Commercial logging and subsequent settler efforts at agricultural conversion change local ecological systems and reduce feedback, which, with the induced population mobility, makes species connections and promotes the emergence of new disease. Land tenure and land use policies can determine who in what place have no “entitlement” to food and so affect regional levels of malnutri-

tion. And so on. Women are especially vulnerable to these changes in their societies generally and in the world sex trade especially.

Geography and anthropology again come together in the study of these ecologies, but anthropologist/physician Paul Farmer (1992, 1997, 1999, 2003) has shaken the public health establishment with his political ecologies. As Farmer points out, globalization of economic and political power is increasing gaps in income, knowledge, and health everywhere, at every scale. In China, more than half the population has been astoundingly lifted out of grinding poverty. Those who remain rural and poor, however, are more economically, socially, and politically marginalized; are less well nourished; are exposed to more dangerous chemicals; and have less access to good health care than before. Those who are now living and working in the industrial cities are exposed to some of the world's worst air pollution from smokestacks and exponentially growing automobiles, as well as socially promoted behaviors like smoking tobacco. Such consequences as asthma and cancer are already increasing.

Political ecology is concerned with understanding these higher-level or "upstream" causes. However, the process of ecological change stemming from higher-level political causation is not limited to the effects of capitalism or globalization. The policies of Soviet communism created industrially poisoned forests and lakes across Eastern Europe. Feudal agriculture responded to population growth with continental deforestation, erosion, filthy crowded cities, malnutrition, and other conditions that facilitated bubonic plague to remove a third of the population. The domestication of plants and animals and the permanent settlements this allowed created the ecological conditions that created most of our contagious diseases. The "downstream" questions of cultural ecology are about what, exactly: What changed ecologically, from what behaviors, altering what feedback and regulation systems, intensifying what, exposing people to what local alterations that affect their lives and to which they must adapt? What information is needed? Where can interventions be made? How can those conditions that produce disease be diminished, and those that improve the population's state of health be promoted? All social scientists concerned with health agree that poverty is the major, primordial "cause of premature death," and that alleviating it "requires that society takes transcendent [policy] actions" (McMichael, 2001, p. 335). As McMichael (2001, pp. 265–267) further explains, two dimensions of complexity about why "this population, at this time" has a health problem involve (1) "deciding the level and type of analysis of disease causation," balancing the upstream and downstream causes; and (2) "envisioning the evolution of health risks over time, over the course of lifetimes."

Public health research is also struggling again with ecological studies, taken as nested systems of multiple interactive variables. As digital data become more available, the capacity to analyze context and to compare relative influence at different scales is a siren song. The neighborhoods-and-health framework for analyzing multiple nested layers of disease causation is described in Chapter 9. It develops out of the power of spatial data management and the integrative power of spatial organization. The use of GISs in public health is already well under way (Cromley & McLafferty, 2002).

Future Directions

The only thing certain about the future of medical geography is that there will be new developments. The earth is warming, and rain very likely will be falling in different spatial patterns. The human population is slowing its growth and may be stopping it in a few decades. The majority of it has also now become urban, and in developed countries and others (e.g., China), the population is getting much older. Global connectivity continues to increase exponentially, along with what have to be considered revolutions in communication and in diffusion of ideas and information. Satellite imaging can monitor and show us the surface of our planet almost foot by foot. The availability of digital data also continues to increase exponentially, as does the need for means to interpret it and to understand what the almost infinite layers of spatial variation mean to real people leading lives in real places. The human genome has been “mapped,” spelled out letter by letter. It turns out that we are all the same, with barely any adaptations of immune systems and other forms of direct interaction with the varied environments to which we have migrated to live. Our differences are our cultural creations—a project that seems to continue.

Infectious disease has not gone away. New diseases are emerging, and old ones are reemerging. We need to study infectious diseases again, this time in the context of higher-level political/economic/ecological processes. The evolutionary biologist Paul Ewald (2000) controversially claims that these studies of infection should include many human degenerative diseases. Even as the physical and social sciences increasingly struggle with the principle of indeterminacy and the precautionary principle, he suggests that the germ theory standards of evidence may need to be changed. The influence of systems theory on ways of thinking in cultural ecology (Moran, 1990) has also had a kind of revival, developing from the way theorization on information and ecology has emerged from Prigogine’s studies of self-organizing systems (Kellert, 1993; Prigogine & Stengers, 1984). Surely some medical geographer will bring complexity theory and modeling of indeterminacy as a theoretical approach to the organization of information and systems of health.

The intensive development of social-theoretical frameworks now enables the study of those social and economic organizations, belief systems, and technologies that constitute the behavioral vertex of the triangle of human ecology. The methodologies and perspectives of health geography deepen and enrich our ability to understand ourselves. Cultural ecology has as a great claim to study culture, but it needs to inform those spatial-analytical studies more. GISs are creating the opportunity to do that and, through community participation, to return the ever-proliferating microspecializations that fracture geography to the ancient integration of the map.

DEFINITIONS AND TERMINOLOGY

Definitions of Health and Disease

Everyone knows what health is, and yet a precise definition of *health* is difficult to come by. This problem is shared by researchers who, in studying health, ironically

need to measure disease. Health, however, is more than the absence of disease. It is a positive thing—the actualization of the creative force of human life for individuals, communities, and societies. We know that greater health is usually equated with lower mortality and morbidity rates. The problem remains of how to define health without reference to disease.

The first major definition to present health as a positive entity—a presence to be promoted and not merely an absence to be regretted—occurs in the 1946 charter (preamble to the constitution) of the World Health Organization (WHO): “Health is a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity.” This influential statement was important for the philosophical position it stated and for the goals it set for government programs and research funding. It has not proved useful, however, for implementing any standards or research designs that require criteria. It is a utopian ideal. Researchers have for specific studies and purposes created various “indexes” of health, but they seem both too particular and too cumbersome for widespread use.

May’s (1961, p. xv) definition of *disease* was for some years referred to by geographers. He stated that disease is “that alteration of living cells or tissues, that jeopardizes survival in their environment.” There are several important points in this definition. The organism has an environment to which it relates. The idea that disease jeopardizes survival implies that there may be different levels of health without there being disease. An office worker, for example, need not have the physique or eyesight of a hunter. One may be born with a physical disability and lead a productive life into old age, depending on the society within which one lives and its technology. May, however, was a physician before he became a geographer, and it shows; the present understanding of and interest in such dimensions as mental health, mind–body interaction, and sociocultural morbidity context are difficult to reconcile to this more biomedical definition.

An influential definition comes from Dubos (1965, p. xvii): “States of health or disease are the expressions of the success or failure experienced by the organism in its efforts to respond adaptively to environmental challenges.” This definition implies a system whose parts can exist in different states of interaction. Health is not necessarily a condition of physical vigor, but a condition suited to reaching goals defined by the individual. The most important word here is “adaptively.” Dinosaurs were highly adapted to their environment but could not cope with environmental changes. There is a dynamic quality to health. Dubos’s definition, however, defines what health results from, not what it is.

In this book, we adhere mainly to J. Ralph Audy’s (1971, p. 142) definition: “Health is a continuing property that can be measured by the individual’s ability to rally from a wide range and considerable amplitude of insults, the insults being chemical, physical, infectious, psychological, and social.” This definition is elaborated on in Chapter 2, but here it should be noted that health is present until death and is a dynamic quality continually engaged in coping with a changing social and physical environment. Audy says that health can be measured, but this has proved difficult except for such specific aspects as immunological or nutritional health. As

noted above, researchers have used many indexes and surrogate measures over recent years, filling government reports with suggested criteria; however, each definition is limited to its narrow purpose and has never become widely accepted. In the end we are still measuring the absence of disease.

Terminology

A familiarity with some terminology is necessary before the availability and limitations of data on health and disease can be appreciated. This section presents some of the most commonly used terms.

Diseases are referred to as *congenital* when they are present at birth. These may be of genetic origin, as hemophilia is; they may be acquired in the womb, as drug addiction or chemical-induced deformity is; or they may be acquired during the process of birth itself, as when severe inflammation of the eyes results from passage through a birth canal infected with the bacteria of gonorrhea. Diseases are referred to as *chronic* when they are present or recur over a long period of time, and as *acute* when their symptoms are severe and their course is short. *Degenerative* diseases are characterized by the deterioration or impairment of an organ or the structure of cells and the tissues of which they are a part. *Infectious* diseases result from the activities of living creatures, usually microorganisms, that invade the body. *Contagion*, transmission of infectious disease agents between people, may be direct through person-to-person contact or indirect through the bites of insect vectors or via *fomites* (vehicles) such as contaminated blankets, money, or water.

Figure 1.3 illustrates one way of looking at the continuum of health and disease. The term *clinical* refers to the appearance of symptoms that can be presented to a physician for observation and treatment. In a *subclinical* condition, an infectious agent may enter the body, multiply, stimulate the production of antibodies, and be eliminated from the body without the person's being consciously aware of any illness. Usually the only way that subclinical infections can be detected is through serology (the identification of antibodies and other immune reactions in the blood). Quite a few diseases produce acute reactions in only a small proportion of those infected. Other common diseases are mild infections when acquired in childhood

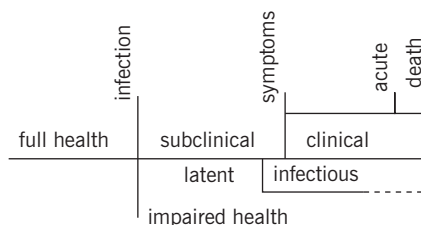


FIGURE 1.3. The health continuum, illustrating terminology and stages of ill health between full health and death.

and often pass unnoticed. In either case, public health officials are sometimes startled to find from serology that a “rare” disease has in fact infected the majority of the population.

Usually an infectious disease has a period of *latency* between the time the infection occurs and the appearance of clinical symptoms. People are sometimes infectious before the disease is manifest. This is because the infectious agent needs time, known as *incubation*, to adapt to its host and multiply and become numerous in its mode of transmission (respiratory droplets, body fluids, etc.). Sometimes, however, it completes this incubation before it causes symptoms (i.e., while it is still latent), and thus can spread before the infected person (host) is treated or becomes bedridden or otherwise restricted in mobility from exposing others. The common cold, for example, usually has an incubation period of 1–3 days, but people may be infectious for 24 hours before their own symptoms appear. Measles has an incubation period of about 10 days until the onset of fever; people are then infectious for the 3–4 days until the rash appears, as well as for several days afterward. People infected with HIV are most infectious in the first months after incubation; the latency period until the symptoms of AIDS appear may be years, allowing untested and unaware people to spread the infection, even through childbirth. Because disease data usually are not produced until clinical symptoms are diagnosed, statistics at any given time usually underestimate the amount of disease in a population. Problems with time lag, apparent health, and time of diagnosis are accentuated when degenerative diseases are studied. The latency period between initial stimulus and the diagnosis of presented symptoms for cancer, for example, is commonly more than 20 years and may vary by several years for individuals, depending at what stage of the disease they are diagnosed.

A disease is *endemic* when it is constantly present in an area. It may occur at low levels, occasionally popping up here or there, as typhoid does in the United States (*hypoendemic*); or it may occur with intense transmission, as malaria does in parts of Africa (*hyperendemic*). Sometimes a hypoendemic disease can flare into rapid spread, perhaps in response to the dislocations of warfare or malnutrition from crop failure. Sometimes diseases that are not endemic are introduced and spread rapidly. A disease is said to be *epidemic* when it occurs at levels clearly beyond normal expectation and is derived from a common or propagated source. Epidemic disease may include an outbreak spreading through a population or widespread degenerative diseases, such as lung cancer.

The terms *incidence* and *prevalence* are often confused or used loosely and wrongly. *Incidence* refers to the number of cases of a disease diagnosed or reported for a population during a defined period of time, most commonly a year. It refers to new cases. *Prevalence* refers to the number of people in a population sick with a disease at a particular time, regardless of when the illness began. Thus the incidence of tuberculosis in Texas in 2000 refers to the number of new cases diagnosed there that year, whereas the prevalence of tuberculosis in Texas in 2000 includes the total number of Texans suffering from the disease that year, no matter what year they acquired it. The incidence and prevalence rates of chronic diseases, especially, can be very different.

THE CHALLENGE OF MEDICAL GEOGRAPHY

“Geography is destiny.” So declared the researchers who analyzed health care and the outcome of treatment by composing *The Dartmouth Atlas of Health Care* (Dartmouth Medical School, 1998) by health service areas across the United States. “Health and our geographies” are inextricably linked, declared Gattrell (2002, p. 3). “Where you live affects your risk of disease, access to basic resources of food, clean water, and housing, and concepts of health.” In the words of Hippocrates, “airs, waters, and places” are still the things from which to start the study of everything else.

It has been said in the context of studies of world regions as syntheses and not just collections of events and facts that if geography did not exist, it would be necessary to invent it. In the face of global transformations of environment, population, economics/politics/society/religion, biotechnology, communication, and so on, scientists concerned with disease globalization and emergence and persistence have, in a way, tried to invent geography. Aron and Patz (2002) target public health graduate students with a global perspective on ecosystem change, including models of population movements, water cycles, and epidemiological design with GIS. Guest (2005) presents globalization, health, and the environment in chapters on poverty, medicinal plants, population dynamics, health ecology of the Nunavut, and urban land use in Baguio City. The best presentation is McMichael’s (2001) look at past patterns and future uncertainties in human frontiers, environments, and disease. It states as its central theme that looking at humankind’s long evolutionary and historical experience shows how the social and natural environments affect patterns of disease and survival. Appreciating the ecological perspective on human population health when such large-scale stresses are appearing in our world is a prerequisite, it says, to achieving a sustainable future. (An obvious successor to Rene Dubos, Tony McMichael is Professor of Epidemiology at the London School of Hygiene and Tropical Medicine, which has given the world so much work in disease ecology.) Admirable and impressive as these works are, they are all composites—patched together from looks at different disciplines, different problems, and seemingly random places. One chapter will recite the hydrological cycle; another will explain cultural expressions of diet or medicine. All the subjects and methodologies they are trying to present are contained within the discipline of geography, which has an historic mandate to integrate them.

As a Request for Action (1999–2003) on “The Ecology of Infectious Disease” from the National Science Foundation and the National Institutes of Health put it, our

capacity to analyze and model biocomplexity and ecological dynamics, and to evaluate spatial and temporal aspects of environmental change has become increasingly sophisticated. However, few of these advances in ecological science have yet contributed to biomedical research or to public health. Similarly, we have improved our ability to define the molecular identity and dynamics of pathogens or infectious agents and their vectors, and our understanding of the defense systems of their hosts. We also understand the importance of genetic systems and evolutionary dynamics of infectious

diseases. These improvements have contributed significantly to our understanding of epidemiology and transmission patterns of diseases. *However, the relationship of these factors to population dynamics of disease reservoirs or the biotic and structural complexity of ecological systems in which transmission occurs remains a poorly understood area. There has also been insufficient attention given to integrating the ecological and evolutionary dynamics of these systems.* (emphasis added)

Too often, geographers have been scrapping over bones and proclaiming themselves loudly to be top dogs in their little pen, while the health and disease of the human population stands at a bifurcation in state of being. Medical geography was at its founding inclusive, open, aiming at synthesis and synergy. This text tries to survey the field and provide an introduction to the geographic concepts and methodologies, the cultural and environmental influences, and the ecological and socio-economic structures that explain the patterns of health and disease. The integrative perspective of geography and all the questions and methodologies of geography's various traditions are needed. In the words of John Hunter (1974, pp. 3–4),

The application of geographical concepts and techniques to health-related problems places medical geography, so defined, in the very heart or mainstream of the discipline of geography. I would suggest that there is no professional geographer, whatever his or her systematic bent or regional interest, who cannot effectively apply a measure of his or her particular skills or regional insights towards the understanding, or at least partial understanding, of a health problem. This is the essential challenge of medical geography.

REFERENCES

- Aron, J. L., & Patz, J. A. (Eds.). (2001). *Ecosystem change and public health: A global perspective*. Baltimore: Johns Hopkins University Press.
- Audy, J. R. (1971). Measurement and diagnosis of health. In P. Shepard & D. McKinley (Eds.), *Environmental: Essays on the planet as a home* (pp. 140–162). Boston: Houghton Mifflin.
- Barrett, F. A. (1980). Medical geography as a foster child. In M. S. Meade (Ed.), *Conceptual and methodological issues in medical geography* (pp. 1–15). Chapel Hill: University of North Carolina, Department of Geography.
- Barrett, F. A. (2000). *Disease and geography: The history of an idea* (Geographic Monograph No. 23). Toronto: Atkinson College, York University.
- Brookfield, H. (2005). Political ecology: A critical introduction. *Geographical Review*, 95(2), 296–300.
- Clark, J. (1998). A social ecology. In M. Zimmerman et al., *Environmental philosophy* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Cliff, A. D., & Haggett, P. (1988). *Atlas of disease distributions: Analytic approaches to epidemiological data*. Oxford, UK: Blackwell.
- Cliff, A. D., & Haggett, P. (1989). Spatial aspects of epidemic control. *Progress in Human Geography*, 13, 315–337.
- Cliff, A. D., Haggett, P., & Smallman-Raynor, M. (2004). *World atlas of epidemic diseases*. London: Arnold.
- Cromley, E. K., & McLafferty, S. L. (2002). *GIS and public health*. New York: Guilford Press.
- Dartmouth Medical School, Center for the Evaluative Clinical Sciences. (1998). *The Dartmouth atlas of health care 1998*. Chicago: American Hospital Association.

- Dear, M. J., & Wolch, J. R. (1987). *Landscapes of despair*. Princeton, NJ: Princeton University Press.
- Drake, D. (1854). *A systematic treatise, historical, etiological and practical, on the principal diseases of the interior valley of North America*. New York: Lenox Hill.
- Dubos, R. (1987). *Mirage of health: Utopias, progress, and biological change*. New Brunswick, NJ: Rutgers University Press. (Original work published 1959)
- Dubos, R. (1965). *Man adapting*. New Haven, CT: Yale University Press.
- Earls, F., & Carlson, M. (2001). The social ecology of child health and well-being. *Annual Review of Public Health*, 22, 143–166.
- Ewald, P. W. (2002). *Plague time: The new germ theory of disease*. New York: Free Press.
- Eyles, J., & Woods, K. (1983). *The social geography of medicine and health*. New York: St. Martin's Press.
- Farmer, P. (1992). *AIDS and accusation: Haiti and the geography of blame*. Berkeley: University of California Press.
- Farmer, P. (1997). Social scientists and the “new tuberculosis.” *Social Science and Medicine*, 44(3), 347–358.
- Farmer, P. (1999). *Infections and inequalities: The modern plagues*. Berkeley: University of California Press.
- Farmer, P. (2003). *Pathologies of power: Health, human rights, and the new war on the poor*. Berkeley: University of California Press.
- Finke, L. L. (1792–1795). *Versuch einer allgemeinen medicinisch–praktischen geographie* (3 vols.). Leipzig: Weidmannische Buchhandlung.
- Fuchs, C. F. (1853). *Medizinische geographie*. Berlin: Duncker.
- Gattrell, A. C. (2002). *Geographies of health*. Oxford, UK: Blackwell.
- Geddes, A. (1978). Report to the Commission on Medical Geography. *Social Science and Medicine*, 12D, 227–237.
- Gesler, W. M. (2003). *Healing places*. Lanham, MD: Rowman & Littlefield.
- Gesler, W. M. (2005). Therapeutic landscapes: An evolving theme. *Health and Place*, 11, 295–297.
- Godlund, S. (1961). *Population, regional hospitals, transportation facilities, and regions: Planning the location of regional hospitals in Sweden* (Lund Studies in Geography Series B: Human Geography No. 21). Lund, Sweden: Department of Geography, Royal University of Lund.
- Guest, G. (Ed.). (2005). *Globalization, health, and the environment: An integrated perspective*. Lanham, MD: Rowman & Littlefield.
- Hawley, A. H. (1950). *Human ecology: A theory of community structure*. New York: Ronald Press.
- Hippocrates. (1886). *The genuine works of Hippocrates* (F. Adams, Trans.). New York: Wood.
- Hirsch, A. (1883–1886). *Handbook of geographical and historical pathology* (3 vols., C. Creighton, Trans.). London: New Sydenham Society.
- Hunter, J. M. (1966). River blindness in Nangodi, Northern Nigeria: A hypothesis of cyclical advance and retreat. *Geographical Review*, 56, 398–416.
- Hunter, J. M. (1974). The challenge of medical geography. In J. M. Hunter (Ed.), *The geography of health and disease* (pp. 1–31). Chapel Hill: University of North Carolina, Department of Geography.
- Inhorn, M. C., & Brown, P. J. (Eds.). (1997). *The anthropology of infectious disease: International health perspectives*. Amsterdam: Gordon & Breach.
- Janson, C.-G. (1980). Factorial social ecology: An attempt at summary and evaluation. *Annual Review of Sociology*, 6, 433–456.
- Jones, K., & Moon, G. (1987). *Health, disease and society*. London: Routledge & Kegan Paul.
- Jusatz, H. J. (Ed.). (1968–1984). *Medizinische landerkunde* (Vols. 1–6, Geomedical Monograph Series). Berlin: Springer Verlag.
- Kearns, R., & Moon, G. (2002). From medical to health geography: Novelty, place and theory after a decade of change. *Progress in Human Geography*, 25(5), 605–625.
- Kearns, R. A., & Gesler, W. M. (1998). *Putting health into place: Landscape, identity, and well-being*. Syracuse, NY: Syracuse University Press.
- Kellert, S. H. (1993). *In the wake of chaos: Unpredictable order in dynamical systems*. Chicago: University of Chicago Press.

- Koch, T. (2005). *Cartographies of disease: Maps, mapping, and medicine*. Redlands, CA: ESRI Press.
- May, J. M. (1950). Medical geography: Its methods and objectives. *Geographical Review*, 40, 9–41.
- May, J. M. (1950–1954). *Atlas of the distribution of diseases*. New York: American Geographical Society.
- May, J. M. (1954). Medical geography. In P. E. James & C. F. Jones (Eds.), *American geography: Inventory and prospect*. Syracuse, NY: Syracuse University Press.
- May, J. M. (1958). *The ecology of human disease*. New York: MD Publications.
- May, J. M. (Ed.). (1961). *Studies in disease ecology*. New York: Hafner.
- Mayer, J. D. (1996). The political ecology of disease as one new focus for medical geography. *Progress in Human Geography*, 20, 441–456.
- McGlashan, N. D., & Blunden, J. R. (1983). *Geographical aspects of health: Essays in honour of Andrew Learmonth*. New York: Academic Press.
- McMichael, A. J. (2001). *Human frontiers, environments and disease: Past patterns, uncertain futures*. Cambridge, UK: Cambridge University Press.
- Moran, E. F. (Ed.). (1990). *The ecosystem approach in anthropology*. Ann Arbor: University of Michigan Press.
- Morrill, R. L., & Earickson, R. (1968). Variation in the character and use of Chicago area hospitals. *Health Services Research*, 3, 224–238.
- National Research Council. (1997). *Rediscovering geography*. Washington, DC: National Academy Press.
- Odum, E. P. (1959). *Fundamentals of ecology*. Philadelphia: Saunders.
- Pattison, W. D. (1964). The four traditions of geography. *Journal of Geography*, 63, 211–216.
- Prigogine, I., & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature*. New York: Bantam Books.
- Pyle, G. F. (1969). The diffusion of cholera in the United States in the nineteenth century. *Geographical Analysis*, 1, 59–75.
- Pyle, G. F. (1971). *Heart disease, cancer, and stroke in Chicago: A geographical analysis with facilities, plans for 1980* (Research Paper No. 134). Chicago: University of Chicago, Department of Geography.
- Robbins, P. (2004). *Political ecology: A critical introduction*. Malden, MA: Blackwell.
- Rodenwaldt, E., & Jusatz, H. J. (Eds.). (1952–1961). *Welt-Seuchen atlas* (Vols. 1–3). Hamburg, Germany: Falk.
- Rothman, K. J. (1986). *Modern epidemiology*. Boston: Little, Brown.
- Shannon, G. W., & Dever, G. E. A. (1974). *Health care delivery: Spatial perspectives*. New York: McGraw-Hill.
- Sorre, M. (1947). *Les fondements de la géographie humaine. Vol I: Les fondements biologiques: Essai d'une écologie de l'homme* (2nd ed.). Paris: Librairie Armand Colin.
- Stokols, D. (1992). Establishing and maintaining healthy environments: Toward a social ecology of health promotion. *American Psychologist*, 47, 6–22.
- Stokols, D. (1996). Translating social ecological theory into guidelines for community health promotion. *American Journal of Health Promotion*, 10(4), 282–293.
- Thomas, R. W. (1992). *Geomedical systems: Intervention and control*. London: Routledge.
- Williams, A. (Ed.). (1999). *Therapeutic landscapes: The dynamic between wellness and place*. Lanham, MD: University Press of America.
- Wolch, J. R., & Dear, M. J. (1993). *Malign neglect*. San Francisco: Jossey-Bass.

FURTHER READING

- Banks, A. L. (1959). The study of the geography of disease. *Geographical Journal*, 125, 199–216.
- Brown, T., & Moon, G. (2004). From Siam to New York: Jacques May and the “foundation” of medical geography. *Journal of Historical Geography*, 30(4), 747–763.

- Cassell, J. (1964). Social science theory as a source of hypotheses in epidemiological research. *American Journal of Public Health*, 54, 1482–1487.
- Gesler, W. M. (1991) *The cultural geography of health care*. Pittsburgh, PA: University of Pittsburgh Press.
- Hawley, A. H. (1950). Human ecology. In *International encyclopedia of the social sciences* (Vol. 4, pp. 328–336). New York: MacMillan.
- Howe, G. M. (1972). *Man, environment, and disease in Britain*. New York: Barnes & Noble.
- Joseph, A. E., & Phillips, D. R. (1984). *Accessibility and utilization: Geographical perspectives on health care delivery*. New York: Harper & Row.
- Learmonth, A. T. A. (1975). *Patterns of disease and hunger*. North Pomfret, VT: David & Charles.
- Learmonth, A. T. A. (1988). *Disease ecology: An introduction*. Oxford, UK: Blackwell.
- McGlashan, N. D. (1972). *Medical geography: Techniques and field studies*. London: Methuen.
- Meade, M. S. (1986). Geographic analysis of disease and care. *Annual Review of Public Health*, 7, 313–335.
- Paul, B. K. (1985). Approaches to medical geography: An historical perspective. *Social Science and Medicine*, 20, 399–407.
- Phillips, D. R., & Verhasselt, Y. (Eds.). (1994). *Health and development*. London: Routledge.
- Pyle, G. F. (1979). *Applied medical geography*. New York: Wiley.
- Shoshin, A. A. (1962). *Principles and methods of medical geography*. Moscow: Academy of Sciences.
- Stamp, L. D. (1964). *The geography of life and death*. Ithaca, NY: Cornell University Press.

CHAPTER 2

The Human Ecology of Disease*

The *human ecology of disease* is concerned with the ways human behavior, in its cultural and socioeconomic contexts, interacts with environmental conditions to produce or prevent disease among susceptible people. This constitutes the *etiology*, or causal evolution, of health and disease. Population genetics, physiology, immunological status, and nutritional status are important to disease processes and must be understood as prerequisites to sound research into these processes. Geography is also important, as its roots are firmly anchored in the study of cultural and environmental interactions.

Geographers have traditionally studied the creation of landscape; the mobility and composition of population; the determinants of economic activity and its location; and the diffusion of things, ideas, and technology. All these are of consequence to medical geography. The landscape is composed of insects, medicinal herbs, and hospitals, as well as topography, vegetation, animals, water sources, house types, and clothing. Mobility is important in exposure to and transmission of disease

*To avoid confusion, let us state succinctly the difference between *human ecology* and a term we have used earlier, *cultural ecology*. *Human ecology* is a broad term used in anthropology and epidemiology, as well as in geography, to denote the patterns of human interaction with the physical environment, including not only behavior but genetic adaptation and physiological reaction to environmental stimuli (such as air pressure or trace elements in water) as well. *Cultural ecology* is more specific and refers to behaviors and belief systems within a particular culture, such as those regarding diet, house construction, or hygiene. It is a frequent source of confusion that a great urban sociologist, Amos Hawley, used the term *human ecology* and its holistic view to build on the theoretical traditions of Robert Parke and E. W. Burgess at the University of Chicago in the 1920s on the structure and development of cities. His influential concepts and writing at the University of North Carolina's Institute for Research in the Social Sciences led directly to the development of a separate tradition of social human ecology in the social sciences—one largely without a biophysical environment. The usage of *ecological* to mean multivariate studies of complex systems stems from the sociological tradition (see Chapter 1).

(discussed in Chapter 5). Elements of population composition include not only age structure and ethnicity, but also immunological and nutritional status and genetic susceptibility.

The main purpose of this chapter is to establish a conceptual framework for understanding why human disease and health vary over the surface of the earth. *Health* is defined in terms of adaptability and is related to complex systems of interaction among habitat (environment), population, and cultural behavior. These three dimensions form a triangular model of human ecology and underlie disease etiology, consequences, and prevention. Each dimension is considered in turn, and is then applied to childhood asthma as an example. Chains of transmission and the particular relevance of water to these are described. A concrete example of disease ecology, cholera, and a complex field of study, the ecology of nutrition, then illustrate the functioning of the conceptual model.

HEALTH

Audy's Definition of Health

J. Ralph Audy (1971) has defined *health* as a "continuing property" that can be measured by an "individual's ability to rally from a wide range and considerable amplitude of insults, the insults being chemical, physical, infectious, psychological, and social" (p. 140). One might prefer the term *stimuli* or *hazards* to *insults*. Such stimuli may be either negative or positive; the crucial thing is that the individual must respond to them.

Infectious insults consist of the *pathogens*—agents that cause disease. Every person is infected at all times with many billions of viruses, bacteria, and protozoa that cause no harm, such as intestinal bacteria. Changes in health status can cause a normally benign relationship to alter and become pathogenic. We also constantly receive nonorganic physical insults, such as from electromagnetic radiation. The trauma of tissue damage and broken bones can result from falls and violence. We live in a chemical soup. Our bodies are chemical systems, quite literally composed of what we eat. Petroleum derivatives and nicotine are now part of our chemistry. The absence of an essential vitamin or an excess in a basic food component, such as cholesterol, can also be a chemical insult. Mental and social insults further influence physiological functioning.

Examples of Insults or Stimuli

Chemical

Carbon monoxide

Drugs

Benzene

Formaldehyde

Calcium deficiency

Oxygen deprivation

Physical

Trauma

Radiation

Light

Noise

Electricity

Air pressure

Infectious

Prions

Viruses

Rickettsiae

Bacteria

Protozoa

Helminths

Psychosocial

Danger

Crowds

Isolation

Anxiety

Community

Love

It is possible to map, at a variety of scales, every kind of insult. The areas of a town could be mapped based on noise, people's fear of walking down the street at night, air pollution, visual blight or beauty, mosquito density, or alcohol consumption. Such maps could be overlaid to show regions of health hazards. These regions of insults form the environments to which individuals are exposed at the micro scale as they move around.

At the micro scale are self-specific environments. Everyone is wrapped in an envelope of heat, humidity, bacteria, fungi, and mites, and may also host lice and fleas. The driver of a car encounters a set of insults that differs by sections of the road and by the number and types of other vehicles around. The infectious and other insults encountered on a bus are quite different from those in a car. Within buildings, one is insulted by microwaves from the walls, magnetism from electricity, light, changes in humidity, infections from other people, and various psychosocial challenges (e.g., from books, television, and conversation). The exact nature and range of insults to which an individual is exposed during a day are unique. Behavioral roles associated with age, sex, class, and occupation create some groupings of insults, however, and geographical location delimits other groupings. These differences in exposure to various health hazards can be modeled.

The idea that health is a "continuing property" and not a characteristic that is either present or absent involves recognition that health exists at various levels. The only absence of health is death. Health can exist at a threshold, marginally, or it can exist amply with great reserves. Audy points out that an insult can have a "training" impact. That is, after the body has successfully rallied from the insult, the body is better able to cope with future insults of that kind. One's first public talk, first date, or first exam in graduate school is more difficult to cope with than the 20th. While a person is reacting to a stimulus, however, the level of health is decreased, and that person becomes less able to cope with another insult.

Audy has suggested that the way insults affect the level of health can most easily be diagrammed for immunological health. In the example illustrated in Figure 2.1, two individuals are conceived, are born, and experience infectious insults. The first becomes infected with a cold virus, and while she is coughing, sneezing, and slightly feverish, her health level declines a little. Soon, however, she is immune to that cold virus, and her health rebounds to a higher level. In this way she proceeds through a succession of infectious episodes. Through time, her level of health continues to increase until her early 20s, and then gradually declines over the next several decades. The second individual also survives the massive insults that attend birth, but poor maternal nutrition has given her a lower birthweight and

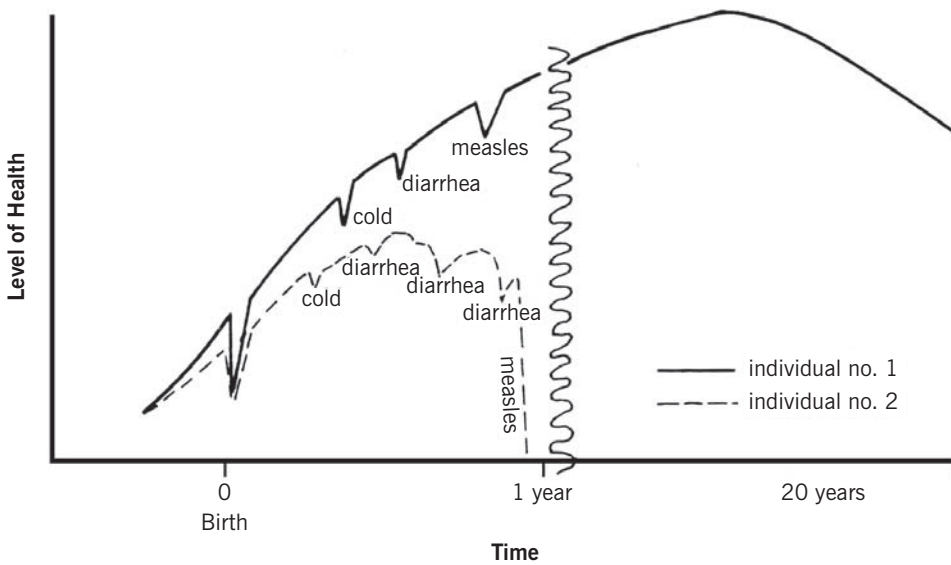


FIGURE 2.1. Health level and insult stress. The discontinuous solid line indicates health level over a lifetime, with peak of health at about age 20. The part before the wavy vertical line indicates levels of health for two individuals within their first year.

level of health. Soon after she rebounds from the cold virus with increased health, she is infected with bacteria that give her diarrhea. She is removed from the food supplements believed to be the cause and even from water, in an effort to stop the diarrhea. Her health level increases slightly as she masters the bacteria, but the episode has precipitated malnutrition because her diet had been marginal even before. With health lowered by malnutrition, before she can rally and restore herself, she is again assaulted by enteric (intestinal) bacteria. Now further malnourished and dehydrated, her health level is greatly decreased. As she struggles to rally and grow, she becomes infected with the measles virus. Depleted of health reserves, she cannot rally, and she dies.

In one study about nutrition and diarrhea, for example, the graph of weight gain by babies and their bouts of infection and illness perfectly reproduced Audy's graph of level of health and stress from insults (Rhodes, 1986). Weighing babies in a sling-scale every month is a method commonly used in well-baby clinics around the world to find early indication that a baby is failing to thrive. Healthy babies usually gain weight and grow in a regular manner, within population-specific norms well known and graphed. Against this curve of normal gain, the ups and downs of the weight of babies over their first 3 years of life were tracked and graphed along with demarcation of each bout of diarrhea, fever, measles, bronchitis, and other illnesses. On the chart of weight curves, one infant with three bouts of diarrhea, a high fever, and the measles at 18 months had reached only the normal weight of an 8-month-old; at the end of 3 years, after further attacks of bronchitis and fevers and repeated diarrhea, it had reached only the weight of a 9-month-old.

The scenario of multiple insults piling up and cumulatively lowering the health level even below the vital threshold occurs frequently in countries with high infant mortality. Mental and physical loads can also exact a long-term, cumulative penalty on individuals (see the discussion of the poverty syndrome in Chapter 8).

The stress of life events such as marriage and divorce, promotion and being fired, moving, losing a spouse, or having a baby can be scored to predict the likelihood of illness events. You may have seen such “check yourself” lists in popular magazines. It has also been noticed that employee absences due to sickness tend to cluster in time. Both these examples of the timing of illness events can be addressed in the terms of the framework that we have developed. Insults require adaptation. Health is lowered during that process, making a person temporarily less able to adapt to the next insult. As final examination time approaches, for example, many students get sore throats as the accumulated stresses of little sleep, poor diet, anxiety, and other insults lower their levels of health until throat bacteria that have been well controlled suddenly cause clinical illness.

THE TRIANGLE OF HUMAN ECOLOGY

Habitat, population, and behavior form the vertices of a triangle that encloses the state of human health (Figure 2.2). *Habitat* is that part of the environment within which people live—that which directly affects them. Houses and workplaces, settlement patterns, naturally occurring biotic and physical phenomena, health care services, transportation systems, schools, and governments are parts of the habitat thus broadly conceived. The following chapters develop other aspects of the habitat; in this chapter the discussion is limited to the constructed part of the habitat, where humans live and work.

Population is concerned with humans as biological organisms—as the potential hosts of disease. The ability of a population to cope with insults of all kinds depends on its genetic susceptibility or resistance, its nutritional status, its immunological status, and its immediate physiological status with regard to time of day or year. The effects of age, gender, genetics, and other population components are pervasive but largely implicit in the remainder of this book.

Behavior is the observable aspect of culture. It springs from cultural precepts, economic constraints, social norms, and individual psychology. It includes mobility, roles, cultural practices, and technological interventions. The triangular ecological model differs from sociological models in its separate consideration of behavior and population. Education, for example, is an element of behavior rather than population status. Education involves behavioral exposures to an opportunity in the habitat; this experience can influence behavior in a way that improves health status by reducing harmful exposures, increasing protective buffering, and inducing alteration of the habitat itself through technology. The social theories of *structuration*, or political/economic/social constructs and perceptions, are concerned with this leg of the triangle.

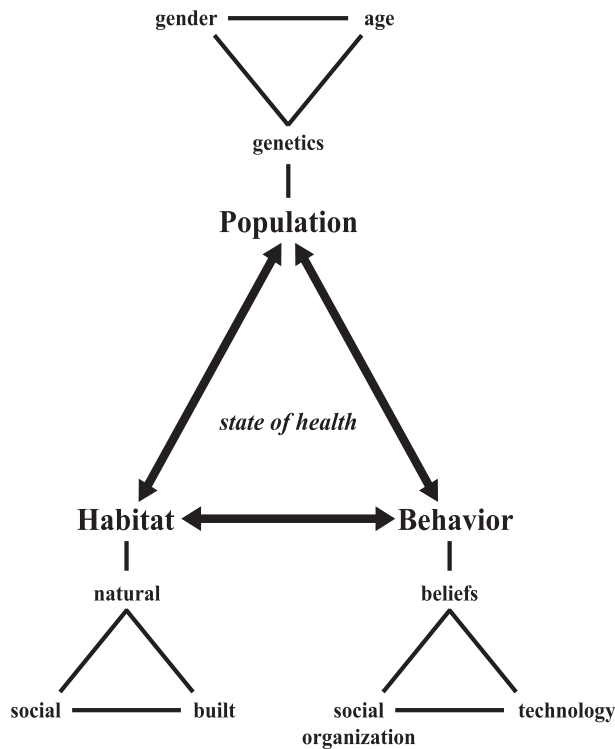


FIGURE 2.2. The triangle of human ecology.

Through their behavior, people create habitat conditions, expose themselves to or protect themselves from habitat conditions, and move elements of the habitat from place to place. The habitat presents opportunities and hazards to the population genetics, nutrition, and immunology. The status of the population affects the health outcome from the habitat stimuli and the energy and collective vigor needed to alter behavior and habitat.

It is possible to elaborate on this basic triangle in many ways. Subsystems can be created for economics, politics, or religion. The effects of global warming (itself affected by human behavior) on natural habitat and behavioral feedbacks can be analyzed. Motivation systems for behavioral alteration can be analyzed. The effects of changes in major cultural paradigms, such as the purpose and role of humans in the environment, can be isolated and spotlighted as insets. This triangle of human ecology is used to discuss women's health in Chapter 8.

Habitat

One leg of the ecological triangle is habitat. It contains sunshine, distance, and many other elements of the physical and biotic environment. We first consider the environment built by humans.

The Built Environment

Asleep or awake, humans spend most of their lives inside their houses or other buildings. Dubos (1965) pointed out that evolutionary stimuli now come more from the environment we have constructed than from nature.

What stimuli do you receive in your house? Is the house heated and cooled, so that the humidity is also affected? Is it well ventilated? Do any insects live in the woodwork or basement, or do you prefer the vapors of insect nerve poison? Does your dog or cat sleep in the house? Are there windows to let in light? Are they screened, or do you usually keep them closed for air conditioning? Are there dark corners and rooms where roaches, spiders, or perhaps mice live? Is the concentration of dandruff, hair, dust, and allergenic materials higher inside or outside your house? Is there perhaps formaldehyde in the insulation or carpet backing, or lead solder in the pipe joints, or asbestos in the shingles? Do you have radiation sources, such as a television or a microwave oven? Or do you have radiation only from electric wiring in the walls? Could your stone foundation emit radon concentrated in the basement? What is the noise level? Do you feel crowded or isolated there?

The types of houses, the presence of domestic animals and their parasites, and the kinds of pens or outbuildings are all of consequence to health. Consider the house types presented in Plate 3, and add your own examples. It matters to insect ecology (and hence to disease transmission) whether roofs are made out of thatch, in which insects nest, or of corrugated iron, which excludes them. It matters what barriers there are to entry of mosquitoes. It matters for the survival and contagion of germs and vectors whether architecture is oriented toward private, shaded, inner courtyards or is open and almost continuous with the outdoors. It matters whether kitchens are inside or outside the dwellings (because of smoke), and whether inside there are piped water and flush toilets. In very developed lands, "sick buildings" can result from sealing the windows in the interest of energy efficiency and thereby trapping fumes from cleaning fluids and paint, plastic emissions, and other noxious gases inside. Plate 4 presents several examples of wells, and Plate 5 presents examples of toilets and other sanitation facilities. Water-borne contagious human diseases have been major causes of death at least since the time of the first cities. What barriers protect water from fecal contamination? What means of access to protected water do people have? As you consider these pictures, think about changes in these conditions from wet to dry seasons or hot to cold. What facilitates disease transmission?

Details matter in the ecology of disease. A certain kind of chimney can cause a room to be smoky. This may repel mosquitoes, but it may chronically irritate eyes and eventually cause blindness. Flooring has one common purpose, but many effects. Floorboards can be spaced so that food and dirt fall through to be scavenged by pigs and chickens below; or they can be placed tightly together or solidly covered, so that one needs to learn to use a dustpan for sanitation; or they can be carpeted, so that fleas can spend their entire life cycle in the living room. Houses can be built out of cold, damp stone and be full of drafts. Alternatively, they can be constructed from

mud and straw and provide good nesting cracks for insects. Humans create much of their disease environment.

Consider the following description by Amato (2000) of the living conditions of poor European peasants (i.e., most of the population) from medieval times until the early 20th century. After explaining how peasants lived “mired in muck” in one-room houses (perhaps with a second room for storage/stable), with dirt floors, no chimneys, and dark, damp, soot-filled interiors, he quotes (p. 26) the Italian historian Camporesi’s assertion that peasants were

dirty, almost always barefooted, legs ulcerated, varicose and scarred, badly protected by meager and monotonous diets, living in humid and badly ventilated hovels, in continuous, promiscuous contact with pigs and goats, obstinate in their beliefs, with dung heaps beneath their windows, their clothes coarse, inadequate and rarely washed, parasites spread everywhere—on their skin, in their hair and in their beds—their crockery scarce or nonexistent, often attacked by boils, herpes, eczema, scabies, pustules, food poisoning from the flesh of diseased animals, malignant fevers, pneumonia, epidemic flus, malarial fevers . . . lethal diarrhea (not to mention the great epidemics, the diseases of vitamin deficiency like scurvy and pellagra, the convulsive fits, so frequent in the past, epilepsy, suicidal manias and endemic cretinism).

Villagers carried around with them a whole fauna of fleas and lice. Not only did they scratch themselves, but friends and relations from all levels in the social scale deloused one another.

The thumb, Amato notes, “was called the louse-killer (*tue-poux*).”

Changes in the built environment can result in profound alterations of disease conditions. We do not know exactly why some diseases have disappeared from Europe and others have increased during the last few centuries. Leprosy, for example, used to be common in Europe but no longer is endemic. Dubos (1965) argued that changes in the built environment, such as cheaply produced window glass and architectural principles that allowed construction of multiple chimneys, flooded even the houses of the poor with light and warmth, drastically changing the habitat for disease agents. In the cities of early industrial Europe, and in New York City early in the 20th century as it received a million immigrants a year from peasant Europe, the construction of dark, unventilated, and crowded tenements provided an ideal habitat for tuberculosis bacteria.

Settlement patterns—the ways people are clustered and distributed on the land—also influence health conditions. On a micro scale, geographers look within settlements at the spatial arrangement of residences and land uses. Usually three general settlement forms are distinguished: *nuclear*, *dispersed*, and *linear*. At a macro scale, geographers look at how the settlements are distributed with regard to each other. A settlement system consists of various sizes of settlements (from dispersed farmhouses to large cities), together with the distances, directions, and connections among them that form the structure of a functional region of trade, ideas, and other interactions. The role played by this hierarchical system in disease diffusion is discussed in Chapter 10. It is also important for the accessibility and specialization of health care.

The most common settlement form is *nuclear*. Most rural people in the world live in houses clustered in a village from which they walk out to the surrounding agricultural fields, with forest and grassland lying beyond. The settlement land use buffers most households from any insect-transmitted diseases from woods and fields. The nuclear form, however, facilitates the fecal contamination of water sources and the spread of directly contagious diseases. Houses in a *dispersed* settlement form are located on the farmland of their owners, and neither air nor water provides much focus of contagion for the scattered population. Each household, however, is exposed to vectored diseases originating in the natural surroundings. A *linear* settlement, in which houses are lined up along both sides of a river, canal, or road, has an intermediary position and often is characterized by the worst conditions of the other two settlement forms. People are only partially buffered from insect-transmitted diseases because the rear of the dwelling is exposed; yet the clustering of houses provides a focus for contagion, especially for those households downstream from other dwellings. In the contemporary U.S. context, the inner city is like a nuclear settlement, exposed to contagion and concentrated pollution. Suburbs, fragmenting the landscapes into which they disperse, reap the hazards of increased tick vectors even as their reduced density retards contagion. Housing sprawled along the highways suffers from the air pollution and vehicular hazards of the city, as well as the mosquito and tick pests of the countryside.

Health services are an integral part of the human habitat. Whether a disease is diagnosed and how it is treated depend partly on whether facilities for urine analysis, X-rays, brain scans, and various blood tests are available to the physician. Some health facilities lack electricity and in many ways constitute a very different health habitat from that presented by a university research hospital. Factors as diverse as international economics and roads washed out by monsoon rains can affect the availability of antibiotics or blood for transfusion. The availability and accessibility of health facilities and health personnel constitute a critical part of the health habitat.

The Social Environment

The social environment consists of the groups, relations, and societies within which people live. Audy's (1971) conception of health at different levels and depths of reserves, and Dubos's (1965) and McKeown's (1988) observations—that (1) all people exposed to an infection or other risk did not get sick, and (2) once they were sick, people followed varied courses of illness and outcome—had come to seem obvious by the mid-1980s but were still mysterious. These authors had pointed out that the major declines in mortality resulting from many infectious diseases had occurred prior to effective therapy. Genetic predisposition, other population attributes, and even such things as public health improvements in sanitation and changes in year-round availability of good nutrition seemed to explain only part. The groundbreaking studies by Marmot and colleagues (Marmot, 1986; Marmot, Kogevinas, & Elston, 1987), known as the Whitehall studies, followed 10,000 British civil servants over two decades. They established the importance to health of relative status and not simply economic deprivation.

In these studies, everyone had abundant food, good shelter, excellent water and sanitation, accessible good-quality health care, public health information, and the general benefits of living in an economically developed country. Yet the clerical/manager civil servants had three and a half times the mortality rates of senior administrator civil servants. The effects of status seemed to work especially through the response to stress. Everyone's blood pressure rose during the work week, for example, but at home over the weekend the blood pressure of senior administrators fell to healthy levels, whereas that of lower-level civil servants declined only slightly. They seemed to remain continually stressed. Social theorists were quick to realize the implications for race discrimination, gender stereotyping, access to higher education, and social status constructs of many kinds. Addressing the question of how apparently equivalent risks to health became manifested so differently, Robert Evans's (e.g., Evans, Barer, & Marmor, 1994) influential research and writing brought the issue to the forefront of Canadian efforts to promote health as a national priority. Many researchers have confirmed the importance of relative differences in status, and not simply deprivation. The old emphases on "unhealthy lifestyle choices" and economic deprivation are clearly incomplete and inadequate. There are larger structural and context processes involved in determining health.

Even "laboratory" studies at the Duke Primate Center (in North Carolina) and elsewhere have confirmed that moving a high-status primate to a group where its status is low has effects on heart disease, cancer, and the course of infections. But what are the plausible mechanisms? Evans and colleagues (1994) elucidated the internal biological responses—especially processes of the immune system and endocrine production—that link health status to the perception of the external environment (note that this is an interaction of the behavior vertex with the population vertex to cope with the habitat vertex). They pointed out that many cultural variables affect how distress and disease become manifest in a particular cultural context. There are implications for such things as how groups respond to the stress of culture change through coping strategies based on different traditional organization and key values, or how epidemiologists conduct community surveys of mental health.

Population

The *nature* of the population—that is, the characteristics, status, and conditions of individuals as organisms—does much to determine the health consequences of any stimulation. Whether the stimulus is a bacterium, light, drug, sound, or thought, the reaction will differ according to the body's biochemical state. This physiology is in part inborn through the genetic code, but it is also influenced by weather, nutrition, previous experience, age, and so on.

Genetics

Genes set the limits of our possible responses to social, biological, and physical experiences. Once it became possible to "read" the deoxyribonucleic acid (DNA) sequence of bases that encodes the structure and processes of life, the sciences of

genetics and biochemistry expanded explosively. DNA, the sum total of all genetic information, is a long sequence of small molecules termed *nucleotides*. There are four of them, which are denoted by the letters A, C, T, and G. Their specific order and position code the genetic information. The two strands of DNA join in the famous structure of the double helix. There are approximately 6 billion nucleotides in the human genome. The DNA of all human beings on Earth is more than 99.9% the same.

Now that the human genome has been read—all those billions of nucleotides have been spelled out—genetic research for marking, determining, intervening, and treating has become very important and well funded.

Our knowledge of how genetic information is stored, transmitted, and activated at the appropriate time is undergoing almost daily revision. The DNA chain can be broken in some places but apparently not in others. Pieces can be switched, overridden, repeated, or deleted. Some genetic information has persisted from ancient times, with no known function today; apparently at least some of this information is that of ancient viruses. Moreover, some genetic information is activated to govern the production of enzymes, hormones, or other proteins only at certain times in life and then is “turned off.” Cancer may turn some of these genes (e.g., for tissue growth) back on, and shut the inhibiting proteins off. The functioning of the genetic code is much more complex, and the encoded instructions much more varied, than had ever been imagined previously.

Several genes responsible for serious diseases have already been identified: those involved in cystic fibrosis, Huntington’s chorea, familial breast cancer susceptibility, and familial early-onset Parkinson’s disease. Recombinant genetic technology—using, for example, common viruses to “vector” normal genes into cells to replace damaged and dysfunctional ones—is already leading to new therapies. Even diseases influenced by several genes (e.g., some types of breast cancer) are yielding to the new technologies, and there is excitement about treatments for diabetes, schizophrenia, and severe forms of obesity, to name a few of the current targets.

Genetically *caused* disease is uncommon, however. Rather than being caused by dominant lethal genes, most genetic diseases are associated with rare recessive traits that, while terribly important to the individuals and families involved, are not a concern at the population level except under conditions of inbreeding. Hemophilia among the royal families of Europe in the 19th and early 20th centuries is the most famous example of such a disease. Harmful traits in a population that is not inbred tend to be eliminated rather than reproduced. When they are found to occur frequently in a population, one looks for some advantage that they bestow. The gene for sickle cell anemia in Africans was the classic conundrum. The sickle cell trait causes high mortality, but wasn’t rare; in much of sub-Saharan Africa, where malaria was and still is most prevalent (see Plate 6), more than half the population carried the trait. Eventually it was found that though having two genes for cell sickling was deadly, having a single gene for the trait was protective against malaria.

Most diseases that tend to be transmitted in families or that occur more frequently in particular population groups result from genetic *susceptibility* rather

than genetic causation. That is, the disease requires a stimulus or “cause” to occur, but another person who is not genetically susceptible will not respond to the same stimulus with disease. Often the ability of the liver or proteins in the blood to block or detoxify a specific insult is involved. Some types of cancer, probably Parkinson’s disease, and the virulence of several diseases are thought to be related to genetic susceptibility or resistance. Great scientific resources are being expended to find and identify these genes for susceptibility, but (alas, in our views) very little for identifying the stimuli (e.g., agricultural chemicals) to which they are susceptible.

Race and Genetic Variability. Research in biochemistry and genetics has been overwhelmed by the recent findings on the variability and plasticity of human inheritance. The paired, rod-shaped chromosomes in the nucleus of each body cell contain the paired genes (composed of those DNA nucleotides) that control heredity. There are specific points, known as *structural loci*, on the chromosomes for genes governing each characteristic (trait). Genes that occupy the same locus on a specific pair of chromosomes and control the heredity of a particular characteristic, such as blood type, are known as *alleles*. When more than one version of the same trait is common (e.g., blue or brown eyes; Type A, Type B, Type AB, or Type O blood), the population is said to be *polymorphic* for that trait. Humans have long been recognized to be polymorphic for blood type, skin color, hair texture, stature, and other traits that used to be categorized by the concept of *race*. There are, at a conservative estimate, more than 50,000 structural loci. About one-third of these are polymorphic. Each individual has two different alleles for about one-third of these polymorphic structural loci, or about one-tenth of his or her entire genetic inheritance.

One way to illustrate the importance of polymorphism for human health is to consider the *histocompatibility* or *human leukocyte antigen* (HLA) system—in short, the genetic control of the body’s immune system. *Histo-* refers to tissue; HLAs are white blood cell substances that induce production of antibodies. The HLA region of the chromosomes has at least four loci. More than 20 alleles may occur at one of these, 40 at another. In total, more than 80 alleles are involved at these four loci, and the possible number of reproductive combinations exceeds 20 million. It has long been known that people have four different types of blood—A, B, AB, or O—and that the Rh factor is positive or negative. Currently, more than 160 red blood cell antigens are identifiable. Most have been implicated in blood transfusion reactions and presumably are involved in mother–fetus exchanges. More than 100 variants of human hemoglobin are also known. This is the type of genetic variability involved in acceptance or rejection of organ transplants, defense against cancer, and resistance to diseases such as malaria or measles.

The concept of race has become obsolete insofar as it denotes a classification with concrete existence, rather than a convenient categorization for some particular purpose. One can visualize thousands, even millions, of maps of the relative frequencies of genetic traits for antigens and enzymes, hormones, bone structure, skin pigmentation, and other coded instructions for forming human beings. Each mapped surface is continuous over the earth, for almost all of these variants occur

everywhere. The frequency of most genetic variations in a population has a spatial gradient (slope), however. The gradients for types of blood antigens, for example, run east–west in Europe, with Type A and Rh factor decreasing to the east. Skin pigmentation and body shape, however, have a north–south gradient in Europe. If people were “racially” grouped by their fingerprints, Western Europeans, sub-Saharan Africans, and East Asians would be grouped together because they more often have “loops” whereas Mongolians and Australian aborigines have “whorls,” and Khoisan and Central Europeans have “arches.” If people were grouped by body hair or size of buttocks, still other different “races” would be classified.

On the other hand, studies of monozygotic (identical) twins have identified an astonishing amount of genetic propensity. Such twins, who have identical DNA, may be raised together under the same social influences or separately in different families and circumstances. Twin studies allow genetic effects and shared environmental effects to be separated. The range of genetic influences includes such things as color preference and favorite foods, occupational choice, form of exercise, alcohol and tobacco usage, and “personality type” in risk of cardiovascular disease.

The manner in which each genetic distribution compares with environmental, cultural, or disease patterns is of interest. Both physical environmental factors (e.g., solar radiation) and cultural factors (e.g., method of maintaining livelihood, perceptions of beauty, marriage ideals, and migration histories) have affected genetic patterns. Geographers have joined the search for associations between blood type and disease susceptibility and between metabolic differences and aspects of cultural evolution, such as cropping systems. Any effort to categorize all distributions, however, has to involve rather arbitrary criteria and great simplification. The concept of race is still used occasionally as a convenient biological categorization for the relative frequencies of many alleles, in full recognition that group boundaries and the traits included are arbitrary.

When race is used as a grouping of alleles to indicate the overall genetic distance or closeness of populations, it sometimes establishes a useful research framework. The Japanese (historically considered Mongoloid) in Japan have different health problems—kinds of cancer, for example—than do white Europeans and Americans (historically considered Caucasian) also living in industrialized countries at the same latitudes. The disease experience of successive generations of Japanese in the United States has been studied to help untangle genetic (changed by intermarriage over generations), cultural (behavioral changes in diet, etc., over decades), and environmental (immediate changes by migration) factors in cancer etiology. In most scientific usage, however, race has generally been replaced by the categorization of *ethnicity*, for it has been many decades since behavioral, linguistic, or mental characteristics were scientifically associated with the genetic inheritance of physical traits. Anyone conducting or reading about research that involves a racial classification needs always to question closely the nature, purpose, and appropriateness of the characteristic used; for example, is careless and lazy usage neglecting socioeconomic or cultural causal associations? *Race* has, of course, become a social construction—a category experienced or studied as a reality, with real consequences for social marginalization, group perception, and health care.

Genetically Based Differences in Metabolism. There is considerable interest today in investigating the genetic basis of differences in human metabolism (energy and material transformation within cells) as these differences interact with culture and health. Lactose intolerance and alcoholism are briefly discussed here.

Lactose intolerance is a classic geographic puzzle in human ecology. Lactose, the only carbohydrate in milk, is split in digestion by the enzyme lactase into glucose (a sugar) and galactose, which are absorbed into the blood stream as nutrients. Lactase appears in the human fetus in the third trimester of pregnancy, reaches a peak at birth, and falls gradually in childhood. The condition of lactose intolerance, or being unable to digest milk fully because of an inability to produce lactase, is the usual condition of most adult mammals, including human beings. Among some populations, notably those of European origin, the gene for producing lactase usually does not shut off, and so most adults are able to digest dairy products. Scientists and others of European descent used to consider this the normal human condition.

Simoons (1969, 1970) proposed a geographic hypothesis of biological and cultural interaction to explain the distribution of lactose intolerance around the world. He estimated that among Asians, about 90% of the population is lactose-intolerant. The question raised by his research is whether, for example, the Chinese lost their ability to digest lactose because they defined cattle keeping as barbaric and consequently excluded from their diet milk and the meat of herd animals (eating the meat of only "dooryard scavengers" such as chickens and pigs), or whether they developed their cuisine because they were unable to digest milk. Similarly, did Europeans develop a high frequency of lactose tolerance because they became herders of cattle, with mixed farming systems to support their draft animals, or did they take to dairy foods and raise the animals that produced milk because they could digest lactose? In Africa the ethnic groups that herd cattle are generally much more lactose-tolerant than the farming groups that do not, but the genetic pattern is complex and highly variable among villages as a result of invasions, migrations, and intermarriage with Muslim traders (e.g., across the Sahara as Islam spread). Lactose intolerance has important ramifications for worldwide emergency relief (usually involving powdered milk) and protein supplementation, agricultural extension, and development aid of all kinds, as well as vitamin D and calcium supplementation.

Studies involving the metabolism of alcohol have shown that Chinese people, even when acculturated to U.S. drinking habits, tend to metabolize alcohol differently than members of European ethnic groups do. A single glass of wine may produce the dizziness, flushing, and nausea that usually characterize much higher levels of alcohol consumption among Europeans. Some ethnic groups, such as Jews, have low levels of alcoholism; others, such as Russians, have rather high levels. Most such differences have been explained in sociocultural terms, as alcohol plays different roles in different societies. But the question arises: Is the different role of alcohol in European and Chinese cultures a result of or a cause of differences in metabolism of alcohol? As the study of alcoholism progresses, it is being found that upbringing, life stress, familiarity, and social custom are important, but that alcoholism also clusters in families. Even when adopted in infancy, the children of alcoholic fathers are more likely to develop alcoholism, and identical twins reared apart and brought

up differently seem to show similar susceptibility (these studies have involved small numbers, however). Alcohol has been produced for millennia in some cultures to store excess grain and to provide an alternative to polluted water; in other societies it has been little used in these ways. Today alcohol is an especially severe problem in ethnic groups as diverse, yet similar, as the Inupiat in Greenland, Australian aborigines, and Native Americans. As in the case of lactose intolerance, geographic associations of culture, alcohol, and population may help establish patterns and promote understanding of a serious human health problem.

Other Influences on Population Status

Age is a critical factor for health status. Those who study geriatrics (diseases of old age) are just beginning to understand the aging process, but many of the biochemical changes are obvious. Metabolism changes in response to different energy requirements when behavioral roles change with age and growth, physical maturation, and completion of reproduction. Hormone and enzyme production; organ function; and deposition and storage of fats and chemicals in the bones, liver, and blood vessels all change with age. Life experience also accumulates, and the immune system recognizes and copes with a greater variety of infectious agents—and is sometimes more sensitized from long exposure to a variety of allergenic substances. The regulation of some homeostatic systems, such as temperature control, tends to become less efficient with age; other systems gain in efficiency, from practice.

The age structure of a population in large part determines consequences as diverse as the spread of an infectious agent and the severity of the illnesses it causes; the effects of changes in the weather or of an air pollution episode; the expression of carcinogenic agents in clinical disease; and the need for health services. Because age affects so many dimensions of health status, it needs to be accounted for in virtually every study of disease ecology. The relevance of life stages, and techniques for adjusting data for age so that populations can be compared, are described later.

Behavior

Human beings are animals that are cultural beings. Culture creates social organization, structuring relationships of power, status, and control of resources. Culture creates belief systems, values, and perceptions. Culture develops technology. The study of this vertex comprehends most of social science and the humanities. Cultural behavior interacts with the triangle of human ecology in four ways:

1. Humans create many habitat conditions.
2. Behavior exposes individuals and populations to some hazards and protects them from others.
3. People move not only themselves from place to place, but also other elements of disease systems.
4. Behavior affects the health of the population by controlling genetics through

marriage customs, nutritional status through food customs, and immunological status through the technology of vaccination and customs of deliberate childhood exposure.

Each of these aspects of behavior is discussed in this section.

Very little of the earth's surface is unaltered by human activity. Vegetation has been burned, removed, and planted. Water has been withdrawn from below the earth's surface and stored and distributed over hundreds of miles on the surface. The chemical composition of the air and water has been altered. The kinds of buildings people construct, the kinds of industries they work in, the kinds of vehicles and roads they build, and the kinds and scales of the settlement systems they live in all help form the hazards to which they are exposed. This cultural creation of disease is as true for malaria as for lung cancer. For example, in the U.S. Northeast over a century ago, people cleared most of the forest for agriculture and growing cities. They thereby eliminated many of the animals. Land has in recent decades reverted to old fields and new forests, and presently is being fragmented by expanding suburban settlement and abandonment of industrial sites. With the restoration of habitat and now without large predators to control them, the deer population has rebounded to the highest level in two centuries. This has promoted the spread of Lyme disease, vectored by the deer tick, and established it as a major public health concern throughout the region.

Besides creating hazards, cultural practices also function to protect people. Many protective customs have been developed. In the Chinese culture realm, people habitually drink tea rather than unboiled water. The European custom of using a handkerchief for blowing the nose is as protective as the Chinese custom of blowing the nose onto and spitting upon the street is endangering. Wearing shoes provides almost total protection against hookworm. Frequent washing of hands, especially after defecation or before handling food, is effective protection from diseases as diverse as the common cold and hydatidosis (invasion of tissue by tapeworms; see Table 4.6). Some of these protective behaviors may be considered a kind of deep cultural ecology, having started to evolve even before our species. Cooking can be considered as an "arms race" with microbes. Garlic, turmeric, cumin, black pepper, chili pepper, and other spices are proving to be antibacterial, antihelminthic, anti-inflammatory, and even anticarcinogenic! Cooking itself, of course, breaks down cell walls and allows humans, who can't digest cellulose, to absorb the nutrients.

We have learned to construct *buffers* against disease deliberately: Consider the chlorination of water supplies, the use of seat belts, and the prevalence of the water-sealed toilet. There are cultural origins and diffusion paths for protective buffers, and sometimes diffusion and adoption take time. There are also socioeconomic barriers to their occurrence. For example, case mortality from breast cancer is higher among African American women than among European American women, in large part because fewer of the former receive mammograms for early detection. Poorer people around the world are smoking more and consuming more sugar than more educated and wealthy people in the same countries are.

Behavioral roles, varying by age and sex and often by ethnicity or class, largely determine who is exposed to what. There is great cultural variation in social norms. In one culture men are exposed to the infections of the crowded marketplace because women are thought not to understand money, whereas in another culture women are exposed because they are judged superior at bargaining and men are considered too unreliable to handle the family money. Who herds animals in pastures and gets hookworm? Who handles dangerous industrial chemicals and gets cancer? Who washes laundry in the morning when *Mansonia* mosquitoes are biting and gets filariasis (a worm infestation; see Table 4.2)? Who tends the orchard in the evening when *Aedes* mosquitoes are biting and gets dengue fever? Who goes to school and gets influenza? Who drives a vehicle and suffers a concussion? Who fights a war and is maimed? Who moves to the city for wage labor and contracts HIV while separated from home? Who leaves the home village for marriage and faces abuse or depression without social support? Globalization may be affecting many socioeconomic structures and the behaviors they channel, but traditional cultural behavior and beliefs remain influential locally.

Mobility patterns are of critical importance for the diffusion and incidence of disease, and they are discussed at some length in Chapter 5. Medicinal plants, as well as insects that transmit disease, have been spread around the world by humans. When people migrate, they carry customs that are adaptive in their place of origin but are often inappropriate in the new place, such as building a house in Hawaii with a steeply pitched roof originally designed to shed snow. Such “cultural baggage” may include ideas about poisonous or nutritious plants. Formerly protective behavior may even be harmful in the new place. A good example is *geophagy*, the practice of earth eating. In parts of Africa pregnant women eat earth formed into molds and sold in the markets. The earth comes from special sites and is often high in calcium, iron, and other minerals. It may provide critical nutrients missing in the diet. It may be eaten to control diarrhea and eliminate parasites that could affect the fetus, just as the clay kaolin is eaten in Kaopectate. The practice persisted late into the 20th century in the United States among a few people of African descent. An old woman in rural Georgia would keep a jar of her favorite clay from a local stream bed in her pantry, although she did not need its minerals (inappropriate behavior); and her daughter who had migrated to a northern city instead consumed laundry starch as a substitute for earth (harmful behavior). With mobility, custom often becomes dissociated from purpose.

It is easy to understand how behavior can affect the genetics of a population. Marriage customs have evolved partly to control genetics. All societies appear to forbid incest, although it is defined in different ways. Some societies prohibit racial intermarriage, whereas others are far more concerned with class status. Some peoples in southern India and in Arabia favor cousin marriages to keep wealth or property in the family, and so promote genetic diseases. Technology can affect the genetics of the population. In one society people with poor vision make poor hunters and providers; in another eyeglasses have rendered eyesight an irrelevant criterion for marriage. Health services and technology enable people with genetic traits for dia-

betes, hemophilia, or deficient immune systems to live and reproduce. The immunological status of a population used to depend almost entirely on its age structure because experience with infectious agents accumulates over time. Now technology allows even newborns to be vaccinated for diphtheria, whooping cough, and other frequently fatal diseases. The degree of disease resistance of the population thus has become a cultural construct.

The Triangle of Human Ecology and Childhood Asthma

The three vertices of the triangle of human ecology can be used to model and summarize what is known about the complex ecology of a disease and to raise or test new hypotheses. Childhood asthma has a political-ecological context of where noxious facilities are located or how health care is made accessible. Figure 2.3, however, presents the cultural ecology of genetics, behaviors (breastfeeding, smoking, and using various chemicals), and the built and social (siblings, with their germs) environment and interactions. The question being addressed is the reason for increasing rates of asthma in rural and suburban areas without roaches, old housing, and other known risks. The hypothesis being tested is the so-called “hygiene hypothesis.” It avers that some of our autoimmune and inflammatory diseases are increasing because we have so thoroughly sterilized our habitat and removed the worms that we evolved with, even from our intestinal tracts, that our immune response cells no longer get properly trained. Instead we bring their powerful armament to bear on inappropriate targets in our own bodies. Do you understand why each variable in the diagram is in the vertex that it occupies?

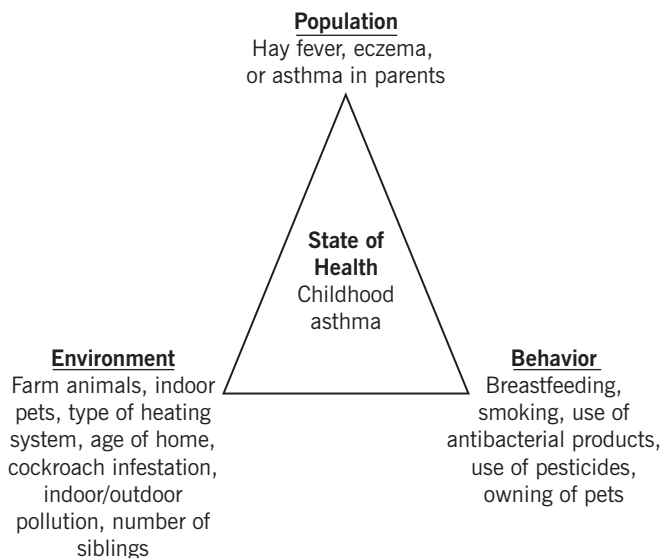


FIGURE 2.3. The triangle of human ecology and childhood asthma. Used by permission of Ashley Ward.

TRANSMISSION AND CREATION OF INFECTIOUS DISEASE

Evolution, Adaptation, and Emergence

We are embedded in life. Within our human body cells themselves live other creatures that, like the mitochondria that energize our every movement, have their own inheritance of their own DNA. Pieces of DNA, prions, and other bits of protein molecules of many kinds pass through generations, across species (e.g., pig, duck, and dog), and among humans alike. In our blood, our tissues, our internal spaces are hundreds of species of and billions of individual viruses and bacteria. Most of them were part of us long before we were people. We are encased in an envelope of life that is dependent on the excretions of our bodies, and their warmth and humidity. Not only do billions of staphylococci and many other types of bacteria cover us, but mites, lice, and perhaps human-specialized fleas as well. Most people have tiny mites that live in the follicle holes of their eyelashes, come out at night, and run around the face harvesting dead skin cells. We have not one but two species of lice—one for head hair and one for pubic hair—so unique to us that they would become extinct if we did. And all this life exists before we even open our mouths to consume the nutrients from other living things or practice agriculture as a process in the biosphere.

It is an adage that diseases represent evolution in progress. Health is adaptability, Dubos (1965) said. Natural selection works on the microbes and it works on us humans, who are facing selective pressures today—mostly from our own culture. Geography is not much interested in intracellular structures, but this ecological continuum of struggle, cooperation, adaptation, mobility, and information is life itself. Whatever we do with our politics and our technology, the microbes will continue their struggles to adapt, evolve, and survive. Just 30 years ago we thought we could eradicate malaria and tuberculosis, and perhaps even eliminate infectious disease altogether. Now we know better. As Dubos (1959) also said, the idea of perfect health without disease is a mirage.

When a disease agent either produces immunity or kills, there must be a population large enough to support its continuous circulation among susceptible people. Otherwise the disease cannot be maintained. This can be seen when infectious diseases such as colds or measles are introduced to small island populations. Military movements during World War II introduced many infections to Atlantic and Pacific islands and caused reported epidemics; yet contact years after the war found the diseases to be again totally absent. A few diseases, such as chickenpox, have evolved special strategies for surviving in small human groups (even during the Stone Age). Its agent, herpes group with *Varicella-zoster*, rarely kills. Instead the virus goes dormant in the nervous system to emerge years later in elders as “shingles,” shed from the skin to infect the next generation.

Most human infectious diseases, however, could not have existed before domestication of plants and animals led to agricultural systems that supported enormously increased human populations with food surplus. These people then created cities with social hierarchies in settlement systems with trade contacts. They transformed

the nature of human interaction. Disease agents that mutated from animal diseases, as must have happened many times previously in isolated areas, now found with cities and settlement systems a large enough interacting population to support their continuing circulation. The measles virus, for example, is closely related to both the virus that causes distemper in dogs and the one that causes rinderpest in cattle. As people domesticated animals and herded them together, microbes that mutated to be contagious among people could spread intensely among them. A settlement system of several hundred thousand interacting people is necessary to maintain the circulation of measles, for example. Because measles kills or leaves permanent immunity, infectious people must have contact with new susceptible people (i.e., children). The domestication of animals and their herding together, increasing intensity of infection among them, thus created conditions optimum for animal diseases to cross to people. Not only measles but smallpox and tuberculosis, for example, have their closest microbe relations in cattle disease agents (it was not accidental that cowpox could be used as a vaccine for smallpox).

The ecological information feedback in relationships is important to both microbe and host. If the infected host does not survive long enough for the disease agent to get to another host, both will die. A virulent agent, which multiplies so quickly it overwhelms its host and kills quickly, will not spread as widely as an agent that becomes attenuated and lets its host carry it around for a while. The English “sweating sickness” terrified Europe from 1485 to 1552, for example, but the microbe was so efficient at killing its hosts that it became extinct instead of emerging as a pandemic. Human hosts, on the other hand, are selected for resistance to the agent and, less directly, for developing successful cultural buffers. The most common model is that disease agents newly transferred to people are most virulent, become less severe through passage and over human generations, and eventually become “mild childhood diseases.” In the human gut, bacteria can cause massive tissue destruction and fatal fluid loss; more commonly, however, intestinal fauna produce symptoms only under physiological stress, or when antibiotics have disturbed the balance with other microbes that gave them feedback. We are still identifying for the first time these commensal residents—species unknown in their billions of organisms, and even truly symbiotic agents that produce vitamins for the human gut to absorb.

The general rule is that the more virulent and less adapted the agent is to people, the newer the disease. There are other microbe survival strategies, however. Disease agents vectored by insects in particular (discussed in Chapter 4) may benefit from concentrating in great numbers to increase their chance of being taken up by a flea or mosquito and inoculated into another person. Their host does not have to be mobile to infect insects, and how long the host needs to survive depends on how many insects there are to feed on him or her and pick up the infectious agent. Similarly, when many people or animals crowd together (people get *herded* too, as in classrooms), they increase the agent’s ease of contacting another susceptible host. Ewald (1993) discovered that the agent can therefore become more virulent (causing more vomiting, coughing, or diarrhea) to maximize transmission of

infection, instead of keeping its host active to maximize contact with others. This discovery and his assertion (Ewald, 2000) that progression toward commensalism over generations did not have to happen—that in fact some human behavior caused, or selected for, microbes' becoming more virulent—controversially shook the complacent human ecology perspective. Most microbes, in this view, mutate or combine with passing bits of DNA to create new strains that constantly probe for the best survival strategy (For the evolution of genetic strains of HIV and their selection by human behavior, see Ewald, 1994.)

Most contagious diseases evolved in separate population settlement systems, within which over centuries hosts and microbes had begun to adapt to each other. Diseases remained regionalized for long periods because people infected with a disease agent either died on the journey or became immune before they could reach another region. Even if a few people made contact, ecological relations usually worked against successful biological invasion (of the microbe). When the yellow fever virus was introduced to the mosquitoes and monkeys of South America, for example, it found a new place to relocate; when (presumably) it was introduced to the mosquitoes and monkeys of tropical Asia, however, it found its niche already occupied by the virus of dengue fever, which induced resistance to it. The protozoan that causes African sleeping sickness must have crossed the Atlantic with the slave trade, but it never found a fly like the tsetse, which exists only in Africa, to transfer it to a new host. Even though most disease agents are species-specific—that is, the dog roundworm does not reproduce in people, only the human roundworm (*Ascaris*) does; the human measles virus does not make a dog sick—adaptation will begin if enough contact is made.

Throughout history the progressive development of transportation and mobility has resulted in exchange of disease agents between formerly separated populations (see, e.g., Crosby, 1972; McNeill, 1976). The contact of the Roman and Chinese empires brought a deadly exchange of disease agents, as did the later contact of Europeans and Native Americans. The so-called “Columbian exchange” brought to the New World the disease agents that had evolved in domesticated animal herds and crossed to people in the cities of the Old World. It also brought a few disease systems, such as malaria and river blindness, that found equivalent insects to vector them and successfully relocated into parallel ecological niches. In the New World were large cities and systems of trade, but domestication had been directed at crops. The equivalent herding of many types of animals, and intensification and cross-over of their infections, did not occur. Syphilis (related to yaws, a tropical disease from soil bacteria spread by contact in humid, warm conditions—which might have adapted to millennia of traveling in very cold lands) probably spread with Columbus from the Caribbean to Europe. From Spain to the New World, however, came smallpox, measles, whooping cough, typhoid, diphtheria, influenza, and malaria, and (with the African slaves) yellow fever, onchocerciasis, schistosomiasis, and more. When a new disease agent spreads in a “virgin population,” in which there is no immunity from age or experience, it strikes down everyone. When everyone is sick, there is no one to gather firewood, cook, or even fetch water for the feverish.

Epidemic after epidemic spread through the Native American population, spreading on indigenous trade routes throughout North America decades before any English settlers appeared. There is a range of estimates, but probably an 80–90% reduction of population (i.e., the loss of 50–80 million people in the greatest holocaust of human existence) resulted. Two centuries later, the people of Hawaii and other Pacific islands did not fare much better.

Today disease agents can easily cross the ocean on a jet airplane. Chapter 5 discusses the accelerating mobility of the human population. Because people can travel and arrive at their destination long before they develop any symptoms, traditional buffers such as quarantine have been rendered moot. Any virus from any place in the world can be anywhere else within a week. The new conditions in an urbanizing world are also creating optimum conditions for the emergence of virulence in old infections and crossing of species by new ones, through the sheer intensity of transmission that has been made possible (see, e.g., the discussion of dengue hemorrhagic fever in Chapter 4). Finally, old diseases are reemerging as the agents learn to cope with cultural buffers. One of the greatest threats to public health today is the emergence of multiple-drug-resistant tuberculosis, some strains of which are now untreatable with any antibiotics. Streptococcus and other hospital infections are becoming very resistant, as are the salmonellas and shigellas of livestock and people. The resistance of malaria protozoa to chloroquine and other prophylactic chemicals has spread around the world. People on the Thai–Cambodian border today are resistant to every known antimalarial drug except the newest, artemisia. Resistance to that has already been induced in a laboratory. We may soon be more helpless before malaria than in the 19th century, when at least quinine worked for those who had it to use. Even lice are reemerging in the hair of American school children! The threat of emerging diseases is producing new research initiatives, new journals, popular books and movies, and new pressure for increased training of scientists and funding of surveillance. The components of the behavioral vertex struggle to adapt to the changing infectious hazard.

Chains of Direct Transmission

Contagious disease agents are passed from one person to another. Three types of transmission are frequently modeled. The first is direct transmission, either through physical contact or through the air. Agents of yaws, ringworm, and trachoma are spread by physical contact, and sometimes cold and other viruses can be communicated by hand as well. Sexual transmission is a specialized form of physical contact. More explosively, direct transmission by means of respiratory aerosol droplets can send pandemics of influenza, whooping cough, measles, or (previously) smallpox whizzing around the world. Some disease agents can also be spread from one person to another through contamination of food and/or water with the disease agents, which are then ingested by another person. This second model is known as the fecal–oral route of transmission. It can involve contact (usually between mother and small child), but the primary danger is water-borne transmission of typhoid,

cholera, leptospirosis, hepatitis, and many other of our most serious infections. This form of transmission is addressed soon. The important differentiation in transmission for the third model is with vectored diseases, agents of which are inoculated into people when they are being eaten by arthropods. This transmission chain is developed in Chapter 4.

Our worst respiratory contagions have met our best counteradaptations. Inevitably, a few now appear to be newly adaptable themselves. Most of our respiratory-transmitted diseases are now adapted to humans only. Smallpox is the only disease that medical science and public health have eradicated from Earth. It infected only people, and we developed an effective vaccine. In the 1960s and 1970s it then became the object of a determined drive by the World Health Organization (WHO) to identify and surround every case and contact with a zone of vaccination. Refugees from war and famines were finally all reached, and the need for smallpox vaccination was eliminated. Measles had higher mortality in much of the developing world than in the United States. Today most developing countries have reached higher levels of vaccination against measles than the United States has. It has become conceivable that measles too might be eradicated, as it is only a human disease. On the other hand, one of the three targeted diseases of the WHO Global Fund is becoming more resistant and dangerous: tuberculosis.

Tuberculosis kills more than 1.5 million people a year. Although incidence is still increasing—9.27 million in 2007 compared to 8.3 million in 2000—its prevalence is slowly decreasing (WHO, 2009). Its 2006 prevalence is shown in Plate 7. Over 55% of cases occur in Asia. There is an amazing story of human ecology behind its reemergence, as the WHO estimates that more than a third of people in the world are infected. Such numbers would indicate a very old disease, and indeed in 2007 a study of the mitochondrial DNA indicated that tuberculosis in people was older than the variant in cattle—a big surprise and contrary to the conventional wisdom of disease ecology. *Mycobacterium tuberculosis*, the “tubercle bacillus,” is necessary for the specific etiology of tuberculosis (TB) but not sufficient. Nutritional status, genetically based immunological resistance, age, and other population characteristics are important to success in “walling off” the bacilli within the lungs. Alteration of this status later in life can bring the bacilli out of dormancy into full-blown infectiousness and sickness. The HIV epidemic resulted in the 1990s in a tuberculosis epidemic. As the immune systems of millions were weakened, they died of tuberculosis. Usually only 5–10% of those infected become sick—an indication of an old disease that has been moving toward commensalism. But with a new person becoming infected every second (WHO, 2009a), this ancient scourge continues to be one of humankind’s greatest dangers. Daily treatment, traditionally prolonged over months, is needed. The “directly observed treatment, short course” (DOTS) public health program of individual contact and observed medicine consumption helped stop and reverse the resurgence of tuberculosis in the United States in 1985–1992. It has saved millions of people in developing countries. Nevertheless, the bad news is that tubercle bacilli are becoming resistant to multiple strains of antibiotics—and, in the still rare strains of XDR tuberculosis (the XDR stands for *extensive drug resistance*), to everything we have.

Known as the “white plague” in 19th-century Europe (because it turns people extremely pale, even ethereal) and as “consumption” in both the United Kingdom and the United States, tuberculosis has been and remains a disease of poverty (Farmer, 1997, 1999). Because tuberculosis is transmitted by respiratory droplets, people who are crowded together in conditions of poor ventilation—crowded tenements, homeless shelters, prisons—are most exposed. Improvements in nutritional levels and housing conditions resulted in a general decrease of tuberculosis in Europe and the United States even before effective antibiotics were found after World War II. Early detection and treatment of the disease prevent a sick person from infecting more than a dozen other people, on average, so lack of health care encourages high endemic levels. The highest prevalence and numbers of deaths are in the slums of Indian cities and among the HIV-positive poor in Africa. The new resistant strains, however, are emerging especially from harsh conditions in overcrowded Russian prisons, where practices of putting all tuberculosis-positive persons together result in successfully treated prisoners’ becoming continually reinfected. Those who don’t die in prison are eventually released to the streets, where the new strains are beginning to walk.

There is a “vaccine” for tuberculosis known as bacillus Calmette-Guérin (BCG). For several decades it has been widely used in developing countries with a high prevalence of tuberculosis. Although it seems to protect children from tuberculosis meningitis, it is of variable help for adult pulmonary tuberculosis. Often foreign students in the United States are repeatedly sent for chest X-rays. BCG has never been used in the United States, a country with a very low prevalence of tuberculosis (in 2005, fewer than 14,000 cases in 300 million), because it eliminates the most useful, cost-effective screening for tuberculosis: using tuberculin skin tests. BCG usually causes those vaccinated to test positive, however; hence the need for X-rays to identify cases. World and U.S. commissions have called for development of a new, effective vaccine. The WHO has tried to launch a new, all-out attack on tuberculosis before the emerging resistant strains march on the world.

Water-Related Infectious Disease

Directly transmitted diseases between people, with the exception of those that are sexually transmitted, need to exist for a period of time in the environment. For respiratory droplets of tuberculosis bacilli, viruses for influenza, and even helminth eggs passed into soil, this means exposure to the ultraviolet rays of sunlight. Outside the built environment, extremely cold temperatures of winter or dry conditions of heat may kill germs. Soil surfaces get flooded; water evaporates. Microbes, larval helminths, and such themselves face parasitism and predation. Such environmental conditions can eliminate or control infectious agents, or at least cause seasonality. Environmental conditions, especially continued warmth and humidity, can also nurture and protect disease agents. In wet-dry climates, the coming of the first rains in months can cause animal and human feces on the ground to wash into wells. Alternatively, the dry season can close wells and concentrate microbial life of all kinds in the remaining wells.

During the WHO's decade of water development in the 1980s, water-related diseases were classified into three development-related categories that continue to be useful: *water-borne* diseases, which are ingested; *water-washed* (or, perhaps more accurately, *unwashed*) diseases, which are preventable by hand/hair/clothes/floor washing and other hygiene; and *water-based* diseases, which are vectored diseases requiring water for the vector to breed (these are discussed in Chapter 4). As Table 2.1 shows, these include most of the major infectious diseases.

Water-unwashed diseases are those that are preventable by washing with water and soap. These include the common cold, which is often transmitted by contaminated hands; conditions caused by infectious agents that encyst when dry and blow as dust in the wind or contaminate money; and flea- or louse-vectored diseases, which hot soapy laundry and floor/rug washing can prevent. One of the most serious unwashed diseases today is trachoma. *Chlamydia trachomatis* now infects 70 million people, 5 million of them in the late stage and 2 million already blind. It infects the mucosa of the eyes, causing a conjunctivitis. Flies, especially *Musca sorbens*, crave the eye secretions for food and spread the infection from child to child. The flies lay their eggs in human feces. Dirty cloths and fingers also spread the infection. Repeated infection irritates and scars the underside of the eyelid, causing the lid to turn inward and perpetually scrape the cornea with rigid lashes. Serotypes of *Chlamydia* can also be sexually transmitted.

Water-borne diseases are “carried” in water, and often in contaminated food, and ingested (i.e., drunk or eaten) by people. Bacteria like diphtheria can spread in this way and kill whole armies. Some water-borne viruses, like hepatitis, continue to cause health consequences worldwide. Poliomyelitis is such a virus with a special ecology. This virus used to be so ubiquitous that it is thought almost all children ingested it very young, when they were still protected by their mothers' antibodies. The maternal antibodies provided a grace period in which children could develop their own antibodies to the polio virus. It wasn't until treated water and sanitation—life-saving public health improvements—prevented infants from getting infected with the polio virus that the horror of the paralyzing disease devel-

TABLE 2.1. Some Major Water-Related Diseases

Water-borne	Water-unwashed	Water-based
Typhoid	Intestinal worms	Malaria (> 1 mil./500 mil.)
Cholera (6,800/380,000)	Amebic dysentery	Filariasis
Hepatitis A	Colds	Japanese encephalitis (11,000/40,000)
Diphtheria	Typhus (louse-vectored)	River blindness
<i>E. coli</i> , salmonella	Plague (flea-vectored)	Schistosomiasis (7,000/200 mil.)
Polio (2,000)	Pesticide residue	Dengue/DHF (15,000/560,000)
Cryptosporidium		

Note. DHF, dengue hemorrhagic fever. Estimates of incidence from Brown (1996, Table 7.1). Mortality data updated from World Health Organization (1999).

oped. Encountering the virus for the first time without antibody protection at 8, or 14, or 39—perhaps while swimming, like Franklin D. Roosevelt—a few were killed outright, but many became paralyzed. Infantile paralysis was a public terror in the United States into the 1950s, when the first effective Salk vaccine was tested. Vaccination has since eliminated the disease in the United States, but as “safe drinking water” has become increasingly available and saved the lives of millions of children in the developing world, so have epidemics of paralysis followed. At the end of the 20th century, the WHO, the Carter Center, Rotary International, and other organizations launched a massive effort to make polio the second disease eradicated from the earth. By 2002 polio was eliminated from three of six WHO regions and reduced from 350,000 paralyzed children a year to 22,021 in 2009.

A few cases, perhaps introduced by refugee and other population movements, sputtered in a dozen countries and simmered in trouble spots in the Congo and Somalia/Ethiopia. That year there were only four countries where polio was clearly still endemic: Northern India, Pakistan, Afghanistan, and northern Nigeria. The eradication effort then sputtered, for reasons best studied as political ecology. Resistance to the vaccination program in northern Nigeria occurred in large part because the mullas thought that it was part of a conspiracy by the West to sterilize Moslem women through the “vaccine” and so to control African population. Disturbances of war and refugee movements from the Horn of Africa and the endless killing in the eastern Congo (already counted in millions of lives and millions displaced) to the frontiers of Pakistan then spread renewed cases. In the largest Muslim country, Indonesia, polio cases soared from none to 349 on Java (with a variant from India and despite its 90% vaccination coverage). As this text goes to print, there have been gains and further setbacks. After a great effort that vaccinated 24 million children in the summer of 2006, Indonesia is again free of polio. There remain 18 countries that in 2008 had 1,731 cases of polio (WHO, 2009b). The remaining foci now include Afghanistan, Pakistan, India, and Nepal (713 cases); countries in the savanna zone of Africa; and another cluster around the Congo basin (totaling, 1,018,861 cases in Nigeria alone).

Although case mortality for diarrhea has been reduced in the past 20 years by the diffusion of a simple and affordable new treatment technology—oral rehydration therapy (ORT), which uses an inexpensive and readily available mixture of salt, sugar, and boiled water to stop dehydration—diarrhea continues to be one of the leading killers in the developing world. About 2 million children a year die from it—more than 3% of total world mortality. The highest rates are in sub-Saharan Africa and South Asia, and of course especially among the malnourished (in accordance with Audy’s definition of health). Even in developed countries, moreover, there are increasing hazards from salmonellas and *Escherichia coli* (*E. coli*), ubiquitous intestinal bacteria that are developing antibiotic-resistant strains. Protozoa, such as *Cryptosporidium* from the intestinal tracts of cattle and *Giardia*, can survive municipal chlorination unscathed and cause epidemics and mortality even in U.S. cities. In the following section, the ecology of cholera is developed as an example of a water-borne infection that takes the direct, fecal–oral route.

The Cultural Ecology of Cholera

Cholera is a disease of poverty, but the bacteria responsible for the disease are linked to the environment. This disease is caused by the colonization of the small intestine with *Vibrio cholerae* 01 or 0139. An infective dose, which is quite large at about a million bacteria, causes massive watery diarrhea within 2 days of infection. Its victims become dehydrated. If they go untreated, the disease will kill as many as half of the people who contract it. ORT (see above) can stop the dehydration and has been a miracle treatment for millions of people in resource-poor countries who are infected with cholera. Everyone is susceptible to infection, with the course of the disease largely depending on the individual level of health reserves (resilience) and nutritional, immunological, and hydration status. *V. cholerae* bacteria inhabit estuaries, brackish waters, rivers, and ponds of coastal areas of the tropical world, especially in South Asia and Africa. In the behavioral vertex, socioeconomic, demographic change, and sanitary practices/hygiene are key drivers of the human dimension of outbreaks.

Figure 2.4 summarizes the ecology of cholera. Humans become infected with the bacteria through contaminated water or food and interaction with the natural aquatic reservoir. Poor people are much more likely to drink contaminated water and to eat contaminated food, as well as to interact with their aquatic environment.

Agent: *Vibrio cholerae* bacteria

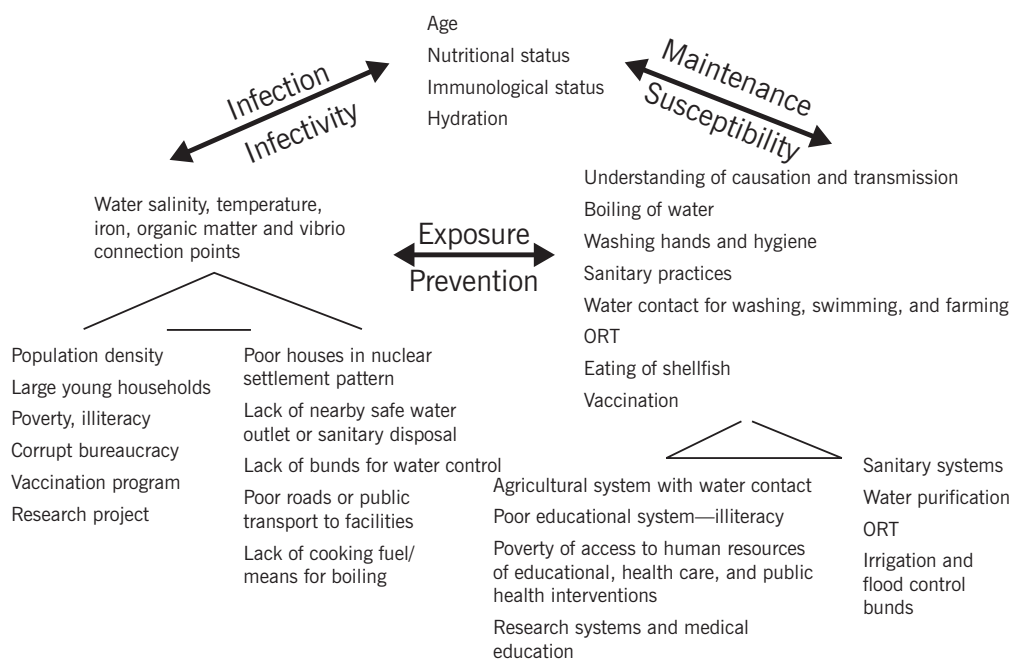


FIGURE 2.4. The triangle of human ecology and cholera.

The habitat of the disease begins with the aquatic flora and fauna, including algae, phytoplankton, zooplankton, and copepods, in which the bacteria flourish and which serve as reservoirs (Islam, Drasar, & Bradley, 1990). *V. cholerae* survival appears to be dependent also on abiotic characteristics, including alkalinity, salinity, iron concentration, temperature, nutrient concentrations, and the number of available attachment sites for the vibrio (plankton). The local sea surface temperature influences the growth of phytoplankton concentrations, which in turn encourage the subsequent multiplication of commensal copepods. Temporal fluctuations in cholera are likely related to variations in physical and nutritional aquatic parameters, including conditions in both ocean reservoirs and the brackish ponds and canals in endemic areas.

These factors are influenced by larger-scale climate variability (e.g., El Niño southern oscillation [ENSO] of the jetstream) and seasonal effects (i.e., sunlight, temperature, precipitation, monsoons), but the exact mechanisms are not well understood. Cholera has been shown to be related to upper troposphere humidity, cloud cover, and top-of-atmosphere absorbed solar radiation. Similarly, ENSO raises water temperature, bringing about increased zooplankton blooms that may influence longer cycles in cholera periodicity. Climate can affect the temperature of the sea surface as well as local ponds and rivers, possibly increasing the incidence of cholera through the faster growth rate of the pathogen in aquatic environments. The heating of surface water may lead to an increase in the phytoplankton blooms that feed zooplankton, encouraging the subsequent multiplication of commensal copepods that house *V. cholerae*. Periodic climatic and temperature cycles such as ENSO have an effect on cholera variability (Koelle & Pascual, 2004). Warm El Niño events are linked to cholera outbreaks in Bangladesh, and also to the emergence of new harmful algal blooms throughout Asia. Cholera outbreaks in Bangladesh between 1966 and 2002 demonstrate a 9- to 14-month lag between Indian Ocean sea surface temperatures on the one hand, and atmospheric temperature changes and subsequent cholera outbreaks on the other (examples include the 1987–1988 El Niño, the 1988–1989 La Niña, and the 1997–1998 El Niño).

The built environment in Bangladesh is based on houses clustered on raised land in rice fields. Some people live on houseboats (see Plate 3). The population density of this land is very high, and the population is very young. Embankments called *bunds* have been built to control water flow, especially to keep floodwaters from the river's flow inundating the fields and villages. One of us (Michael E. Emch) has done much of his research about the alterations in concentration and persistence of the vibrio associated with this water control (Emch, Feldacker, Islam, & Ali, 2008; Emch et al., 2008). Water sources and sanitation systems are illustrated in Plates 4 and 5. Some wells are better protected than others, but there are few outlets people can access for safe drinking. Even many of the tube wells protected from surface contamination produce arsenic poisoning. Roads and transport facilities to get sick children to hospitals for intravenous rehydration are few and poor. The local fuel for heating in this land of few trees is fuelwood, which is thus expensive, relatively speaking. The social environment includes the population density and large young households (although the average fertility rate of Bangladesh's total population has been lowered below three children today).

Although the environment is important for whether cholera exists in a particular area of the world, the magnitude of epidemics is related to socioeconomic factors (Emch, 1999). Secondary transmission occurs when the infection is passed by the persons directly infected by the original source to other people. Perhaps the original persons contaminate a different water site, handle food or dishes, touch a child, or are helpfully touched in their own illness. Factors affecting secondary transmission control the extent to which the disease is present (i.e., whether it will reach epidemic proportions). The severity of secondary transmission depends primarily on water sources for household consumption. People who use contaminated surface water for drinking, cooking, and bathing are more likely to contract cholera than those who do not. In other words, there is an inverse relationship between cholera and access to improved water sources and sanitation. People living on less than a dollar a day, of course, cannot afford fuelwood for boiling water. We could virtually eliminate cholera throughout the world if drinking water and sanitation in the poor countries of the world could be improved to the standards of rich countries such as the United States. In the 1800s there were large cholera epidemics in the United States and Europe, but it has been virtually eliminated in these areas since then. In contrast, cholera has rapidly expanded throughout Africa during the past decade.

Other cultural buffers to cholera have been developed and practiced or utilized. The oldest, in a new promotion, is illustrated in Figure 2.5. It shows a cholera education billboard in Malawi, with a heading that can be roughly translated as "Get Rid of Cholera in Malawi." The main message states that people should wash their hands every time they use the toilet, after they change a baby's diaper, before they prepare food, before they eat food, and before they breastfeed. There are other,



FIGURE 2.5. Cholera education advertisement in Malawi.

newer cultural buffers as well. Water is sometimes filtered through multiple layers of old cloth in the household. New ceramic filters have been designed (by graduate students in public health at the University of North Carolina); in the future these may be made locally and provide employment as well as safer water. Efforts to design and distribute locally functional solar cookers may help boil water also.

NUTRITION AND HEALTH

“Good nutrition is a cornerstone for health.” The interaction between nutrition and disease is two-way. In themselves, deficiencies of minerals and vitamins constitute diseases, but malnutrition also makes the body more susceptible to disease. Antibodies are not produced quickly or abundantly, tissues heal slowly, bones are more easily broken, and membranes become permeable. A malnourished person is more likely to get sick, and the course of the sickness is more likely to be severe and drawn out; the level of health is lowered by the malnutrition, and the person is less able to rally from new insults. Sickness may cause malnutrition. The body commonly responds to fever and nausea with a loss of appetite. Feverish people tend to sleep for long periods and not to eat regular meals. If diarrhea or vomiting occurs, the body’s absorption of nutrients is hindered. Prolonged diarrhea drains the body of salts and disturbs the electrolyte balance. When the body copes with infection, high levels of nitrogen are eliminated in the urine. Even the mild reaction resulting from vaccination will cause nitrogen to be eliminated, so that the body requires protein supplementation for recovery. It is easy to understand how a newly weaned child with borderline nutrition can develop kwashiorkor (a form of severe protein deficiency; see below) after measles, or pneumonia after malnutrition.

Dietary excess can also result in disease conditions. Sugar consumption, for example, is related to the occurrence of diabetes and dental caries. Cholesterol, an animal fat, is related to atherosclerosis. Several salts can cause hypertension. The “Western diet” that is spreading around the world is increasingly associated with obesity and its multiple complications for health.

Human Nutritional Needs

We humans need most basically (1) protein for the substance of our bodies, from DNA and enzymes to blood, muscle, and brain; and (2) energy for life, to breathe and move, and to maintain body heat. Lack of calories sufficient to maintain life results in skeletal figures that have consumed their own tissues; this is the form of starvation known as *marasmus*. Lack of protein sufficient to maintain substance, even in the presence of enough calories of energy, results in the extended bellies and bloated bodies (with water substituted to fill cells) of the form of starvation known as *kwashiorkor*, which is an African word for a protein deficiency disease. Since people with any protein at all will consume it to supply energy when there is none, the deficiencies are usually measured together as *protein–energy malnutrition* (PEM).

Besides energy and protein, of course, people need minerals and vitamins. Some of these are needed in relatively large quantities (macronutrients like iron and calcium), and some as mere traces (micronutrients like biotin and cobalt). There are new dietary concerns today related to protective elements against degenerative diseases, such as antioxidants. Table 2.2 lists most human nutritional needs, sources of nutrients, and deficiency diseases.

Until the global trade, marketing, and food information or promotion of recent times, the health benefits and deficiencies of traditional diets varied regionally, as discussed below. Kwashiorkor is most common in the wet tropics. There are many reasons for this, including the leaching and low nitrogen content of soils, as well as the agricultural importance of starchy root crops such as cassava and taro. Although some dwarf, resistant breeds of cattle have been developed, and there are today vigorous, internationally supported breeding programs, over the centuries trypanosomiasis (sleeping sickness) made it impossible to raise livestock in large areas of tropical Africa. This has meant loss of animal protein from milk (for those who are not lactose-intolerant) and meat. Traditional weaning foods are not high in protein. In some tropical regions, especially places near the sea, fish are an important source of protein. In tropical Asia, dried fish made into a sauce is a high-protein condiment universally used on rice and traded far inland (a similar sauce is made out of soybeans instead of fish within the Chinese culture realm).

Ecologies of Deficiencies

Beriberi

The cause of *beriberi*—a disease involving loss of temperature sensation, progressive weakness, paralysis, and eventual heart failure—was a great mystery for centuries. The deficiency association of beriberi with milled rice is so well known today that the marvel of the mystery's solution has largely been lost. Beriberi results from the lack of thiamine, vitamin B1 (the 1 signifies that this was the very first vitamin ever discovered), which is needed for carbohydrate metabolism. Just at plowing time, when the need for carbohydrate energy is high, beriberi would emerge to paralyze farmers. Rice has thiamine in its husk. When rice is milled at home with mortar and pestle, some husk and thiamine remain in the finished product. Milling rice by hand is arduous and time-consuming, however; and the pretty, white, polished mill rice is greatly preferred. The polished rice also stores better.

By the end of the 19th century, it was known that beriberi befell institutionally crowded populations (such as monks and soldiers, and especially sailors and prisoners); that it was most common on the coasts, whereas people in the interior of countries seemed immune; that newly arrived people usually took several months to develop the disease; that people such as scholars who led sedentary lives, poor laborers eating rice, and soldiers fed on meat and eggs were all prone to develop beriberi; that it was epidemic at the beginning of the rainy season, when plowing was done, but was unrelated to climate because it had spread in the Amazon basin and Southeast Asia and the climate had not changed; that it occurred from Brazil to

TABLE 2.2. Human Nutritional Needs, Sources, and Deficiencies

Nutrient	Major body function	Major sources	Deficiency disease
Balanced amino acids	Precursors of structural protein, enzymes, antibodies, hormones	Good sources: meat, fish, dairy products, legumes Adequate sources: rice, corn, wheat Poor sources: cassava, yams	Kwashiorkor
Fatty acids	Cell membrane function, regulation of digestion and hormones	Vegetable oils	Skin lesions
<u>Water-soluble vitamins</u>			
B1 (thiamine)	Removal of carbon dioxide	Organ meats, whole grains, legumes	Beriberi (nerve changes, weakness, heart failure)
B2 (riboflavin)	Energy metabolism	Widely available	Lip cracks, eye lesions
Niacin	Oxidation reduction reactions	Liver, lean meats, legumes (can be formed from amino tryptophan)	Pellagra (skin rash, nerve and mental disorders)
B6 (pyridoxine)	Amino acid metabolism	Meats, vegetables, whole grains	Irritability, convulsions, kidney stones
Pantothenic acid	Energy metabolism	Widely available	Fatigue, sleep disturbances
Folacin	Nucleic acid and amino acid metabolism	Legumes, green vegetables, whole grains	Anemia, diarrhea
B12	Nucleic acid metabolism	Muscle meats, eggs, dairy products (not in plant foods)	Pernicious anemia, nerve disorder
Biotin	Fat synthesis, amino acid metabolism	Legumes, vegetables, meats	Fatigue, depression
Choline	Precursor of neurotransmitter	Yolk, liver, grains, legumes	Not known
C (ascorbic acid)	Maintains cartilage, bone, dentine; collagen synthesis (= about 30% of body protein)	Citrus fruits, tomatoes, green vegetables	Scurvy (degeneration of skin, teeth, blood vessels)
<u>Fat-soluble vitamins</u>			
A (retinol)	Constituent of visual pigment; maintenance of epithelial tissue	A or pro-A (beta-carotenes) in green vegetables, dairy products	Xerophthalmia, night blindness, blindness
D	Growth and mineralization of bone, absorption of calcium	Cod liver oil, eggs, fortified dairy products	Rickets (bone deformity), osteomalacia in adults
E	Antioxidant to prevent cell membrane change	Green leafy vegetables, seeds, shortenings	Anemia
K	Blood clotting	Green leafy vegetables	Internal hemorrhages, severe bleeding

(cont.)

TABLE 2.2. (cont.)

Nutrient	Major body function	Major sources	Deficiency disease
<u>Major minerals</u>			
Calcium	Bones and teeth, blood clotting, nerve transmission	Dairy products, dark green vegetables, legumes	Stunted growth, rickets, osteoporosis, convulsions
Phosphorus	Bones and teeth, body pH balance	Dairy products, meat, poultry, grains	Weakness, demineralization of bones
Potassium	Body pH balance, water balance, nerve function	Meats, milk, fruits	Muscular weakness, paralysis
Chlorine and sodium	pH balance, water balance, formation of gastric juices	Common salt	Muscle cramps, apathy
Magnesium	Activates enzymes, protein synthesis and nerve transmission	Whole grains, green leafy vegetables	Growth failure, weakness, spasms, behavioral disturbances
Iron	Constituent of hemoglobin and enzymes in energy metabolism	Eggs, organ meat, meats, legumes, whole grains, green leafy vegetables	Iron deficiency anemia
Fluorine	Bony structure and teeth	Water, seafood	Higher frequency of tooth decay
Zinc	Constituent of many enzymes, tissue healing, digestion	Widely distributed	Growth failure, small sex glands, slow healing
Copper	Constituent of enzymes of iron metabolism	Meats, water	Anemia
<u>Trace minerals</u>			
Iodine	Constituent of thyroid hormones	Marine fish, dairy products, vegetables (depending on soil)	Goiter
Selenium	Association with vitamin E	Seafood, meat, grains	Anemia, cardiovascular disease
Chromium	Glucose and energy metabolism	Fats, oils	Impaired ability to metabolize glucose
Manganese, molybdenum, cobalt, silicon, vanadium, tin, nickel	Constituents of enzymes; functions poorly known in humans; essential to animal nutrition	Organ meats, fats; some widely available	Not reported in humans

Note. From Scrimshaw and Young (1976, pp. 62–64). Copyright 1976 by *Scientific American*. Adapted by permission.

Argentina and in the U.S. South, but was responsible for up to half the hospitalizations in Kuala Lumpur, Manila, and Tokyo. Researchers looked for a toxin in bad food, for a chemical contaminant, and especially for the newly discovered germs. The Dutch discovered that by feeding their prisoners in Java brown rice, they could stop the deaths—but world medical scientists did not read Dutch. Finally Robert R. Williams, working with the military in what was then the U.S. colony of the Philippines, discovered that he could cure the disease with rice bran water. This discovery made it clear that beriberi was due to a nutritional deficiency, not a positive toxin. Williams later identified and synthesized the missing life chemical, the first vitamin, B1.

The distribution pattern of the disease was explained by industrial milling of rice and by its shipment and storage along coasts and to cities and institutions, but also by specific cultural behaviors like cooking techniques. Beriberi occurred around the world, but historically it was especially concentrated in Tokyo and parts of northern China, which were at the end of a long transportation chain of rice from southern regions of the country. The rice was thoroughly milled prior to shipment by canal boat or wagon. To help protect it from weevils, talcum powder was commonly added. This in turn led to the cooking custom of thoroughly washing the rice and even throwing out excess cooking water—customs that further diminished the thiamine, which is water-soluble. In contrast, in the southern part of India, medical researchers had puzzled over the absence of beriberi despite dependence on rice; the customary treatment there was to partially cook the rice before storage and before milling. This “parboiling” saturated the grain with the thiamin and so preserved the vitamin.

As part of economic development, power mills in the 1960s and 1970s spread along new railroad lines and roads into formerly rural and isolated areas in tropical Asia. Power milling or polishing almost completely removes the thiamine, so that without compensating changes in the population’s diet, the spread of economic development also spread beriberi in almost epidemic form. Enrichment of rice at the mills has proved to be an answer to the deterioration of nutritional ecology, but this did not happen easily. Because of international opposition to companies’ making money from poor farmers by selling their “quick fix” (supplementary vitamins) instead of improving the overall diet by reduction of poverty, it took decades and thousands of lives before the staple enrichment model of the United States was followed.

Pellagra

The best known of the staple-associated deficiency diseases in the United States is *pellagra*. The disease is first characterized by a distinctive butterfly-shaped rash on the back, which is easily reversed by consuming vitamin B3, or niacin. At advanced stages it results in nerve damage and insanity. Maize, or corn, has little niacin. The way that Native Americans prepared maize (grits) made the most of its niacin, but diet still needs to be supplemented with niacin-rich foods. In the South, settlers’ con-

tact with the Cherokees and other Native Americans was prolonged. Maize came to have a very important role in the Southerners' diet. By the end of winter there was often little left for the poor to eat except cornmeal—a situation that became more common in the late 19th century, when “King Cotton” was grown all the way up to sharecroppers' doors, and vegetables in the home garden were less cultivated. A niacin-deficient diet resulted. The majority of patients in southern mental institutions in the 1930s were suffering from pellagra. As poor people in parts of Africa, South America, and Asia turned to maize for most of their calories, pellagra became a problem in those areas also.

Vitamin A Deficiency

Deficiency of vitamin A continues to be one of the leading causes of blindness; this type of blindness is known as *xerophthalmia*. As Table 2.2 shows, diets deficient in dairy products and especially in green leafy vegetables and orange carotene are at risk for causing xerophthalmia. The cultivation of vegetables (many domesticated in temperate climes or altitudes) is not practiced by all cultures; sometimes it is no longer practiced for economic reasons.

Pastoralists tend to have ample protein but inadequate B vitamins, and sometimes deficient vitamin A also. For children one common result of vitamin A deficiency is blindness because the cornea becomes softened (keratomalacia). Historically, African pastoralists traded meat for vitamin-B-rich grains from sedentary farmers and for vitamin-A-rich palm oil from the inhabitants of chronically protein-deficient rain forest areas. New national borders and trading systems oriented toward Europe have changed the spatial structure of such nutritional exchanges.

Vitamin D Deficiency

The only vitamin humans make themselves is vitamin D, and this is largely what skin pigment variation is all about. The sun's ultraviolet radiation stimulates the human body to produce vitamin D, which is necessary for calcium metabolism and plays a role in immune systems and other cell system functions. The deposition of melanin under the skin serves in tropical lands to keep out the most harmful of the sun's rays, but sufficient ultraviolet still penetrates for people to produce vitamin D. At high latitudes, with low sun angles and short sun seasons, there was high selection pressure to eliminate pigment and absorb what ultraviolet was available. The rather elegant evolutionary solution to living in Mediterranean lands and other locales where sunlight varies a lot seasonally was the genetically given capacity to deposit pigment under ultraviolet rays and remove it in their absence—in other words, tanning. Human mobility and technology have, of course, greatly disrupted this solution! People from northern Europe, without the ability to tan, moved to places like Florida and Alabama; people with black skin moved to high latitudes where all rays were blocked out. Cultural buffers like wearing long sleeves and

big hats helped for a while to prevent skin cancer in persons of European origin, but these habits seem to have been lost, with dietary as well as skin-related consequences. The impact of the Industrial Revolution, with its coal dust pollution and indoor drudgery, was another matter entirely.

Rickets, a bone deformity disease caused by poor calcium metabolism, used to be common in high-latitude countries. Wrapped in heavy clothing when outdoors and spending most of the low-light winter period indoors, growing children did not get enough sunlight. The situation became worse in the coal-polluted cities and dark tenements of the early Industrial Revolution. Rickets not only bowed the legs; it affected the pelvis and thus complicated childbirth. But health, you recall, is adaptability. Fish liver is one of the few sources of vitamin D, and so through processes of cultural ecology, cod liver oil became a common tonic. Vitamin D is fat-soluble, and eventually the happy solution (for lactose-tolerant populations) of adding the vitamin to milk, where it could be ingested with the calcium that needed metabolizing, was found. This has not proved such a simple solution at the same latitudes in China, which does not have a dairy cuisine.

The human ecology of vitamin D continues to evolve. As the world population has aged (see the discussion of the demographic transition in Chapter 5), osteoporosis in elderly persons has become a great concern. Hips and other bones break easily and heal slowly because of inadequate calcium and, of course, inadequate vitamin D. In recent years deficiency of vitamin D has been implicated in everything from susceptibility to prostate cancer to low birthweight. American women have been determined to be severely deficient in the vitamin—African Americans even when taking supplementation.

Iodine Deficiency

Deficiency of iodine—an element needed by the thyroid to make the important metabolic control hormone, thyroxine—results commonly in the gross thyroid enlargement known as *goiter*. With more severe iodine deficiency, mental retardation in general and *cretinism* (a thyroid deficiency that causes dwarfed stature, bone dystrophy, and low basal metabolism, as well as mental retardation) in particular develop as the growth and development of young brains and bodies are impeded. Historically, regions with glaciated soils deficient in iodine have been most affected, as plants grown there could not take up the element for people to consume. Trade has been important for overcoming this deficiency, especially trade for sea fish (seawater being the major source of iodine). Sometimes there was even a political geography to goiter, as when the goiter problem in Nepal's population varied with whether Nepal had friendly relations with India (which could provide sea salt) or China (which could not). Universal salt iodization—for salt used in agriculture and food processing, as well as in households—has been “the agreed strategy.” In a “great achievement in the field of noncommunicable disease,” currently 68% of households in previously deficient areas have access to iodized salt (Delange, de Benoist, Pretell, & Dunn, 2001).

As recently as 1996, more than 10 million people with mental retardation and tens of thousands with cretinism were regionalized within China—a generation lost by the destruction of public health iodized salt projects as Mao's Cultural Revolution ravaged China. The ancient salt sources of these areas are dry basins in central Asia. Some current American–Chinese efforts involve adding iodine to irrigation water in the most severely affected districts. There are other regional concentrations, such as those in upland but tropical central Java, where causation is more complicated and involves goitrogenic plants in the diet that prevent the absorption of what iodine is consumed.

Radioactive isotopes of iodine are among the first consequences of nuclear radiation release, as happened at Chernobyl. One of the emergency public health responses that has been developed is immediate iodine table salt supplementation for children. This ecology, too, continues to evolve.

Regionalization of Diet and Deficiencies

Diet can often be associated with cropping pattern, season, or culture realm. Jacques May devoted the last 20 years of his life to studying the ecology of malnutrition and left us a voluminous literature on the subject (May, 1977; see also “Further Reading”). He showed appreciation of both the environmental and the cultural aspects of nutritional patterns.

Environmental parameters—of rainfall and evapotranspiration potentials, sunlight and seasonal changes, freezing and heat, soil and topography—put limits on what can be cultivated. Where droughts are frequent and soils tend to be saline, the various kinds of millets are reliable and yield good protein. Where water is abundant and temperatures are warm, rice returns the highest number of calories per acre and provides good nutrition. Winter wheat yields more than spring wheat, but winter wheat usually cannot be grown in Siberia, Heilongjiang (Manchuria), and much of Canada. Many of our grains evolved in places with long hours of sunlight during the growing season. They either fail to ripen or yield poorly under tropical conditions of an unvarying 12 hours of sunlight. The agricultural patterns of the world formed within these parameters. (These statements are generalizations, and exceptions of course exist; for example, rice is grown even in northern Japan, which uses specially adapted strains.) For most of human history 20 inches of rainfall marked the geographical limit of cultivation and the beginning of pastoralism, but in recent years irrigation from tube wells has greatly extended the amount of arable land.

To these physical parameters must be added the cultural patterns that have formed broad dietary regions. Food crops were domesticated in specific cultural hearths and then spread over paths of cultural contact. The natives of Middle America domesticated corn, squash, beans, tomatoes, chili peppers, and numerous vegetables. They did not domesticate many animals and had no equivalent of the cow, but the collective amino acids in corn, squash, and beans are well balanced and provide complete protein. In the Middle East wheat was domesticated and became the

staff of life. It was balanced by the domestication of cattle, sheep, goats, and other dairy and meat sources. In Southeast Asia rice was domesticated, and the fish cultivated in the rice fields provided the balancing protein. In West Africa the cocoyam and oil palm were domesticated and combined with millet.

The originally domesticated crops and animals continue to constitute the core regional diets. Each diet is balanced and potentially sufficient for people who can afford to get enough of it. Only the domesticated starch roots, especially manioc (otherwise known as tapioca or cassava), are so deficient in anything but calories that they constitute a poor diet. For most people the cereal grains and legume vegetables such as soybeans or chickpeas are the major sources of protein, although medieval Europeans and their Mediterranean predecessors developed a mixed farming system that used grain cultivation and animal husbandry to support each other. Diversity and variety in diet, however, are everywhere necessary to provide the full range of nutrients. A diet limited to staple foods is often the lot of the poor, and so regional diets are often associated with the specific nutritional deficiencies discussed above.

Famine also has its regional patterns. Starvation can claim the poor even in well-fed lands, of course, as a matter of social structure. Crop failures, floods, earthquakes, and hurricanes can strike anywhere. Hundreds of local famines have been recorded in Europe in the last few centuries. Repeatedly during the 20th century, the Huang (Yellow) River, "China's Sorrow," escaped its levees to spread across the north China plain and bring famine to tens of millions. The Indian monsoons can fail to produce an adequate period of rain several years in a row. Famines tend to occur in areas of unreliable and highly variable annual rainfall. The grasslands and marginal agricultural lands of the African savanna, the land under the rain shadow of the Western Ghats and margins of monsoon penetration in India, and the degraded scrublands of northeastern Brazil are all famine-prone. Famines are also commonly the results of warfare, as guerrillas or soldiers and bombers make cultivating fields and storing seed grain impossible, and armies of refugees are displaced. With international response, there is actually little outright starvation on earth today, despite the problems of population growth. There is mass malnutrition of various kinds, however, and most of that is due to the lack of means or right to get the food that does exist. Social scientists refer to this as *vulnerability* to malnutrition, resulting from the lack of *entitlement*, by social right or purchasing power. The political/economic structures of power are seldom more starkly revealed.

There is also the power of cultural preference, prescription, and prohibition (in the behavioral vertex). Swine would fit in well with Middle Eastern agriculture, but there are religious injunctions in both Islam and Judaism against eating pork. Valuable though cattle are for draft power, milk, dung, and hides in India, the herds would be better fed and stronger if they were culled for meat, but Hinduism proscribes it. The poor and homeless in the United States could get high-quality protein from the excess dogs killed daily by the animal pounds. The majority of the human race eats tasty dog meat, but in Western civilization it is taboo. Some people have vitamin A and iron deficiencies simply because they culturally do not want to grow

vegetables, not because of inadequate land. Most cultures have strict rules about what pregnant women, sick people, or infants can and cannot eat. Sometimes these prescriptions promote good nutrition, and sometimes they hinder it (according to scientific criteria, which themselves constitute a belief system). There is the widest possible range of opinions about what constitutes good food supplementation for a nursing infant and what children should eat after being weaned. Some people culturally refuse their children eggs, the best protein available to them, because eggs are believed to cause worms. Some infant mortality and kwashiorkor are caused by the lack of income to purchase anything but the local starchy staple, but much of these are caused by ignorance of what an infant needs for nourishment.

World dietary patterns are undergoing many changes. In the last few centuries, crops and animals have spread from their hearths to be incorporated into other dietary regimes. Potatoes have gone from South America to Europe, wheat and soybeans to the Americas, and maize to China. Food storage and processing, transportation and finance, and types of ovens and cooking fuels have changed. Local famines resulting from isolated valley floods, hurricanes, or insect outbreaks have largely been eliminated by regional and national economic and political cooperation as new scales of organization have developed. International trade; refrigeration; veterinary science; and the development of new protein drinks, vitamin supplements, and enrichment targets have changed world nutrition. All these nutritional changes affect the health status of populations: Because of radical changes in diet, the Japanese have grown taller, and as people in the United States change their diet, less heart disease has extended life expectancy. Human nutrition around the world is generally better than in past centuries. One result of this improvement is earlier age of menarche and its consequences.

The issues of food distribution can be studied and dealt with at a wide range of scales. The purchasing power of the Netherlands or Japan allows these nations to buy the food they need in the international market, whereas countries such as Bangladesh, which are far more self-sufficient in food, are considered to have a population problem. Issues of social equity have become important worldwide at the national scale as countries struggle with globalization of trade, resources, capital, pollution, and occupational restructuring. For example, the United States exports food to earn a large part of its foreign exchange, yet has its own malnourished people. As Americans begin to address global warming seriously, one immediate response of producing more corn for ethanol fuel has had the result of driving up the price of the grain, and hence the price of products ranging from Mexican tortillas to New England milk. At the family scale, a household may have adequate food, but if it is customary for the male head of the family to eat first and get his choice of any eggs, fish, or special foods, then the dietary intake of pregnant women and small children may be inadequate.

Measurement of the availability of nutrients and diets is difficult. Methods used include 24-hour recall, maintenance of food diaries, the weighing and sampling of foods as they are cooked, and the less intrusive feeding of a "ghost" visitor—an additional plate at the table, which can be emptied daily by an investigator. All methods have their drawbacks and deficiencies. Individual and household studies

are expensive. One result is that small-area data for nutrition are almost nonexistent. For some purposes state-level marketing data may be adequate, but few data of value are available for medical geographical research in nutrition. Today in the United States, the political ecology of food and diet is changing dramatically. Suburban and educated people and their food stores are trying to eat food grown locally, usually organically. On the other side, however, inner-city populations have limited access to any fresh produce. Some local efforts have tried to establish community vegetable gardens. Efforts to change diets to fight obesity have run into the expense of organically grown food, the paucity of retailers willing to open stores in high-crime areas, and the difficulties of local transportation and hours of availability. Building healthier diets for poor Americans will not be easy.

CONCLUSION

The customs, beliefs, and socioeconomic structures that characterize each global culture realm and local ethnic group create the environmental conditions and exposure patterns that result in geographic distribution of health and disease. Genetic factors, to be sure, often underlie susceptibility and resistance, but the geographic distribution of genes is also a result of adaptation to environment, population mobility, and cultural selection. Influences of the physical environment and of insect/other animal communities and habitats are crucial, but they are buffered and even formed by human agency. Entirely natural landscapes scarcely exist. Transmission chains of disease agents are especially related to water sources, their safe potability, and their use for hygiene and sanitation. Later, in Chapter 6, we'll see that these needs face dire futures with climate warming.

Medical geography explains the distribution of health and disease, and can identify efficient ways to intervene and distribute trained personnel and technology. Every disease and quality of health has its cultural ecology, its geographic regionalization, and its patterns of diffusion and change. We discuss how these patterns change in time and are affected by technological and economic development in Chapter 5. This has already been a long and complex chapter, presenting concepts and ideas as well as a lot of information. It should serve as a basis and continuing reference for the considerations of health and disease in the following chapters.

REFERENCES

- Amato, J. A. (2000). *Dust*. Berkeley: University of California Press.
- Audy, J. R. (1971). Measurement and diagnosis of health. In P. Shepard & D. McKinley (Eds.), *Environmental: Essays on the planet as a home* (pp. 140–162). Boston: Houghton Mifflin.
- Bias, W. B. (1981). Genetic polymorphisms and human disease. In H. Rothschild (Ed.), *Biocultural aspects of disease* (pp. 95–131). New York: Academic Press.
- Brown, L. R. (Ed.). (1996). *State of the world, 1996*. New York: Norton.
- Colwell, R. R., & Huq, A. (1994). Environmental reservoir of *Vibrio cholerae*—the causative agent of cholera. *Annals of the New York Academy of Sciences*, 740, 44–54.

- Crosby, A. (1972). *The Columbian exchange: Biological consequences of 1492*. Westport, CT: Greenwood Press.
- Delange, F., de Benoist, B., Pretell, E., & Dunn, J. T. (2001). Iodine deficiency in the world: Where do we stand at the turn of the century? *Thyroid*, 11(5), 437–447.
- DeLong, G. R., Leslie, P. W., Wang, S. H., Jiang, X. M., Zhang, M. L., Rakeman, M., et al. (1997). Effect on infant mortality of iodination of irrigation water in a severely iodine-deficient area of China. *Lancet*, 350, 771–773.
- Dubos, R. (1965). *Man adapting*. New Haven, CT: Yale University Press.
- Dubos, R. (1987a). *Mirage of health: Utopias, progress, and biological change*. New Brunswick, NJ: Rutgers University Press. (Original work published 1959)
- Dubos, R. (1987b). *The white plague: Tuberculosis, man, and society*. New Brunswick, NJ: Rutgers University Press.
- Elliott, D. E., Urban, J. F., Jr., Argo, C. K., & Weinstock, J. V. (2000). Does the failure to acquire helminthes parasites predispose to Crohn's disease? *FASEB Journal*, 14(12), 1848–1854.
- Emch, M. (1999). Diarrheal disease risk in Matlab, Bangladesh. *Social Science and Medicine*, 49, 519–530.
- Emch, M., Feldacker, C., Islam, M. S., & Ali, M. (2008). Seasonality of cholera from 1974 to 2005: A review of global patterns. *International Journal of Health Geographics*, 7(31), 1–33.
- Emch, M., Feldacker, C., Yunus, M., Streatfield, P. K., Thiem, V. D., Canh, D. G., et al. (2008). Local environmental drivers of cholera in Bangladesh and Vietnam. *American Journal of Tropical Medicine and Hygiene*, 78(5), 823–832.
- Evans, R. G., Barer, M. L., & Marmor, T. R. (1994). *Why are some people healthy and others not?: The determinants of health of populations*. New York: Aldine de Gruyter.
- Ewald, P. W. (1993). The evolution of virulence. *Scientific American*, 268, 86–93.
- Ewald, P. W. (1994). *The evolution of infectious disease*. New York: Oxford University Press.
- Ewald, P. W. (2000). *Plague time: The new germ theory of disease*. New York: Free Press.
- Farmer, P. (1997). Social scientists and the “new tuberculosis.” *Social Science and Medicine*, 44(3), 347–358.
- Farmer, P. (1999). *Infections and inequalities: The modern plagues*. Berkeley: University of California Press.
- Hunter, J. M. (1973). Geophagy in Africa and in the United States. *Geographical Review*, 63, 170–195.
- Islam, M. S., Drasar, B. S., & Bradley, D. J. (1990). Survival of toxigenic *Vibrio cholerae* O1 with a common duckweed, *Lemna minor*, in artificial aquatic ecosystems. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 84, 422–424.
- Koelle, K., & Pascual, M. (2004). Disentangling extrinsic from intrinsic factors in disease dynamics: A nonlinear time series approach with an application to cholera. *American Naturalist*, 163, 901–913.
- Marmot, M. G. (1986). Social inequalities in mortality: The social environment. In R. G. Wilkinson (Ed.), *Class and health: Research and longitudinal data*. London: Tavistock.
- Marmot, M. G., Kogevinas, M., & Elston, M. A. (1987). Social/economic status and disease. *Annual Review of Public Health*, 8, 111–135.
- May, J. M. (1958). *The ecology of human disease*. New York: MD Publications.
- May, J. M. (1977). Deficiency diseases. In G. M. Howe (Ed.), *A world geography of human diseases* (pp. 535–575). London: Academic Press.
- McKeown, T. (1988). *The origins of human disease*. Oxford, UK: Blackwell.
- McNeill, W. H. (1976). *Plagues and peoples*. New York: Doubleday.
- Morse, S. (Ed.). (1993). *Emerging viruses*. New York: Oxford University Press.
- Newman, J. L. (1995). From definition, to geography, to action, to reaction: The case of protein-energy malnutrition. *Annals of the Association of American Geographers*, 85, 233–245.
- Race, R. R., & Sanger, R. (1975). *Blood groups in man* (6th ed.). Oxford, UK: Blackwell.
- Rhode, J. E. (1986). Diarrhea is a nutritional disease. In L. Ladislaus-Sanei & P. E. Scully (Eds.), *ICORT II proceedings: Second International Conference on Oral Rehydration Therapy* (p. 36). Washington, DC: U.S. Agency for International Development.

- Scrimshaw, N. S., & Young, V. R. (1976). The requirements of human nutrition. *Scientific American*, 235, 50–64.
- Simoons, F. J. (1969). Primary adult lactose intake and the milking habit: A problem in biological and cultural interrelations. *American Journal of Digestive Diseases*, 14, 819–836.
- Simoons, F. J. (1970). Primary adult lactose intolerance and the milking habit: Part 2. A cultural historical hypothesis. *American Journal of Digestive Diseases*, 15, 695–710.
- Simoons, F. J. (1976). Geographic perspective on man's food quest. In D. Walcher, N. Kretchmer, & H. Barnett (Eds.), *Food, man, and society* (pp. 31–53). New York: Plenum Press.
- Smith, C. J., & Hanham, R. Q. (1982). *Alcohol abuse: Geographical perspectives* (Resource Publications in Geography). Washington, DC: Association of American Geographers.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science*, 291, 1304–1351.
- Ward, A. (2007). *The cultural ecology of asthma in Orange County, North Carolina*. Unpublished master's thesis, University of North Carolina at Chapel Hill.
- Watterlond, M. (1983). The telltale metabolism of alcoholics. *Science*, 83, 72–76.
- Watts, E. (1981). The biological race concept and diseases of modern man. In H. Rothschild (Ed.), *Biocultural aspects of disease* (pp. 1–25). New York: Academic Press.
- Whitmore, T. M. (1992). *Disease and death in early colonial Mexico*. Boulder, CO: Westview Press.
- Wills, C. (1996). *Yellow fever, black goddess: The coevolution of people and plagues*. Reading, MA: Addison-Wesley.
- World Health Organization (WHO). (1999). *World health report 1999*. Geneva: Author.
- World Health Organization (WHO). (2004). *Vitamin and mineral requirements in human nutrition* (2nd ed.). Geneva: Author.
- World Health Organization (WHO). (2009a). *Global tuberculosis report 2009*. Geneva: Author.
- World Health Organization (WHO). (2009b). *Vaccine-preventable diseases: Monitoring system 2009 global summary* (August update). Geneva: Author.
- Xu, L.-Q., Yu, S.-H., Jiang, Z.-X., Yang, J.-L., Lai, C.-Q., Zhang, X.-J., et al. (1995). Soil-transmitted helminthiasis: National survey in China. *Bulletin of the World Health Organization*, 73, 507–513.

FURTHER READING

- Abrahams, P. W., & Parsons, J. A. (1996). Geophagy in the tropics: A literature review. *Geographical Journal*, 162, 63–72.
- American Association for the Advancement of Science. (1994a). Genes and behavior issue. *Science*, 264, 1637–1816.
- American Association for the Advancement of Science. (1994b). Genome issue. *Science*, 265, 1981–2144.
- Boyden, S. V. (Ed.). (1970). *The impact of civilisation on the biology of man*. Toronto: University of Toronto Press.
- Cockburn, A. (1967). *Infectious diseases: Their evolution and eradication*. Springfield, IL: Thomas.
- Dando, W. A. (1976a). Man-made famines: Some geographical insights from an exploratory study of a millennium of Russian famines. *Ecology of Food and Nutrition*, 4, 219–234.
- Dando, W. A. (1976b). Six millennia of famine: Map and model. *Proceedings of the Association of American Geographers*, 8, 29–32.
- Diamond, J. (1997). *Guns, germs, and steel*. New York: Norton.
- Garrett, L. (1994). *The coming plague: Newly emerging diseases in a world out of balance*. New York: Farrar, Straus & Giroux.
- Gopalan, C. (1992). *Nutrition in developmental transition in South-East Asia*. New Delhi: World Health Organization.
- Howe, G. M. (1972). *Man, environment, and disease in Britain*. New York: Barnes & Noble.
- Learmonth, A. (1988). *Disease ecology*. Oxford, UK: Blackwell.

- Lederberg, J., Shope, R. E., Stanley, C., & Oaks, J. (1992). *Emerging infections*. Washington, DC: National Academy Press.
- Logan, M. H., & Hunt, E. E., Jr. (Eds.). (1978). *Health and the human condition: Perspectives on medical anthropology*. North Scituate, MA: Duxbury Press.
- Matossian, M. K. (1989). *Poisons of the past: Molds, epidemics, and history*. New Haven, CT: Yale University Press.
- May, J. M. (1961). *The ecology of malnutrition in the Far and Near East*. New York: Hafner.
- May, J. M. (1963). *The ecology of malnutrition in five countries of eastern and central Europe*. New York: Hafner.
- May, J. M. (1965). *The ecology of malnutrition in middle Africa*. New York: Hafner.
- May, J. M. (1966). *The ecology of malnutrition in central and south-eastern Europe*. New York: Hafner.
- May, J. M. (1967). *The ecology of malnutrition in northern Africa*. New York: Hafner.
- May, J. M. (1968). *The ecology of malnutrition in French-speaking countries of West Africa and Madagascar*. New York: Hafner.
- May, J. M., & McLellan, D. L. (1970). *The ecology of malnutrition in eastern Africa and four countries of western Africa*. New York: Hafner.
- May, J. M., & McLellan, D. L. (1971). *The ecology of malnutrition in seven countries of southern Africa and in Portuguese Guinea*. New York: Hafner.
- May, J. M., & McLellan, D. L. (1972). *The ecology of malnutrition in Mexico and Central America*. New York: Hafner.
- May, J. M., & McLellan, D. L. (1973). *The ecology of malnutrition in the Caribbean*. New York: Hafner.
- May, J. M., & McLellan, D. L. (1974a). *The ecology of malnutrition in eastern South America*. New York: Hafner.
- May, J. M., & McLellan, D. L. (1974b). *The ecology of malnutrition in western South America*. New York: Hafner.
- Schofield, C. J., Briceno-Leon, R., Kolstrup, N., Webb, D. J. T., & White, G. B. (1990). The role of house design in limiting vector-borne disease. In S. Cairncross, J. E. Hardy, & D. Satterthwaite (Eds.), *The poor die young* (pp. 189–212). London: Earthscan.
- Stannard, D. (1989). *Before the horror: The population of Hawaii on the eve of Western contact*. Honolulu: University of Hawaii Press.
- Stearns, S. C. (1999). *Evolution in health and disease*. New York: Oxford University Press.
- Thomas, L. (1974). *The lives of a cell*. New York: Viking Press.

VIGNETTE 2.1. Biological Classifications of Importance to Health

Medical and biological terminology can be very intimidating. The first stop on the way to read something in a medical journal is often the reference shelves to find a medical dictionary. It can still be confusing when a single parasite is referred to in two or three ways in different sources. This introduction to the classification system may be a useful reference.

An infectious disease's causative organism—variously known as a germ, microbe, pathogen, or parasite—is the *agent* of the disease. It may be a virus, a rickettsia, a bacterium, a protozoan, or a worm. The type of disease agent is important for understanding

the possibilities of intervention in the disease cycle. At the smallest and simplest end of the infectious agent continuum are various bits and strings of protein; tiny short loops of DNA; and similar mysterious molecules called prions, plasmids, and new things seemingly every year. They do not seem to meet the historic definitions of *life*. They have no nucleus of control, no container of their being, apparently no being to contain as they drift among individual animals and species of animals and even between animals and plants. But they can be replicated. They can change the behavior of DNA and, for example, make a strain of bacteria more virulent, or cause the destruction of brain cells. Prions are the agent of Creutzfeldt-Jakob disease in people, and of bovine spongiform encephalopathy (mad cow disease) in cattle. Next are the viruses, simple loops of information that formerly were not considered living, either. Outside living tissue, viruses cannot reproduce. Pharmaceutical drugs have little effect on them, except for the potential of some of the new pharmacopeia developed through research on AIDS. The best weapon against viruses is a vaccine.

Rickettsiae cannot live outside cells, but in other respects they are similar to bacteria and have been regarded as degenerated bacteria that have lost their containers. They live very well within arthropods and get introduced to people during "blood meals." Rickettsiae and bacteria both are sensitive to antibiotics, which means that infected people can be treated. Mass treatment programs based on a single injection are sometimes possible. Protozoa have more DNA information and more than one life stage, generally defying vaccines and learning to resist drugs. The ameba that causes dysentery, for example, exists as a trophozoite which can act commensally or invasively, causing massive pathology; or it can exist as a durable cyst, the infectious form that is resistant to chlorine, drying, or sun as it travels through water, food, or even dust on money. The protozoan that causes malaria, a plasmodium, has four life stages that have defied vaccine production; these include sexual reproduction inside an *Anopheles* mosquito. The more complicated organisms, protozoa and especially helminths (flukes, filaria, and intestinal worms), are difficult to treat on a mass basis because powerful drugs that have many adverse side effects must be used, requiring hospitalization. Hospitalizing half the population to treat schistosomiasis, for example, is not only too costly in agricultural work time and medical care, but can seem useless when people return home only to be reinfected. Treatment of individuals, such as returned American tourists, is usually possible. Helminths infecting various body organs are even more difficult to treat, especially when they exist in heavy loads of hundreds and thousands of organisms. When they die, the body attacks them. Only intestinal worms have a ready means of egress from the body; the others, dead and unable to disguise themselves, become masses of internal foreign matter (clogging the liver, lymph nodes, veins, etc.). Currently, there are no successful vaccines for any of the complicated organisms. They have enough genetic material that they seem able to adapt and even prevent complete immunity to them by their hosts.

Multiple terminologies are often used to refer to a specific insect or parasite. Usually this draws on genus or family classification to invoke known characteristics (of breeding, attaching, eating, etc.) that separate the groups. For example, an article might refer to "the trombiculid mite" or "the trematode," and, in the same paragraph, to "the fluke"

or “the schistosome.” People commonly say “insect bite” when they mean “arthropod bite” (e.g., ticks are not insects). A *filarial worm* is a long, thread-like worm that may live in lymph glands or under the skin, but that reproduces and disseminates itself by producing *microfilaria* (microscopic larvae) that arthropods can take up with a blood meal. Elephantiasis and river blindness are filarial diseases vectored by arthropods; roundworm and tapeworm infestations are not. To refer to a bacteria by its classification as a spirochete, or a vibrio, or a bacillus, is to comment on the bacteria’s mobility, its ability to survive in the soil or water, its contagion. Vignette 2.1, Table 1, shows the relationships of different scales of classification for some of the most important agents and vectors.

VIGNETTE 2.1, TABLE 1. Some Medically Important Taxonomy

Phylum	Order	Family	Genus	Species	Disease caused or vectored
Arthropoda	Acarina (class Arachnida)	Trombiculidae Ixodidae	<i>Trombicula</i> (mites)	<i>T. akamushi</i>	scrub typhus
			<i>Dermacentor</i> (ticks)	<i>D. andersoni</i> <i>D. variabilis</i> <i>I. dammini</i> <i>S. scabiei</i> <i>X. cheopis</i> <i>P. irritans</i>	RMSF, tularemia, etc. RMSF Lyme disease scabies plague tapeworm
	Siphonaptera	Sarcoptidae Pulicidae (fleas)	<i>Ixodes</i> <i>Sarcoptes</i> (mites) <i>Xenopsylla</i> <i>Pulex</i>	<i>P. humanus</i> <i>P. varians</i> <i>S. damnosum</i> <i>G. palpatis</i> <i>C. discalis</i> <i>C. fatigans</i> <i>C. pipiens</i> <i>Ae. aegypti</i> various, any	typhus, relapsing fever leishmaniasis river blindness African sleeping sickness tularemia filariasis encephalitis yellow fever, dengue fever malaria
		Diptera	<i>Pediculus</i> <i>Phlebotomus</i> (sandflies) <i>Simulium</i> (blackflies) <i>Glossina</i> (tsetse flies) <i>Chrysops</i> (deer flies) <i>Culex</i>		
		Psychodidae Simuliidae Tipulidae Tabanidae Culicidae (mosquitoes)	<i>Aedes</i> <i>Anopheles</i>		
			<i>Schistosoma</i> (blood fluke)	<i>S. hematobium</i> <i>S. japonicum</i> <i>S. mansoni</i> <i>F. buski</i> <i>G. hominus</i> <i>C. sinensis</i> <i>F. hepatica</i> <i>O. felineus</i> <i>P. westermani</i> <i>T. solium</i> <i>T. saginata</i> <i>E. granulosus</i>	schistosomiasis schistosomiasis schistosomiasis intestinal fluke intestinal fluke liver fluke liver fluke liver fluke lung fluke pork tapeworm beef tapeworm hydatidosis (dog tapeworm)
			<i>Fasciolopsis</i> <i>Gastrodiscoides</i> <i>Clonorchis</i> <i>Fascia</i> <i>Opisthrochlis</i> <i>Pargonomus</i> <i>Taenia</i> <i>Echinococcus</i>		
			<i>Ascaris</i> <i>Trichinella</i> <i>Ancylostoma</i> <i>Necator</i> <i>Strongyloides</i> <i>Wuchereria</i> <i>Onchocerca</i> <i>Dracunculus</i>	<i>A. lumbricoides</i> <i>T. spiralis</i> <i>A. duodenale</i> <i>N. americanus</i> <i>S. intestinalis</i> <i>W. bancrofti</i> <i>O. volvulus</i> <i>D. medinensis</i>	ascariasis (roundworm) trichinosis hookworm hookworm strongyloides bancroftian filariasis (elephantiasis) river blindness dracunculiasis (guinea worm)
	Nematodes (roundworms)	Ascaridoidea Trichurata Strongylata			
Platyhelminthes (flatworms)	Trematoda	Digenea			
Nematodes (roundworms)	Cestoda				
Nematodes (roundworms)	Filarioidea				

(cont.)

VIGNETTE 2.1, TABLE 1. (cont.)

Phylum	Order	Family	Genus	Species	Disease caused or vectored
Protozoa		Mastigophora (flagella) Trypanosomatidae	<i>Leishmania</i>	<i>L. donovani</i>	kala-azar
			<i>Trichomonas</i>	<i>T. vaginalis</i>	vaginitis
			<i>Trypanosoma</i>	<i>T. gambiense</i>	African sleeping sickness
				<i>T. rhodesiense</i>	African sleeping sickness
	Sarcodina (pseudopods) Sporozoa (no adult locomotion)	Entamoeba		<i>T. cruzi</i>	Chagas disease (American trypanosomiasis)
			<i>Giardia</i>	<i>G. lamblia</i>	diarrhea
			<i>Entamoeba</i>	<i>E. histolytica</i>	amebic dysentery
			<i>Plasmodium</i>	<i>P. falciparum</i>	falciparum malaria
				<i>P. vivax</i>	benign tertian malaria
				<i>P. malariae</i>	quartan malaria
			<i>Vibrio</i>	<i>V. cholerae</i>	cholera
			<i>Bacillus</i>	<i>B. anthracis</i>	anthrax
			<i>Clostridium</i>	<i>C. tetani</i>	tetanus
Bacteria		Mycobacteriaceae	<i>Yersinia</i>	<i>Y. pestis</i>	plague
			<i>Mycobacterium</i>	<i>M. tuberculosis</i>	tuberculosis
				<i>M. leprae</i>	leprosy
				<i>B. recurrentis</i>	relapsing fever
			<i>Borrelia</i>	<i>B. burgdorferi</i>	Lyme disease
			<i>Treponema</i>	<i>T. pallidum</i>	syphilis
			<i>Rickettsia</i>	<i>R. rickettsii</i>	Rocky Mountain spotted fever
				<i>R. prowazekii</i>	typhus
			<i>Coxiella</i>	<i>C. burnetii</i>	Q fever
Rickettsiae		Spirochaetales			
Virus	RNA viruses		<i>Picornavirus</i>		enteric viruses, rhinoviruses (intestinal, respiratory)
			<i>Arbovirus</i>	arthropod-borne	encephalitis, hemorrhagic fevers (various, numerous)
			<i>Myxovirus</i>		influenza, mumps, measles
			<i>Adenovirus</i>	<i>Retrovirus</i> HIV	AIDS
			<i>Herpesvirus</i>	<i>Varicella-zoster</i>	respiratory
					chickenpox-zoster (shingles)

CHAPTER 3

Maps and Geographic Information Systems in Medical Geography

Geography has a fundamental interest in variations over space and in how things are put together at different scales of thinking about them (see Chapter 12 for a more comprehensive discussion of scale). Maps have been used for hundreds of years to describe spatial variations of health and disease. Koch (2005) provides a meticulously documented history of disease mapping, from the Naples plague in the 1690s to the contemporary AIDS epidemic in Thailand. As historians are concerned with actions and influences considered over time, analyzed in periods of various lengths, so geographers are concerned with actions over distance and interrelationships considered across space, analyzed at various scales. The capabilities of computers and the capacity for digital transmission of information available today have transformed not so much the questions geographers ask as their ability to address them. Although, like other social scientists, geographers have for decades used statistics, only recently have technological developments allowed the power of the computer to be directed at spatial data management through new software packages. The map has become the tool known as a *geographic information system* (GIS). This chapter initially offers a brief history of the cartography of disease, followed by an introduction to basic cartographic principles and then by an overview of how one can use a GIS in health and disease studies. Chapter 9, "Neighborhoods and Health," and Chapter 12, "Scale, Spatial Analysis, and Geographic Visualization," build on this chapter and discuss more advanced concepts related to spatial distributions of health and disease, using examples introduced in other chapters of the book. This chapter is an overview of how spatial data are represented and visualized; Chapters 9 and 12 describe how spatial distributions are analyzed.

CARTOGRAPHY OF DISEASE

The use of GIS has its roots in the field of cartography, which is a field highly developed as both an art and a science. Cartography has advanced through the centuries along with mathematics, astronomy, and navigation. It has invoked the curiosity and imagination of countless minds. It has recently been computer-automated and is currently being revolutionized by satellites, including those that take complex digital pictures of the Earth's surface and those that constitute the *global positioning system* (GPS). People working in health sciences need to become aware of more and different ways they can spatially analyze and portray their data, and of the ways such maps not only can be used but can be abused to lie and distort, or simply to err.

The beginnings of medical geography in the United States lie in mapping, specifically in Jacques May's project of mapping major vectored, parasitic, and nutritional diseases at the world scale for the American Geographic Society in New York. As discussed in Chapter 1, between 1950 and 1954 he produced a series of 17 maps that were published with limited commentary as supplements to the *Geographic Review* (see the May entries in "References"). Several maps were later reduced and published in May's book *The Ecology of Human Disease* (1958). These maps were for most geographers the first ones that had ever brought distributions of disease and their ecological relationships with cultural patterns and physical environments to their attention. A few years later, Rodenwaldt and Jusat (1952–1961) edited the more comprehensive, three-volume *Welt-Seuchen Atlas* (*World Atlas of Epidemic Disease*) in German and English. This atlas included global and regional maps (with a special emphasis on Europe) and a commentary on the diseases. It had a wide impact in public health and international medical and development centers. Disease mapping has come a long way since May created his maps. Disease maps are now readily available on websites from governments and other sources. The *Atlas of United States Mortality* (National Center for Health Statistics, 1997), which presents color maps, text, and adjustment methods, is available on the Centers for Disease Control and Prevention (CDC) website. The last section of the present chapter, entitled "Disease Maps on the Web," describes many other online sources of disease maps.

In the GIS era, disease maps often display the results of sophisticated spatial-analytical operations performed in a GIS, such as the work by Openshaw and colleagues (Openshaw, Craft, Charlton, & Birch, 1988) that identified clusters of childhood leukemia in Great Britain. Also in Great Britain, Peter Haggett, Andrew Cliff, and their colleagues have written several books and atlases that use spatial-analytical techniques to reveal patterns of infectious disease transmission. These volumes include *Spatial Aspects of Influenza Epidemics* (Cliff, Haggett, & Ord, 1987), *Atlas of Disease Distributions* (Cliff & Haggett, 1988), *London International Atlas of AIDS* (Smallman-Raynor, Cliff, & Haggett, 1992), *Measles: An Historical Geography* (Haggett, Cliff, & Smallman-Raynor, 1993), *Deciphering Global Epidemics* (Haggett et al., 1998), *Island Epidemics* (Haggett et al., 2000), and *World Atlas of Epidemic Diseases* (Haggett et al., 2004). In the United States, Peter Gould (1993) has documented the

progression of the AIDS epidemic in the United States during the 1980s in *The Slow Plague: A Geography of the AIDS Pandemic*. These aforementioned cartographers of disease were the pioneers; there have been many since who have used maps and GIS to understand epidemics, and some of their work is listed in the “Further Reading” section of this chapter.

As an abstraction of reality, a map is a distillation of complex inter-relationships that the map maker *chooses* as important to represent. That is, all of reality cannot be put on a map. The cartographer must *generalize*. The teacher or researcher must decide which elements to represent: If roads, what type? If vegetation, how specific? If water courses, what size or permanence? If houses, what type, and in what detail? Although disease data are usually collected according to administrative organization, such as counties, a map can portray the patterns of habitat, wind-borne pollutants, or water sources and flows that render the administrative borders dangerous irrelevancies. Again, a map is a model of what a map maker believes crucial for analyzing etiology or communicating important relationships.

Geographers sometimes forget the power of a simple map. Yet much of the early geographic disease studies’ impact on our understanding of health stems from the use of maps. To quote an 1852 cholera study:

Geographical delineation is of the utmost value, and even indispensable; for while the symbols of the masses of statistical data in figures, however clearly they might be arranged in the Systematic Tables, present but a uniform appearance, the same data embodied in a Map, will convey at once, the relative bearing and proportion of the single data together with their position, extent, and distance, and thus, a Map will make visible to the eye the development and nature of any phenomenon in regard to its geographic distribution. (Petermann, 1852, cited in Gilbert, 1958, p. 178)

Modern medical geography (frequently called *medical topography* during the 1800s) began in Europe during the late 18th century, when most of today’s academic disciplines were not separately defined. Investigators, who were usually medical practitioners, described a place’s topography and climate as they related to health and disease. This, of course, was in the Hippocratic tradition as exemplified by the quotation at the start of this book. Although disease distributions were often described in detail, and the reports sometimes did contain detailed topographic maps, they did not contain disease maps. The yellow fever epidemics of the late 18th/early 19th centuries and the cholera outbreaks of the 19th generated the first disease maps. Dot maps of the distribution of yellow fever victims (which were clustered in filthy port and sailor lodging areas) were used by both contagionists and anti-contagionists in their argument over the nature of what caused that dreaded disease. Contagionists considered it a single disease brought by travelers from places already afflicted with yellow fever, whereas anticontagionists thought that it simply emerged from crowded, filthy urban areas. Apparently the first such map was produced by Dr. Valentine Seaman in his anticontagionist treatise on yellow fever in New York City in 1798. His work was continued over the next half century by many other people.

In 1852, Heinrich Berghaus published his *Physikalischer Atlas*. One of its eight sections included a number of medical maps and charts. These were the first medical maps included in an atlas, the first to show the distribution of a variety of epidemic and endemic diseases, and the first published by a major cartographer. The extraordinary quality of these maps represents a singularly important development in medical cartography. The most famous 19th-century disease map, used to argue the contagionists' analysis, was John Snow's (1855) dot map of cholera around the Broad Street water pump in London in 1854 (Figure 3.1). The clustering of cholera in the vicinity of the well supported Snow's contention that cholera was a water-borne disease, with the pump the local source of infection. He urged that the handle be removed and the pump shut down. It was, and the local incidence of cholera declined quickly.

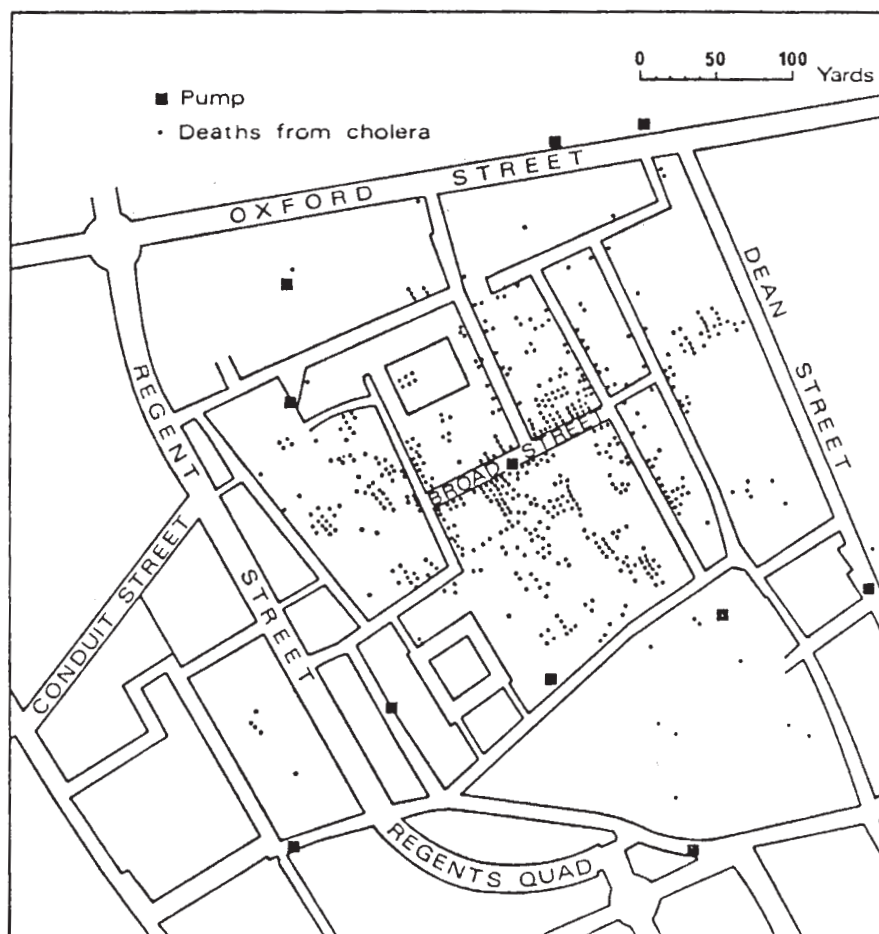


FIGURE 3.1. Snow's map of cholera. The affected well is clearly identified by the concentration of cases in its vicinity. From Howe (1972, p. 178). Copyright 1972 by G. M. Howe. Reprinted by permission.

TYPES OF MAPS: SOME ISSUES

Although a map can be visually effective and authoritative, it really can be no better than its data. The importance of political or administrative borders for the collection of data, and their frequent irrelevance to the patterns of distributions on the surface of the earth, have been intensely spotlighted by the availability of digitized satellite imagery of such distributions. The effort to reconcile this difference is at the core of much research in GIS and medical geography today. The origins, however, lie in the basic question of scale of analysis and of how different types of information are portrayed.

Map Scales

The type of projection used to portray the round globe on a flat sheet of paper remains important for global mapping of distributions or diffusion. Most health researchers of course will—and can—just use the base maps of the city blocks, counties, or countries that are readily available to them. However, the most convenient map may not be of the best scale, and the easiest route may not yield the best outcome. Base maps—blank maps waiting for data—need to be sought out for the scale most appropriate for aggregation and analysis of the topic at hand.

- A small scale (1:250,000 to 1:1,000,000 or more) is appropriate for continents, realms, and other large regions.
- A medium scale (1:25,000 to 1:250,000) is appropriate for states, state economic areas, health service areas, standard metropolitan statistical areas, and some minor civil divisions (counties, cities); topographic maps and coverage from the U.S. Geological Survey (USGS); and satellite coverage of vegetation, watersheds, and geological formations.
- A large scale (1:200 to 1:25,000) is appropriate for some minor civil divisions (cities, census tracts, etc.); satellite imagery of transportation routes, buildings, agricultural fields and local vegetational features; USGS topographic coverage; air photography special coverage; U.S. Census Bureau Topographically Integrated Geographic Encoding and Referencing (TIGER) files; and engineering blueprints (of electrical, sewerage, water, and other infrastructure).

Types of Data

Point data represent information that has no spatial dimension, but of course this is a matter of scale. An individual with a case of tuberculosis may be represented by a point on a city map; a city's population can become a point on a smaller-scale national map. These kinds of data are most commonly presented in *dot distribution maps* (like Snow's; see also Figure 7.4). The dots can become quite elaborate. A set of graduated circles indicating population or incidence size (Figure 3.2), a set of pie graphs with proportions for outcomes (Figure 3.3; see also Figure 5.3) or types of

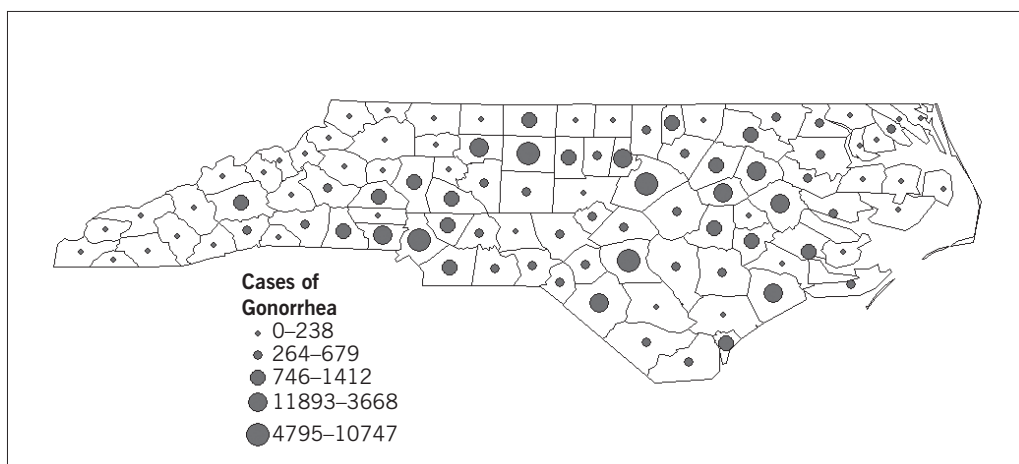


FIGURE 3.2. Graduated circle map of gonorrhea cases by county in North Carolina.

people, and even a climagraph showing seasonal distribution of data can all be used as point symbols.

Lines are used to show direction of movement (*flow maps*), often with the thickness of the line being *scalar* (proportional to the amount that is moving) (see Figure 10.3). Most commonly, lines are used to connect points and so portray new information. *Isolines* (also known as *isopleths* or *isarithms*) can connect points of equal data value (see Figures 7.7 and 10.6). These are most familiar as lines of equal altitude on topographic sheets or lines of equal atmospheric pressure on weather maps. In potential maps lines connect points of equal *population potential* (see Vignette 11.1, Figure 2); in diffusion maps they connect equal times of arrival; in health service

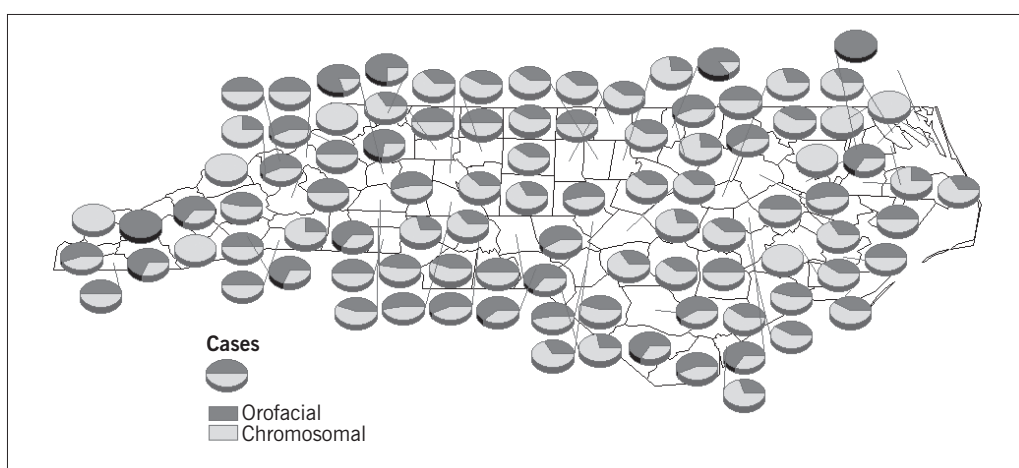


FIGURE 3.3. Pie graph map of birth defect type by county in North Carolina.

maps they can connect equal proportions of the population using a facility or equal distances traveled to the facility. Isolines are two-dimensional representations of three-dimensional reality because they have both extensions in space and values (i.e., X-, Y-, and Z-axes). The area that lines separate into different spaces is continuous, and points are discrete. To connect the points with a line through area with no data points, values are *interpolated*. For example, a line connecting points of value 10 could be drawn midway between points of value 5 and value 15, as well as through points of value 10 itself. This used to be a laborious as well as a specialized art, but computer programs do the calculations routinely now. Careful inspection and human art are still required, however, as there can be more than one way to configure such interpolated line patterns, and the simplest computer solution is not always the most sensible. Figure 3.4 is an isoline map of blood lead levels determined by laboratory screening tests, allocated to the residential locations of the children. The pattern of distribution identified is much richer and more intricate than a mere dot distribution of cases that exceeded a certain limit.

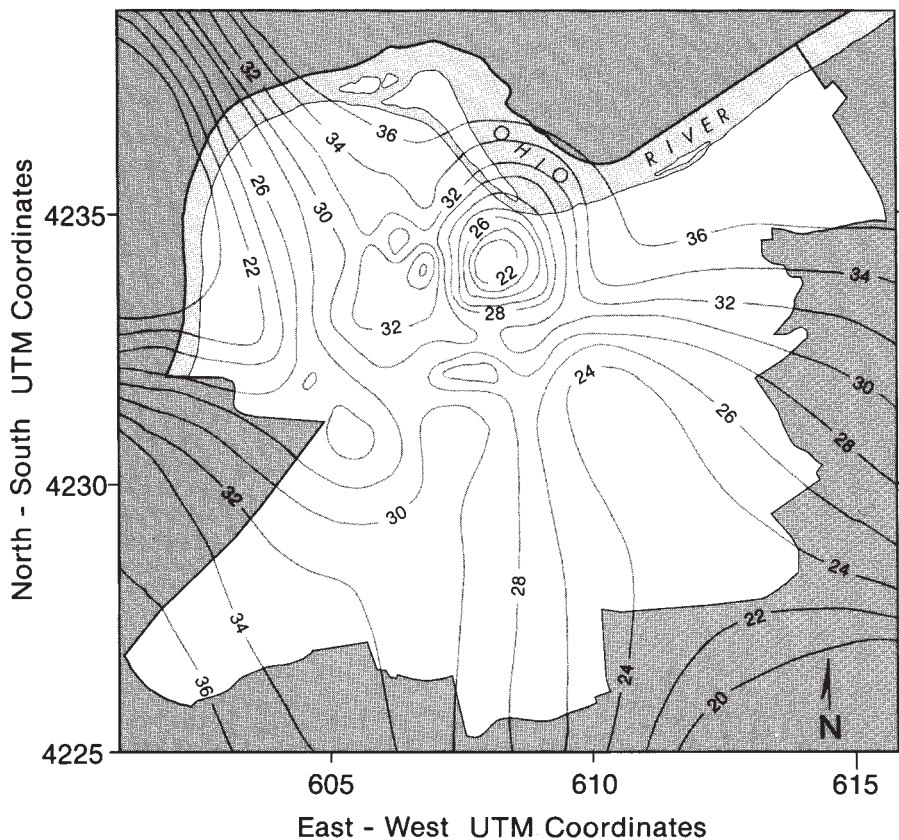


FIGURE 3.4. Lines of equal pediatric blood levels, Louisville, Kentucky, 1979. Values are micrograms per deciliter of blood.

Area information is symbolized on thematic maps by shades, patterns, or colors covering the area and symbolizing the class or category to which the information value belongs. This kind of map is known as a *choropleth map*. Choropleth maps are the familiar maps of data distribution by county, state, or nation, but the areas do not have to be politically delimited. Vegetation regions, soil regions, activity areas, or population circulation regions—any areas that can be discretely defined—may be appropriate units for choroplethic thematic mapping.

Critical Intervals: Mapping Patterns

Choropleth maps are the most commonly used maps for presenting disease or health service information. The question is how to divide this data distribution into intervals (i.e., create classes) to be symbolized with different shading on a choropleth map. Figures 3.5a through 3.5d are examples of choropleth maps of North Carolina infant mortality rates classified in different ways (see also Figures 7.2 and 8.2, and Plates 6, 7, 8, 12, 13, and 15).

A map made well can communicate a lot of information. The information, however, can be distorted if the intervals are not appropriate to the data distribution. First, there can be too many intervals. Students sometimes think there will be more detail in the distribution if they map 10 or 15 categories, but cartographic research has determined that the mind can only differentiate patterns from six or seven categories in black and white, and fewer in color. If too few categories are used, of course, detail and information will be lost. Regardless of how many categories are used, selecting shading, colors, or patterns to grade continuously from one class to another will eliminate the need to consult the key for every observation and will allow whatever pattern there is to appear. When a series of maps compares distributions over time periods or different areas, consistent class selection is necessary for comparison; otherwise, clusters and patterns will seem similarly significant, even though actual prevalence varies enormously. The most popular method in health sciences is to map *quantiles* (quantitative divisions), especially *quintiles* (20% categories) or *quartiles* (25% categories). There can be confusion about whether the data range (from the lowest point to the highest) or number of observations (i.e., geographical units) is being so divided and classified. Sometimes the default option of computerized mapping programs does the former (i.e., divides the range of data into five categories), whereas the researcher intends the latter (i.e., to divide the number of observations into five groups).

Figure 3.5A illustrates the most intuitive way of categorizing classes of infant mortality rates: division by natural breaks. When a frequency distribution is plotted with the values on one axis and the number of observations accumulated on the other, “breaks” usually appear in the curve. That is, many observations accumulate over a range of values, and then the rate of accumulation slows and even levels out until a new range of values is reached. These natural breaks in the data distribution can sometimes be more interesting in an etiological sense than the standardized breaks are; the problem with this method is the difficulty of comparing the patterns on one map with those on another done the same way.

Figure 3.5B shows the classification of infant mortality rates into five equal divisions of the range of the data from lowest to highest values. These equal divisions are a common default setting on commercial mapping programs and are easily interpreted if the data distribution is normal. Otherwise, this method can result in categories that are either empty or have few observations. In this map, the interval from 22.6 to 30.0 contains only one county.

Figure 3.5C is truly a map of quintiles—classes containing 20% of data observations (counties). Figure 3.5D shows the infant mortality from measles as mapped by standard deviations, not by acres of corn or trees or desert. A county-level map of the United States communicates the disease rates of small numbers of people over large areas west of the Mississippi, and crams the rates of millions of people in places on

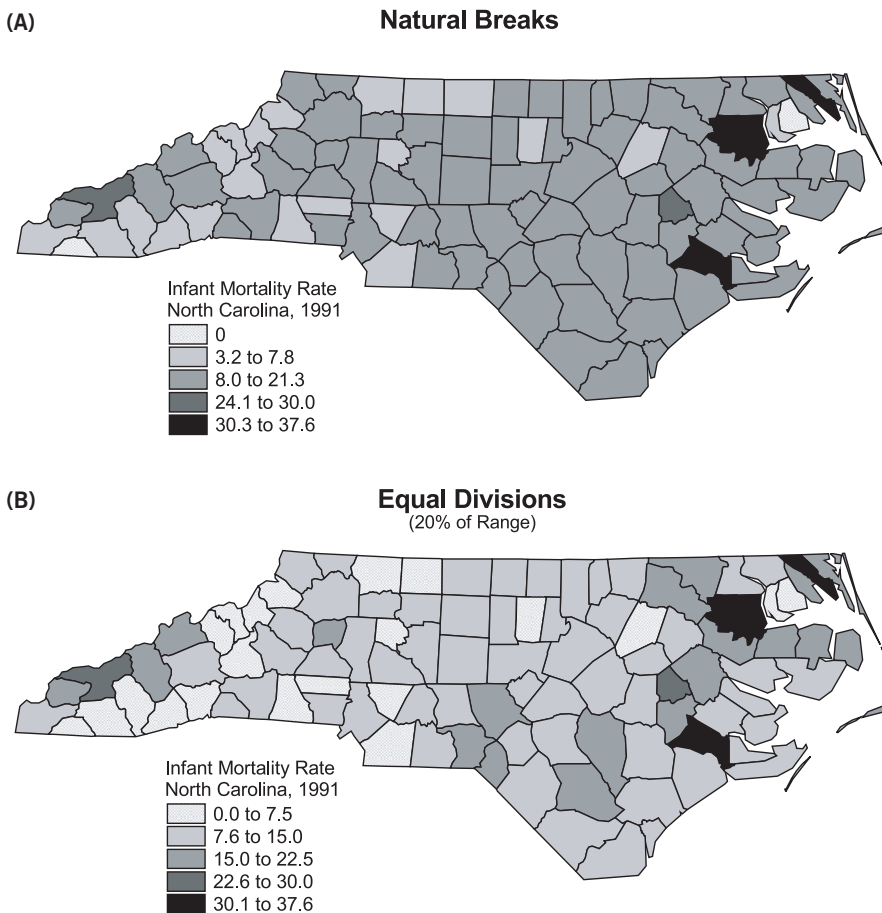


FIGURE 3.5. Choropleth maps of infant mortality in 1991 in North Carolina according to different means of interval determination. (A) Intervals are determined by natural breaks. (B) The range of rates is divided into five equal divisions. (C) The number of observations (100 counties) is divided into five equal groups, or quintiles. (D) Standard deviations from the mean value for infant mortality determine the intervals.

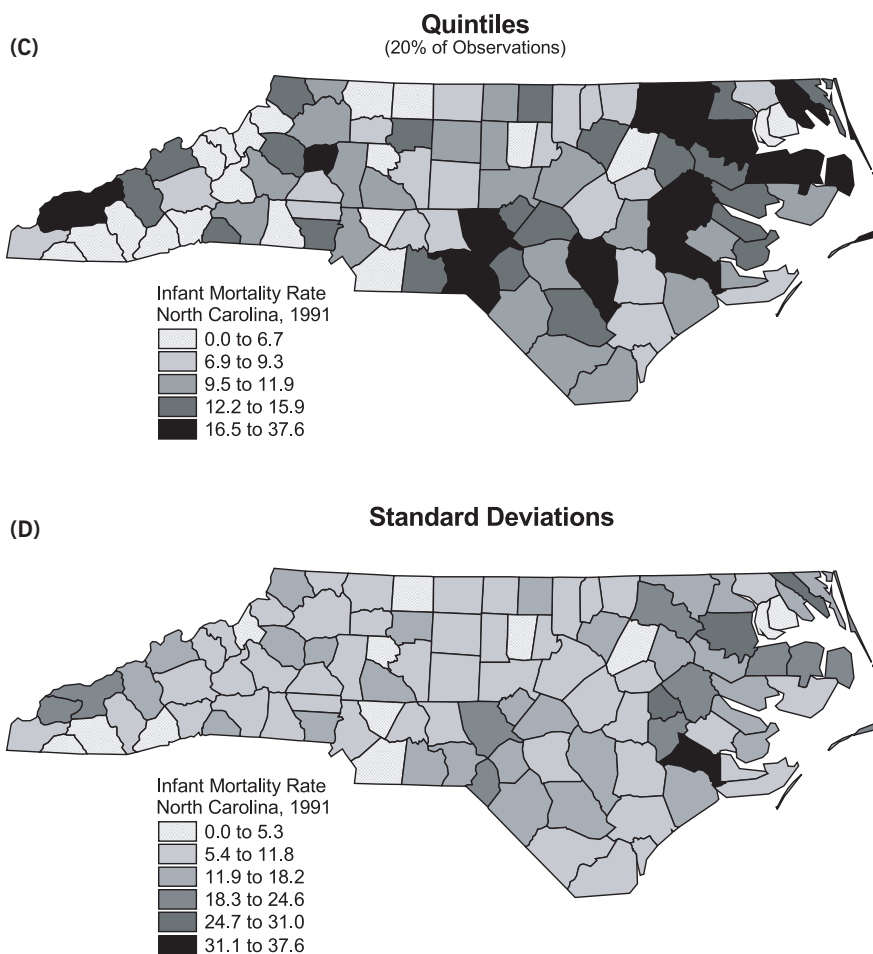


FIGURE 3.5. (*cont.*)

the East Coast into fractions of an inch. Especially when the disease or health service concern is targeted at a subgroup of the population—whether this subgroup is defined by age and life stage, ethnicity, or behavioral activity—it makes better sense to construct a map in which the size of the units is proportional to population instead of land area. Such demographic base maps are types of cartograms.

Demographic base maps are used much less often than they should be because they are time-consuming to construct and often difficult for people who are unfamiliar with them to interpret. Figure 3.6 shows the infant mortality of Figure 3.5 mapped on a demographic base. Because high rates often apply to small populations that are less accessible or economically developed, a population base often portrays conditions very differently. On a global map of infant mortality mapped on a demographic base of births, for example, it is clear that many more infants are dying in South Asia than in Africa.

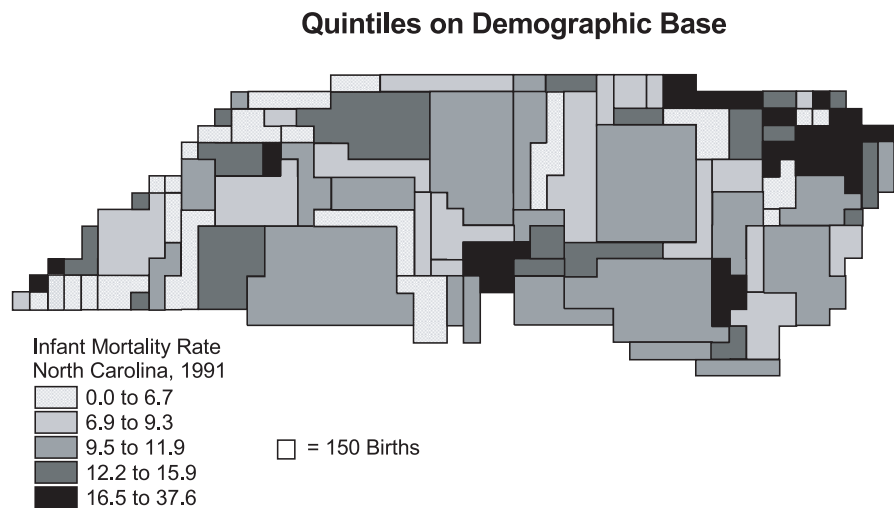


FIGURE 3.6. Demographic base map of infant mortality in North Carolina, 1991. Quintiles of data are mapped on a cartogram of county size proportional to county births.

GEOGRAPHIC INFORMATION SYSTEMS

A GIS is an information system specifically created for maps. GISs are much more than computerized cartography systems, since they facilitate the manipulation of spatial data in many different ways. They are automated systems for the capture, storage, retrieval, analysis, and display of spatial data. Spatial data are stored in either *vector* (i.e., points, lines, and polygons) or *raster* (i.e., cell-based) data structures, allowing them to be integrated on the basis of common geography. In Figure 3.7a each raster cell has a value that has a corresponding land cover class represented as vector polygons in Figure 3.7b.

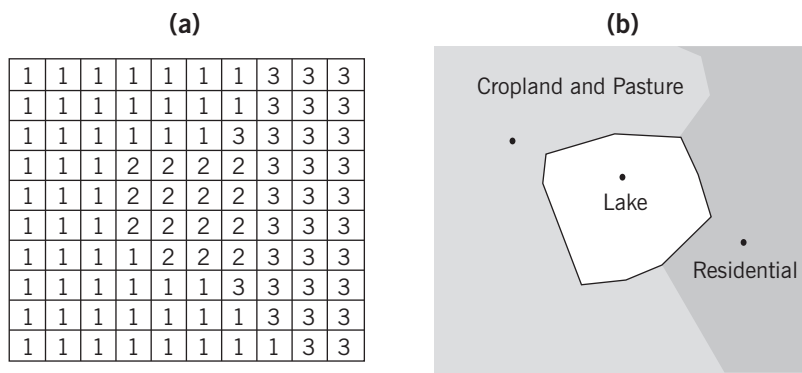


FIGURE 3.7. (a) Raster data structure. (b) Vector data structure. From the USGS (erg.usgs.gov/isb/pubs/gis_poster).

The following discussion uses Figure 3.8 (see also Table 3.1), a map of the imaginary town of Search, as its reference to describe how a vector GIS system is structured. The location of the lake and its shore, the ground cover of Stamp Forest and Fonaroff Swamp, much of the road system, and the altitude of Haggett Eminence would be determined by using GPS or satellite imagery. One geocoded census block (from TIGER files) has been included as an inset that enlarges the shaded area and provides more detail. The location of the hospital has been digitized, as well as the locations of smelters and smokestacks, and those of cases of an unknown disease.

One of the main sources of geocoded health data in the United States is based on block-level street files that are part of the census. Block 103 is bounded by four streets. Each of the corners is treated as a node. The first node—the corner of Howe and Shannon Streets—has a precise location in space, which the grid determines is 2.0 on the X-axis and 8.0 on the Y-axis. The node at the corner of Howe Street and Hunter Boulevard defines the other end of the line segment (which will have its own arc identification number), and has its own location in space at $X = 3$, $Y = 8$. If one continues “chaining” from node 2 to 3, node 3 to 4, and node 4 back to 1, a polygon that is block 103 will be defined. This specific locational information for the nodes, line segments (or arcs), and polygons, however, also comes associated with certain topological attributes. Table 3.1 shows the so-called “topology tables” (including the from-to nodes and the left-right polygon tables) that give lines direction. The software “knows” that polygon 102 is across the line segment from polygon 103. If house numbers have been associated with that line segment—say, odd numbers 7 through 15 on the left, and even numbers 8 through 16 on the right—then the home addresses of cases in the cancer registry can be allocated to the separate blocks. Furthermore, if a line segment of Hunter Boulevard from Prothero Street to McGlashan Street is blocked by a reported fallen tree, the GIS can be used to determine the minimum route (distance along line segments) necessary to route an ambulance around the blockage to the hospital. The GPS navigation instruments used in cars routinely calculate minimum distances, and their rerouting functions can easily determine alternate routes in such situations.

Separate files are created of attributes associated with these points, lines, and areas (polygons), and it is the job of the GIS to manage and interrelate all these. Thus there are files that give each line segment its context of other line segments and polygons, so that an ambulance can be routed. Whether the polygons are census tracts, portions of marketing areas, or congressional districts, all the attributes a researcher desires and can specify are linked to it. Block 103, for example, is a part of a census tract. Individual household data are not released, but information for the block may include median family income, numbers of single female-headed households, numbers of white children ages 5–9, room density, age of housing, percentage of households with air conditioning, proportion of the population that has graduated from high school, and numbers speaking Spanish at home. Some information may be confidential at the block level, especially by sex-age subdivisions of the population, which can get very small (and so inappropriate to use for mapping incidence rates); however, because the block is part of a tract, the data for the larger census tract can be attached to it.

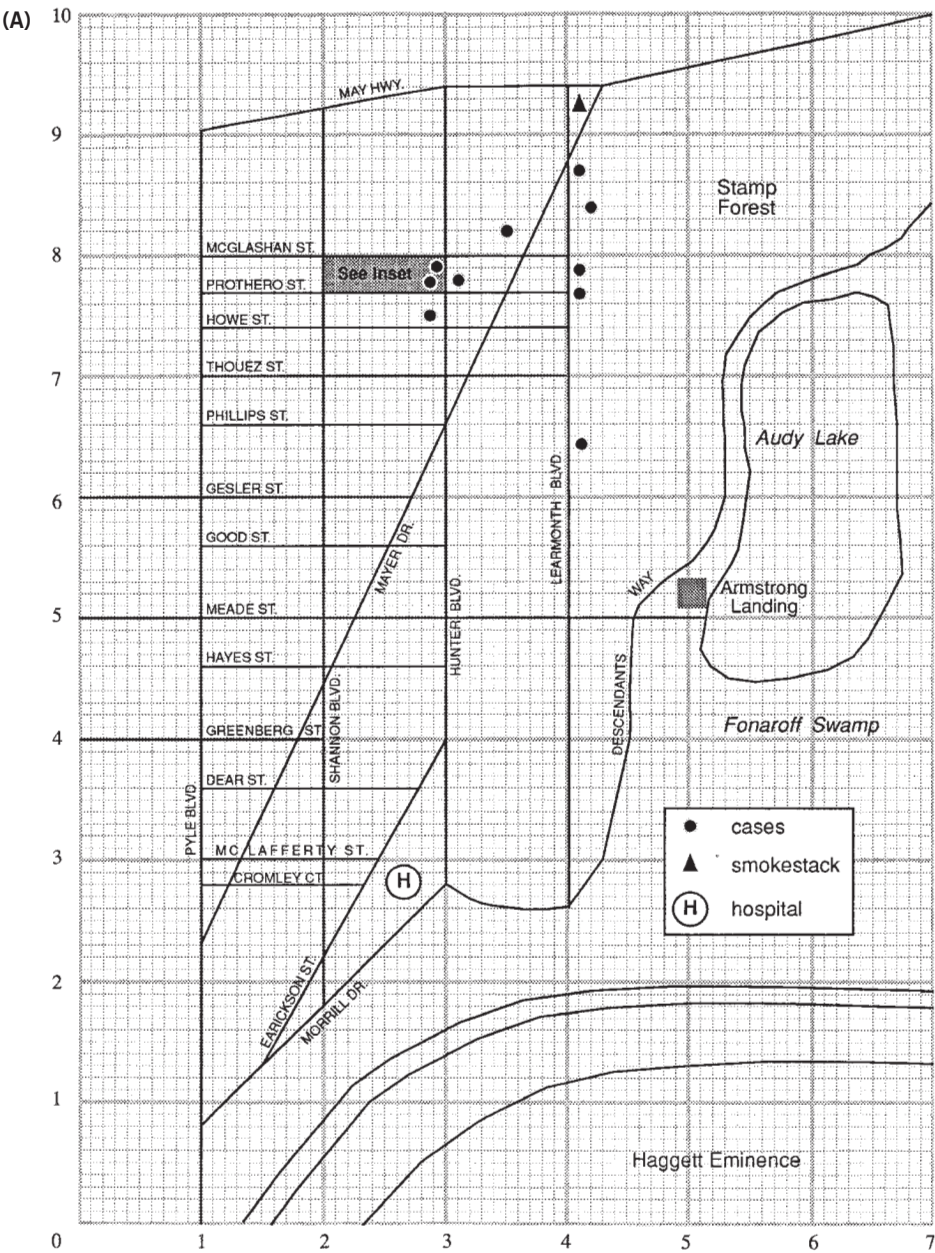


FIGURE 3.8. (A) The town of Search. (B) Locations relative to tracts and blocks.

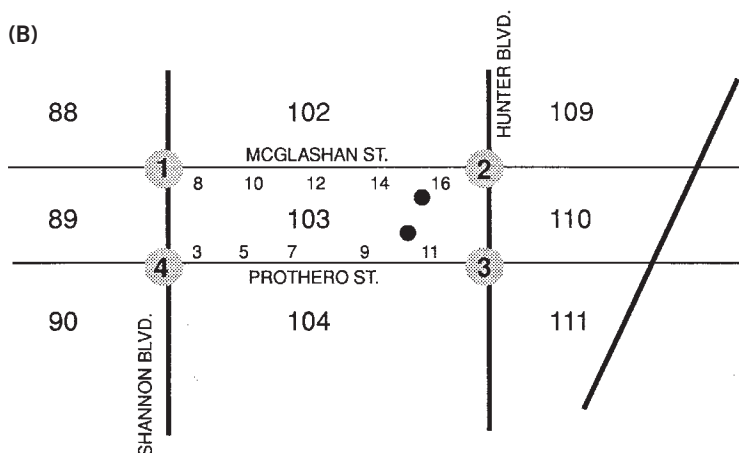


FIGURE 3.8. (cont.)

Knowing where the smelter is located, the GIS can do distance searches to the cases that have been allocated to blocks. Or “buffers” of specified width can be established along roads and cases in them counted. If the municipal water pipelines have been digitized, age, material, or other characteristics can be matched to the line segments of the pipes, and the movement and siting of water along the segments and at various distances to houses can be modeled. If there is more than one source of water for the people of Search, the separate water delivery sheds can be geocoded and then overlain as areas, and the population (by age) and cases in the various watersheds can be counted and compared for risk. People attending the hospital for related reasons can be allocated by their addresses to the blocks where they live; iso-lines can be constructed of the proportion of people in blocks attending the hospital; and service areas with different distance decay functions can thus be determined.

Spatial data integration is one of the fundamental reasons for using a GIS. A nonspatial information system stores data via a common identifier, called a *key* in relational database management terminology. An example of a common identifier is a student number by which a university may keep track of each student. Records such as grades are maintained in academic offices, and financial information such as whether a student has paid his or her tuition is stored in another office. Information from the different databases can be linked via the database key (the student

TABLE 3.1. Topology of Block 103 of the Town of Search

Coordinate			Coordinate			Polygon to	
From node	X	Y	To node	X	Y	Right	Left
1	2.0	8.0	2	3.0	8.0	103	102
2	3.0	8.0	3	3.0	7.7	103	110
3	3.0	7.7	4	2.0	7.7	103	104
4	2.0	7.7	1	2.0	8.0	103	89

number, in this case) for different purposes at different times, such as before graduation. A student who has gotten high enough grades in the right classes and paid his or her tuition bills will be given a diploma, and this information requires integration of databases from the different offices before the student graduates.

Now consider trying to link two different databases together that have no common identifier. People live in different communities with varying social and biophysical environments. If we have information about a person's disease status and address, and we want to link this information to the presence of a septic tank or sewerage system around the person's residence, then we can integrate these two types of information based on common geography. In other words, we can integrate two different maps: one of the locations of homes and the disease status of people living in them, and the other of sewerage and septic systems. Having all this information for many people will allow us to investigate whether there are associations between sanitation practices and disease distributions.

The other main purpose of using a GIS is to facilitate spatial analysis. Simple spatial calculations can be automated, such as distance measures to a source of contamination (e.g., a toxic waste incinerator). However, more complex techniques can also be employed that take into account wind currents from such a plant at different times of day. Disease clusters can be located, and hypotheses about their causes can be formulated; given the right ancillary spatial data layers, these hypotheses can also be tested. Although spatial data integration and analysis are the main reasons for the rapid expansion of GIS technology, during the last two decades spatial visualization has also blossomed.

As described in the previous sections of this chapter, spatial visualization of health and disease is not new. However, with a GIS it is easy to create a basic disease map, and digital spatial databases are more readily available to make them. Disease maps can be simple representations of univariate distributions of the prevalence of one disease. Figure 3.9 shows two zip-code-level morbidity ratio maps, standardized for age and gender, of anxiety-related *International Classification of Diseases*, ninth revision (ICD-9) diagnoses before and after the September 11, 2001, terrorist attacks in New York City (DiMaggio, Galea, & Emch, in press). These maps and the associated research project show that distance from the World Trade Center site in the postattack time period was associated with increased risk of anxiety-related diagnoses. In fact, each 2-mile increment in distance closer to the World Trade Center site was associated with a 7% increase in anxiety-related diagnoses in the population and no similar association was found during a similar time period in the year prior to the attacks. The map reader can develop hypotheses about the reasons for the spatial distribution of one of these maps or for changes between the two dates. (ICD-9 is used to code and classify morbidity data from inpatient and outpatient records, physicians' offices, and National Center for Health Statistics surveys.)

Disease maps can also summarize distributions. The geometric center of a group of points is the mean center that might be used to describe the center of an epidemic. A *centroid* is the geometric center of a polygon that is represented as a point. These measures of central tendency are easily calculated in GIS software, such as ESRI's ArcGIS package. A GIS can also be used to calculate clusters of disease (e.g., breast

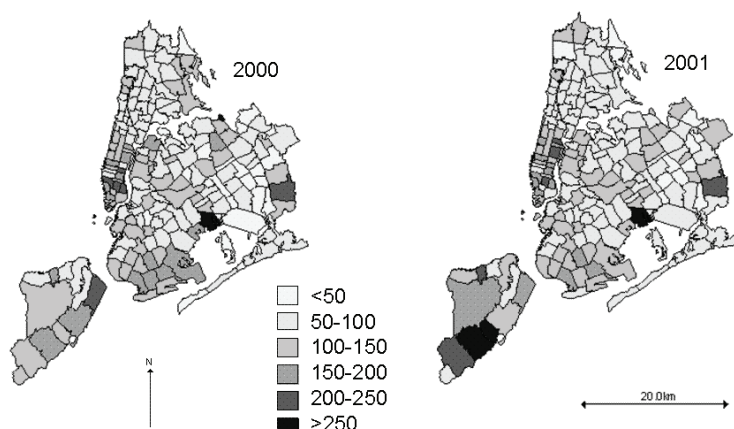


FIGURE 3.9. Fitted standardized morbidity ratios, anxiety-related diagnoses, New York City, September 11–December 30, 2000 and 2001. New York City outpatient Medicaid data. From DiMaggio, Galea, and Emch (in press).

cancer clustering on Long Island). The idea behind identifying a cluster is to determine where there are more cases in an area than would be expected due to chance. Many types of methods can be used to identify spatial clusters; some of these are available in commercial GIS software such as ArcGIS (e.g., Anselin local Moran's I , Getis–Ord G_i^*), whereas others require specialized spatial statistical software (e.g., the spatial scan statistic implemented in SaTScan software; www.satscan.org).

Overlay analysis is a fundamental method in a GIS and simply means that the user is mathematically integrating multiple map layers together. Figure 3.10 shows how two layers are integrated through raster overlay: The spatial intersection of the two map layers results in a composite layer. For example, if the cells in layer 1 represent areas near toxic waste dumps, and the cells in layer 2 represent childhood leukemia cases, then the composite can pinpoint the locations where there are both. For example, if you have two layers—one of disease outcomes and another of soils—the GIS operator can determine whether the disease outcome is more common in areas with a particular type of soils. There are numerous mathematical overlay methods (e.g., intersection, union), and these are discussed in further detail in Chapter 12.

As mentioned earlier, a GIS is a spatial data integration tool, and the data it can integrate are now available from a variety of sources—including satellite images, the GPS, geocoding address data, and analog maps that have been digitized. Satellite images are complex digital photographs taken from satellite sensors orbiting the earth. However, satellite sensors do not take these digital photographs only in the visible part of the spectrum that humans can see; they can also sense electromagnetic wavelengths (see Figure 6.1) that humans cannot detect, such as near-infrared wavelengths. Because different surface features of the earth reflect electromagnetic energy differently, these satellite images can be used to distinguish between them and therefore to map them. Satellite imagery is made up of *pixels* (rasters), which represent small square areas of the earth's surface and define the spatial resolution

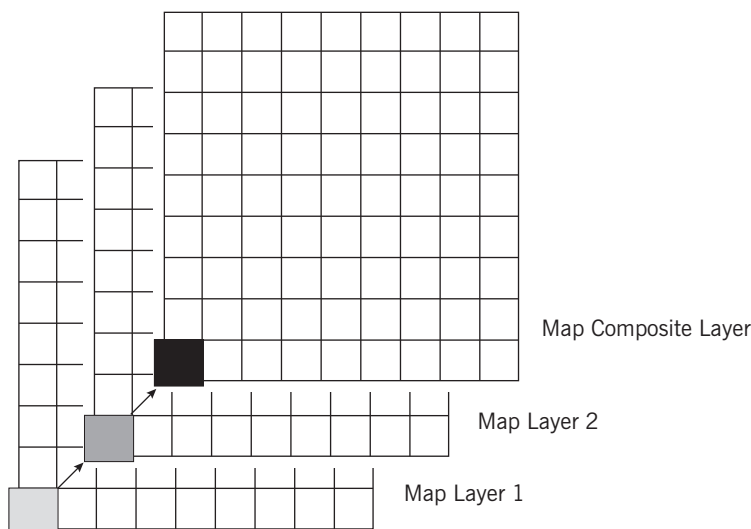


FIGURE 3.10. Overlay analysis in a raster GIS.

of the satellite. Figure 3.11 is a Landsat ETM+ satellite image of a study area where there is a large health and demographic surveillance system for health research. The imagery has a spatial resolution of 30 meters, which means that each pixel represents 30 meters on the ground. This image shows information from electromagnetic radiation that has been sensed in the near-infrared and visible parts of the spectrum.

Other types of resolutions that define the quality of satellite imagery are the temporal and spectral resolutions. The *temporal resolution* is how often the satellite images a particular area of the earth's surface; the *spectral resolution* is the detail of the electromagnetic energy measured, and therefore represents how well the imagery can distinguish between earth surface features. Satellite imagery is a raster GIS data source.

The GPS is a satellite-based system made up of a network of 24 satellites placed into orbit by the U.S. government. A GPS receiver takes this information and uses *triangulation* (more precisely, something similar to triangulation, called *trilateration*) to calculate the receiver's exact location. A GPS can be used to create a GIS map of points or combinations of points (i.e., lines or polygons). The GPS is therefore a vector GIS input device. Inexpensive GPS receivers are accurate to approximately 10–15 meters, and more expensive receivers can be accurate to the centimeter. The GPS has revolutionized GIS database creation by making it easy and inexpensive.

Another very important GIS data source is geocoded address data. *Geocoding* is the process of relating an address to a geographic location (such as latitude–longitude coordinates) or geographic area (such as census tract, block group, block, or zip code), as discussed for Figure 3.8. There are now several Web-based geocoding services that can be used to geocode a small number of addresses. For those with a large number of addresses to match, options include geocoding the addresses by

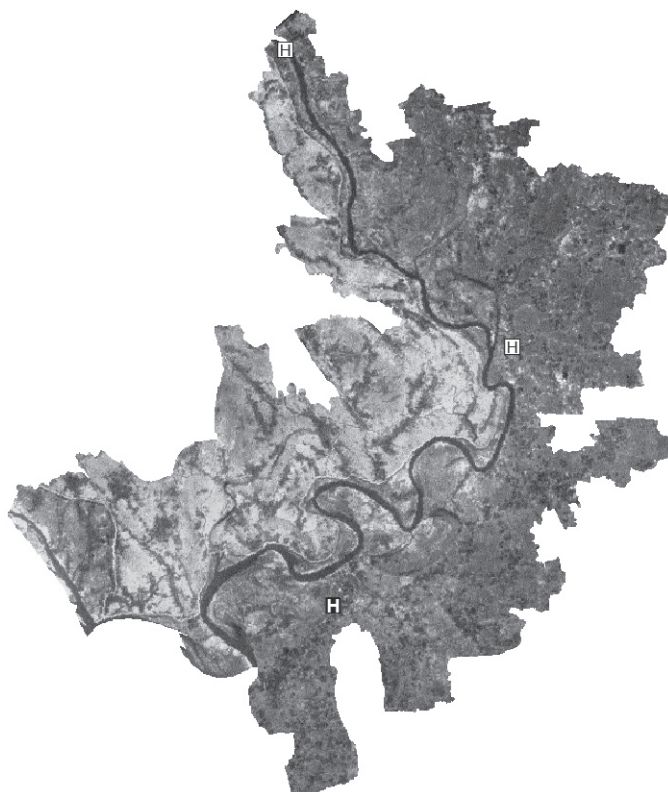


FIGURE 3.11. Landsat ETM+ image of Matlab, Bangladesh, with hospital and health center locations.

using GIS software, or paying a commercial vendor to do the geocoding. In developed countries such as the United States, many health and disease data include the addresses of individuals, and through geocoding a locational component can be added by geocoding these addresses.

The last important GIS data source to be discussed in this chapter consists of legacy data that were originally collected as analog data (i.e., on paper maps) and subsequently digitized. Almost the entire history of cartography occurred before the computer age, and disease maps were originally created as paper maps. Many of these paper maps represent important sources of data on spatial distributions of health and disease. Thus many such maps have been converted to GIS files by using one of a couple of available technologies. Digitizing boards are now an obsolete technology, but much of the digitization of paper maps was done by using these boards for the conversion process. Digitizing boards were drafting tables with electronic grids inside them; as the digitizer dragged a mouse along the points, lines, and polygons of the map features, these were electronically captured. The modern technology for digitizing consists of a large-format scanner and subsequent “heads-up digitizing.” A large-format scanner is a raster input device through which the

user feeds a map and it captures a digital picture, which is displayed on a computer screen. The digitizer then drags the computer mouse over the screen and identifies the points, lines, and polygons that make up the map, creating a vector GIS map.

DISEASE MAPS ON THE WEB

The GIS, in conjunction with the World Wide Web, has made it possible to display disease maps inexpensively and sometimes interactively to a wide audience. For instance, the World Health Organization (WHO) Communicable Disease Global Atlas (apps.who.int/globalatlas) allows users to map infectious diseases at country, regional, and global levels, using data and mapping software available on their server via the Web. Figure 3.12 displays a map of reported global cholera cases by country for 2007, created on the WHO Global Atlas website. The system allows users to browse, view, query, and search the contents of the WHO's communicable disease global database, and to output data in reports, charts, and maps. It has a user-friendly mapping interface that can be used to select geographic areas of interest and create maps of disease outbreaks; the locations of health facilities, schools, and roads; and other geographic features. The website also provides access to static maps of infectious diseases. Other data that can be interactively displayed include demographic variables, socioeconomic conditions, and environmental factors; this capability allows the user to explore the broad range of determinants that influence patterns of infectious disease transmission.

Another example of disease-mapping capabilities available on the Web is the USGS Disease Maps site (diseasemaps.usgs.gov). This site provides close to real-time statistics on several vector-borne diseases in the United States, including West Nile virus (WNV). It provides a series of county-level maps that display confirmed WNV cases in humans, birds, mosquitoes, and animals throughout the United States. For example, the grey counties in Figure 3.13 are those that had at least one confirmed human WNV case in 2008. The online map allows the user to zoom in on individual states displaying the numbers of confirmed cases in a particular county. The USGS Disease Maps site also provides similar information for other vector-borne diseases, including St. Louis encephalitis, Eastern equine encephalitis, Western equine encephalitis, La Crosse encephalitis, and Powassan virus. This site and others like it are valuable tools to communicate information to the public.

Another WHO website (www.who.int/apoc/cdti/remo/en/index.html) describes a GIS-based tool that was developed by the WHO to determine the geographical distribution of onchocerciasis and to identify which communities to treat. The tool, called *rapid epidemiological mapping of onchocerciasis* (REMO), identifies communities at high risk for onchocerciasis so that interventions can be carried out. By 2007 REMO had been largely completed in West Africa, and this facilitated the creation of national onchocerciasis control plans. The REMO website includes a series of maps that shows the disease distribution in different countries. It also describes the REMO methodology.



FIGURE 3.12. Global cholera distribution in 2007. From the WHO Communicable Disease Global Atlas website (apps.who.int/globalatlas).

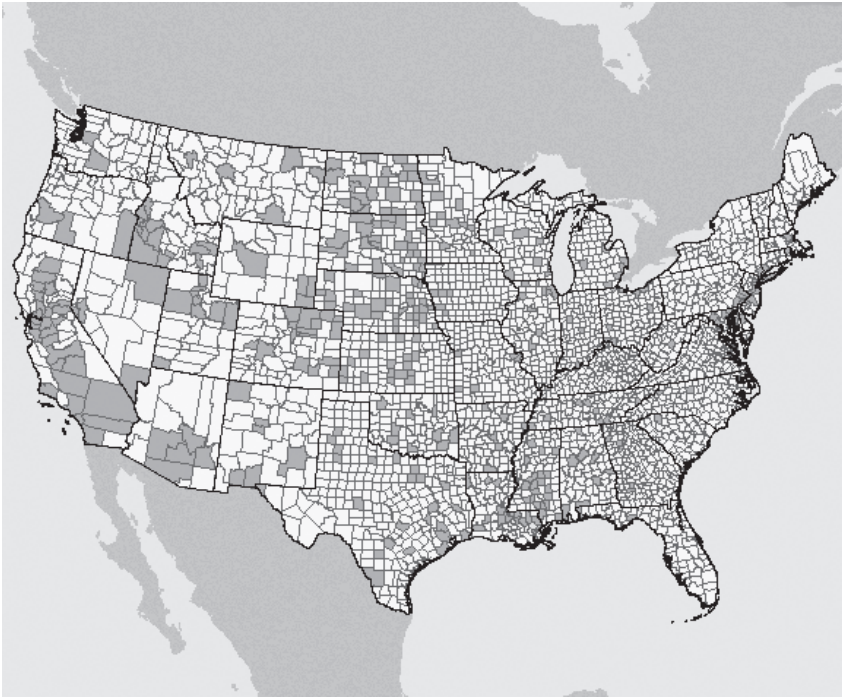


FIGURE 3.13. Human West Nile virus (WNV) cases in 2008. From the USGS Disease Maps website (diseasemaps.usgs.gov).

CONCLUSION

Geographic methodology has been revolutionized by the information now available and the speed and complexity with which it can be manipulated in GIS. Software has been developed to make GIS available on a point-and-click basis in windows and menus. Uses are spreading rapidly into a variety of disciplines, including epidemiology and other areas of public health. It is now regularly used internationally by the WHO and other organizations for surveillance of vectored diseases. Annual GIS and public health conferences now have sessions that focus on spatial applications in health, with contributions by government epidemiologists and statistical analysts (from the CDC and the National Center for Health Statistics in the United States), academics, and applied analysts. Such sessions facilitate planning and health services across disciplines.

The present state is one of euphoria over possibilities. The power of a GIS lies in the quality of the data and the innovation of the operator. Just as cartography is both an art and a science, so is the spatial analysis performed in a GIS. An expert in GIS is someone who not only has an understanding of the breadth of spatial-analytical tools available, but also can put them together in innovative ways to answer spatial questions. These types of questions are further explored in Chapter 9 and Chapter 12, which not only cover some of the more advanced spatial-

analytical methods, but also discuss some of the challenges and potential misuses of this technology.

REFERENCES

- Berghaus, H. (1852). *Physikalischer atlas*. Gotha, Germany: J. Perthes.
- Burrough, P. A., & McDonnell, R. A. (1998). *Principles of geographical information systems*. Oxford, UK: Oxford University Press.
- Clarke, K. C. (1995). *Analytical and computer cartography* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Clarke, K. C. (1997). *Getting started with GIS*. Upper Saddle River, NJ: Prentice Hall.
- Cliff, A. D., & Haggett, P. (1988). *Atlas of disease distributions: Analytical approaches to epidemiological data*. Oxford, UK: Blackwell.
- Cliff, A. D., Haggett, P., & Ord, K. (1987). *Spatial aspects of influenza epidemics*. London: Routledge & Kegan Paul.
- Cliff, A., Haggett, P., & Smallman-Raynor, M. (2004). *World atlas of epidemic diseases*. London: Arnold.
- DiMaggio, C., Galea, S., & Emch, M. (in press). Spatial proximity and the risk of psychopathology after a terrorist attack. *Psychiatry Research*.
- Gilbert, E. W. (1958). Pioneer maps of health and disease in England. *Geographical Journal*, 124, 172–183.
- Gould, P. (1993). *The slow plague: A geography of the AIDS pandemic*. Oxford, UK: Blackwell.
- Haggett, P., Cliff, A., & Smallman-Raynor, M. (1993). *Measles: An historical geography of a major human viral disease from global expansion to local retreat, 1840–1990*. Oxford, UK: Blackwell.
- Haggett, P., Cliff, A., & Smallman-Raynor, M. (1998). *Deciphering global epidemics: Analytical approaches to the disease records of world cities, 1888–1912*. Cambridge, UK: Cambridge University Press.
- Haggett, P., Cliff, A., & Smallman-Raynor, M. (2000). *Island epidemics*. Oxford, UK: Oxford University Press.
- Howe, G. M. (1972). *Man, environment, and disease in Britain*. New York: Barnes & Noble.
- Koch, T. (2005). *Cartographies of disease: Maps, mapping, and medicine*. Redlands, CA: ESRI Press.
- May, J. M. (1950). Map of the world distribution of poliomyelitis. *Geographical Review*, 40, 646–648.
- May, J. M. (1951a). Map of the world distribution of cholera. *Geographical Review*, 41, 272.
- May, J. M. (1951b). Map of the world distribution of malaria vectors. *Geographical Review*, 41, 638–639.
- May, J. M. (1952a). Map of the world distribution of dengue and yellow fever. *Geographical Review*, 42, 282–286.
- May, J. M. (1952b). Map of the world distribution of helminthiasis. *Geographical Review*, 42, 98–101.
- May, J. M. (1952c). Map of the world distribution of plague. *Geographical Review*, 42, 628–630.
- May, J. M. (1953a). Map of the world distribution of leprosy. *Geographical Review*, 43, 89–90.
- May, J. M. (1953b). The mapping of human starvation. *Geographical Review*, 43, 253–255.
- May, J. M. (1954a). Map of the world distribution of leishmaniasis. *Geographical Review*, 44, 583–584.
- May, J. M. (1954b). Map of the world distribution of rickettsial diseases. *Geographical Review*, 44, 133–136.
- May, J. M. (1954c). Map of the world distribution of some viral encephalitides. *Geographical Review*, 44, 408–410.
- May, J. M. (1958). *The ecology of human disease*. New York: MD Publications.

- National Center for Health Statistics. (1997). *Atlas of United States mortality*. Retrieved from www.cdc.gov/nchs/products/pubs/pubd/other/atlas/atlas.htm
- Openshaw, S., Craft, A. W., Charlton, M., & Birch, J. M. (1988). Investigation of leukemia clusters by use of a geographical analysis machine. *Lancet*, *i*, 272–273.
- Rodenwaldt, E., & Juszat, H. J. (Eds.). (1952–1961). *Welt-Seuchen atlas* (Vols. 1–3). Hamburg, Germany: Falk.
- Smallman-Raynor, M., Cliff, A., & Haggett, P. (1992). *London international atlas of AIDS*. Oxford, UK: Blackwell.
- Snow, J. (1855). *On the mode of communication of cholera*. London: John Churchill, New Burlington Street.

FURTHER READING

- Ali, M., Emch, M., von Seidlein, L., Yunus, M., Sack, D. A., Holmgren, J., et al. (2005). Herd immunity conferred by killed oral cholera vaccines in Bangladesh. *Lancet*, *366*, 44–49.
- Ali, M., Emch, M., Yunus, M., Sack, D., Lopez, A. L., Holmgren, J., et al. (2008). Vaccination of adult women against cholera protects infants and young children in rural Bangladesh. *Pediatric Infectious Disease Journal*, *27*(1), 33–37.
- Ali, M., Emch, M. E., Tofail, M., & Baqui, A. H. (2001). Implications of health care provision on acute lower respiratory infection mortality in Bangladeshi children. *Social Science and Medicine*, *52*(2), 267–277.
- Ali, M., Goovaerts, P., Nazia, N., Haq, M. Z., Yunus, M., & Emch, M. (2006). Application of Poisson kriging to the mapping of cholera and dysentery incidence in an endemic area of Bangladesh. *International Journal of Health Geographics*, *5*(45), 1–11.
- Barnett, R., Pearce, J., & Moon, G. (2005). Does social inequality matter?: Changing ethnic socio-economic disparities and Maori smoking in New Zealand, 1981–1996. *Social Science and Medicine*, *60*(7), 1515–1526.
- Blackburn, J., McNyset, K., Curtis, A., & Hugh-Jones, M. (2007). Modeling the geographic distribution of *Bacillus anthracis*, the causative agent of anthrax disease, for the contiguous United States using predictive ecological niche modeling. *American Journal of Tropical Medicine and Hygiene*, *77*(6) 1103–1110.
- Clarke, K. C., McLafferty, S. L., & Tempalski, B. J. (1996). On epidemiology and geographic information systems: A review and discussion of future directions. *Emerging Infectious Diseases*, *2*, 85–92.
- Cleek, R. K. (1979). Cancer and the environment: The effect of scale. *Social Science and Medicine*, *13D*, 241–247.
- Comrie, A. C. (2005). Climate factors influencing coccidioidomycosis seasonality and outbreaks. *Environmental Health Perspectives*, *113*, 688–692.
- Curtis, A. (1999). Using a spatial filter and a geographic information system to improve rabies surveillance data. *Emerging Infectious Diseases*, *5*(5), 603–606.
- Curtis, A., Mills, J. W., & Blackburn, J. (2007). A spatial variant of the basic reproduction number for the New Orleans yellow fever epidemic of 1878. *Professional Geographer*, *59*(4), 492–502.
- Emch, M., Ali, M., Park, J. K., Yunus, M., Sack, D., & Clemens, J. D. (2006). Relationship between neighborhood-level killed oral cholera vaccine coverage and protective efficacy: Evidence for herd immunity. *International Journal of Epidemiology*, *35*, 1044–1050.
- Emch, M., Ali, M., & Yunus, M. (2008). Risk areas and neighborhood-level risk factors for *Shigella dysenteriae* 1 and *Shigella flexneri*: Implications for vaccine development. *Health and Place*, *14*, 96–105.
- Emch, M., Ali, M., Yunus, M., Sack, D., Acosta, C., & Clemens, J. D. (2007). Efficacy calculation in randomized vaccine trials: Global or local measures? *Health and Place*, *13*, 238–248.

- Emch, M., Feldacker, C., Islam, M. S., & Ali, M. (2008). Seasonality of cholera from 1974 to 2005: A review of global patterns. *International Journal of Health Geographics*, 7(31), 1–33.
- Emch, M., Feldacker, C., Yunus, M., Streatfield, P. K., Thiem, V. D., Canh, D. G., et al. (2008). Local environmental drivers of cholera in Bangladesh and Vietnam. *American Journal of Tropical Medicine and Hygiene*, 78(5), 823–832.
- Emch, M. E. (1999). Diarrheal disease risk in Matlab, Bangladesh. *Social Science and Medicine*, 49, 519–530.
- Emch, M. E. (2000). Relationships between flood control, Kala-azar, and diarrheal disease in Bangladesh. *Environment and Planning A*, 32, 1051–1063.
- Hanchette, C. L. (1999). GIS and decision-making for public health agencies: Childhood lead poisoning and welfare reform. *Journal of Public Health Management and Practice*, 5(4), 41–47.
- Hanchette, C. L., & Swartz, G. G. (1992). Geographic patterns of prostate cancer mortality: Evidence for a protective effect of ultraviolet radiation. *Cancer*, 70(12), 2861–2869.
- Hohl, P., & Mayo, B. (1997). *ArcView GIS exercise book* (2nd ed.). Santa Fe, NM: On-Word Press.
- Howe, G. M. (1970). *National atlas of disease mortality in the United Kingdom* (2nd ed.). London: Nelson.
- Jerrett, M., Burnett, R. T., Kanaroglou, P., Eyles, J., Finkelstein, N., & Giovis, C. (2001). A GIS–environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environment and Planning A*, 33, 955–973.
- Jerrett, M., Burnett, R. T., Ma, R., Pope, C. A., III, Krewski, D., Newbold, K. B., et al. (2005). Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology*, 16(6), 727–736.
- King, P. E. (1979). Problems of spatial analysis in geographical epidemiology. *Social Science and Medicine*, 13D, 249–252.
- Kolivras, K. N., & Comrie, A. C. (2003). Modeling valley fever (coccidioidomycosis) incidence based on climate conditions. *International Journal of Biometeorology*, 47, 87–101.
- Matthews, S. A. (1990). Epidemiology using a GIS: The need for caution. *Computer, Environment and Planning*, 17, 213–221.
- Mayer, J. D. (1981). Problems of spatial analysis in geographical epidemiology. *Social Science and Medicine*, 13D, 249–252.
- Mayer, J. D. (1983). The role of spatial analysis and geographic data in detection of disease causation. *Social Science and Medicine*, 17, 1213–1221.
- McLafferty, S., & Grady, S. (2004). Prenatal care need and access: A GIS analysis. *Journal of Medical Systems*, 28(3), 321–333.
- McLafferty, S., & Grady, S. (2005). Immigration and geographical access to prenatal clinics in Brooklyn, NY: A geographic information systems analysis. *American Journal of Public Health*, 95(4), 638–640.
- Meade, M. S. (1979). Cardiovascular mortality in the southeastern United States: The coastal plain enigma. *Social Science and Medicine*, 13D, 257–265.
- Meade, M. S. (1983). Cardiovascular disease in Savannah, Georgia. In N. D. McGlashan & J. R. Blunden (Eds.), *Geographical aspects of health* (pp. 175–196). New York: Academic Press.
- Moon, G., & Barnett, J. R. (2003). Spatial scale and the geography of tobacco smoking in New Zealand: A multilevel perspective. *New Zealand Geographer*, 59(2), 6–15.
- Pearce, J., Hiscock, R., Moon, G., & Barnett, R. (2009). The neighbourhood effects of geographical access to tobacco retailers on individual smoking behaviour. *Journal of Epidemiology and Community Health*, 63(1), 69–77.
- Pope, K. O., Rejmankova, E., Savage, H. M., Arredondo-Jimenez, J. I., Rodriguez, M. H., & Roberts, D. R. (1994). Remote sensing of tropical wetlands for malaria control in Chiapas, Mexico. *Ecological Applications*, 4, 81–90.
- Rushton, G., Armstrong, M. P., Gittler, J., Greene, B. R., Pavlik, C., West, M. M., et al. (2006). Geo-coding in cancer research: A review. *American Journal of Preventive Medicine*, 30(2, Suppl. 1), S16–S24.
- Rushton, G., Peleg, I., Banerjee, A., Smith, G., & West, M. (2004). Analyzing geographic patterns

- of disease incidence: Rates of late stage colorectal cancer in Iowa. *Journal of Medical Systems*, 28(3), 223–236.
- Shisematsu, I. (1981). *National atlas of major disease mortalities in Japan*. Tokyo: Japan Health Promotion Federation.
- Vine, M., Degnan, D., & Hanchette, C. (1997). Geographic information systems: Their use in environmental epidemiologic research. *Environmental Health Perspectives*, 105, 598–605.
- Vine, M. F., Stein, L., Weigle, K., Schroader, J., Degnan, D., Tse, C., et al. (2000). Effects on the immune system associated with living near a pesticide dump site. *Environmental Health Perspectives*, 108(12), 1113–1124.
- Washino, R. K., & Wood, B. J. (1994). Application of remote sensing to arthropod vector surveillance and control. *American Journal of Tropical Medicine and Hygiene*, 50(6, Suppl.), 134–144.
- Ziegenfus, R. C., & Gesler, W. M. (1984). Geographical patterns of heart disease in the northeastern United States. *Social Science and Medicine*, 18, 63–72.

CHAPTER 4

Landscape Epidemiology

The landscape that distinguishes a place is a complex expression of physical, biotic, and cultural processes. When one knows how to analyze its elements and patterns, one can usually determine what diseases can occur. This is true at every scale, from the microecology of a house (see Plate 3) and its backyard to the transcontinental migratory paths of birds and the viruses they disseminate. As the world population grows and the world economy changes, landscapes are being altered in ways that increase the risk of disease or enhance protection from it.

This chapter is about the cultural ecology of transmissible disease, or the ways in which regions impart pattern to disease distribution. The broad term *communicable disease*, as used by the Centers for Disease Control and Prevention (CDC), includes all diseases due to a specific infectious agent that is transmitted from an infected host to a susceptible host, whether directly or indirectly through an intermediary vector. We address the directly transmitted agents, contagious ones, in Chapter 2 and mainly in Chapter 10 on diffusion. This chapter is about vectored diseases, which parasitologists and entomologists call *transmissible*. The cultural ecology of vectored diseases needs to address the biogeography of the arthropod vectors and their animal hosts. The Russian parasitologist/geographer Eugene N. Pavlovsky (1966) called this biotic regionalization of vectored diseases *landscape epidemiology*. Today studies of the regionalization of vectored diseases, often using geographic information systems (GIS) and sometimes satellite imagery, are increasing internationally. Most of them use the following concepts, but few of them apply Pavlovsky's term.

Before we take up the elaboration of disease systems and their regionalization, however, basic terminology and system parameters must be established. Then the cultural ecology of several diseases—yellow fever, bubonic plague, malaria, dengue hemorrhagic fever (DHF), schistosomiasis, Lyme disease, and Rocky Moun-

tain spotted fever (RMSF)—are discussed at length. The latest available data on the prevalence of these diseases and more, complete with detailed descriptions and usually maps, are available on the World Health Organization (WHO) website. The use of GIS methodology by WHO scientists in disease surveillance and control is summarized.

Each of these diseases is of major importance not only to human health, but also to the advance of scientific understanding about the nature and occurrence of different types of transmission systems. There are many “models” available in various disciplines for predicting secondary infections and epidemic size, or for analyzing niche competition among species. The importance of human agency (i.e., cultural behavior) in creating the conditions with which people come in contact, modifying them, exposing susceptible individuals to them, and preventing and treating different disease outcomes is almost always ignored. In Chapter 5 the emergence of new diseases and movement of people is related to landscape change. In this chapter we want to understand these vectored disease systems as a whole. We look at why these disease systems occur where they do; how they are relocated, contained, eradicated, or maintained; and how cultural buffers can be used for intervention or prevention. The use of landscape epidemiology allows us to integrate and examine the disease expression in place.

REGIONS

There are at least four types of regions involved in the study of landscape epidemiology. The first is biotic. Climate, altitude, and latitude combine to create broad biotic regions, *biomes*, with predictable locations (see Chapter 6, Vignette 6.3). These are themselves compounded of other regions, of course—climatic, geomorphic, soils, and glaciers, to name a few—and include not only plants and animals, but also particular types of insects and microbes. If the biome’s natural fauna and flora are displaced through land development, their replacements tend to be crops and domestic animals also distinctive to that biome, even when historically not that cultural region.

The second type of region is also biotic: *realms of evolution*. Because oceans, deserts, and great mountain ranges formed barriers to the exchange of genetic information as the continents drifted apart, complexes of plants and animals followed separate paths of evolution in South America, in Africa south of the Sahara, in Asia south of the Himalayas and the Yangzi (Yangtze) River, in Australia, and in the remainder of Eurasia (Palearctic) and North America (Nearctic). These realms of evolution include distinct microbes and insects, as well as the more visible mammals and birds (e.g., kangaroos or giraffes) that popularly demarcate the regions. As trade and political empires established contacts beyond their culture realms, the biotas of the evolutionary realms were exchanged. Deliberately, in the case of food crops, and unintentionally, in the case of disease agents, sweeping processes of global readaptation were set in motion. Historians have only relatively recently

paid attention to the health consequences that followed upon connection of these realms (McNeill, 1976). When the maritime and land-based transportation and trade networks of the Roman empire connected with those of the Chinese empire, for example, terrible epidemics swept both worlds as they exchanged smallpox, measles, diphtheria, typhoid, and much more. The best-studied connection of evolutionary realms is that between the Old World of Eurasia and Africa and the New World of the Americas. The so-called “Columbian exchange” involved the two-way exchange of scores of crops. Animals and the communicable diseases that evolved from them and with them—smallpox, measles, diphtheria, whooping cough, tuberculosis, cholera, typhoid, schistosomiasis, river blindness, malaria, yellow fever, and other pestilence with catastrophic consequences for the Native Americans—went one-way to the New World.

The third type of region, *culture realm*, is a broad cultural area delimited by the extent of particular cultural practices and beliefs that largely originated in the primary culture hearths. China and the United States, for example, occupy vast territories with the same latitude and often similar climatic patterns, but because their inhabitants use the land differently and have different settlement forms, house types, and technologies, their landscapes are distinctive. The use of human waste for fertilizer in the Chinese culture realm (which includes Korea, Japan, and Vietnam) and the use of cows for milk in the European and North American cultural spheres are examples of regional patterns associated with health consequences: intestinal parasites and dysentery in the former, and tuberculosis of the bone and lymph system, Q fever, and brucellosis in the latter, as well as nutritional advantages and disadvantages in each.

The fourth type of region, *natural nidus* (focus or cluster), has been defined by Soviet and German geographers but is probably unknown to most of our readers. A natural nidus is a micro-scale region constituted of a living community, among the members of which a disease agent continually circulates, and the habitat conditions necessary to maintain that circulation in the disease system. The more familiar term *focus* is frequently used instead of the unfamiliar term, but it does not carry the connotation of microecology of soil, vegetation, and animal movement as well as *nidus* does. The life cycles and transmission chains of nidal diseases are complex. This chapter describes several examples of different types of systems.

TRANSMISSIBLE DISEASE SYSTEMS

The biological *agent* of a disease is the infectious information that is necessary (but seldom sufficient) to cause it. It is commonly known as the *germ*, *microbe*, or *parasite*. The organism infected by a disease agent is called the *host*. When animals are the ordinary hosts, the disease is known as a *zoonosis*. A disease that often infects both people and animals is an *anthropo-zoonosis*. When animal hosts serve as a continuing source of possible infection for human beings, our anthropocentric name for them is a *reservoir*, and we humans regard ourselves as the hosts (although the animals

are actually the primary hosts). A disease always present in, or *endemic* (literally “in the population”) to, an animal population is said to be *enzootic*, and a disease epidemic (spreading in a population) among animals is said to be *epizootic*. Some diseases are transmitted directly by contact or inhalation. Rabies, for example, is directly transmitted by an infected animal’s biting a susceptible one. Others are transmitted through contaminated food or water, such as the canine tapeworms of hydatid disease, which go through herbivore–carnivore–feces–grass–herbivore cycling. This chapter is mainly concerned with disease agents that are transmitted by arthropods.

An arthropod that transmits a disease agent between hosts, and in which the agent multiplies and often goes through life cycle changes in form, is known as a *biological vector*. *Biological* distinguishes these vectors from flies or inanimate objects that may merely transport the agents mechanically (as vehicles or fomites). Many people speak of vectors as “insects,” but ticks and mites are not, strictly speaking, insects. The more inclusive category is the phylum, Arthropoda (see Vignette 2.1). The term *arbovirus* includes all arthropod-borne viruses. The major vectors are mosquitoes, biting black flies, ticks, mites, sand flies, fleas, and lice, but gnats, midges, and other arthropods are occasionally involved. Most disease agents are strictly limited to transmission by a single species, or at most a genus, of vector. This establishes limits to their distribution because vectors have specific habitat requirements.

Intermediate hosts are organisms that are necessary to some stage of an agent’s life cycle. Hosts in which the agent attains maturity or its sexual life stage are the *primary hosts* or *definitive hosts*; hosts for the agent’s asexual or larval stage are *intermediate hosts*. The fluke that causes schistosomiasis, for example, must alternately infect people and snails. The snails do not transmit the agent to people; they are not vectors, but intermediate hosts that are eaten and die themselves. Nevertheless, it is often convenient to treat intermediate hosts as though they were vectors because the same kinds of population dynamics and intervention strategies are involved. Sometimes this concept is even further enlarged, as when dogs are treated as “vectors” of rabies to people. Such careless usage may nevertheless be helpful for mathematical modeling of a disease system from our anthropocentric viewpoint.

Transmission Chains

The first chain of transmission diagrammed in Figure 4.1 involves direct communication of the agent by contact between people. Animals also communicate contagion (chain 2). It is possible for people to get a zoonotic infection—for example, through contact with skin fungi or consumption of raw milk.

Chain 3 illustrates a vectored human disease that has no animal reservoir. The absence of a reservoir raises a possibility that does not exist for most anthroozoonoses, disease eradication. Although birds, reptiles, and monkeys have their own forms of malaria, for example, human malaria is solely a human disease vectored by *Anopheles* mosquitoes. There was therefore some hope in the 1950s that it could be eradicated. We consider later some of the reasons why it could not.

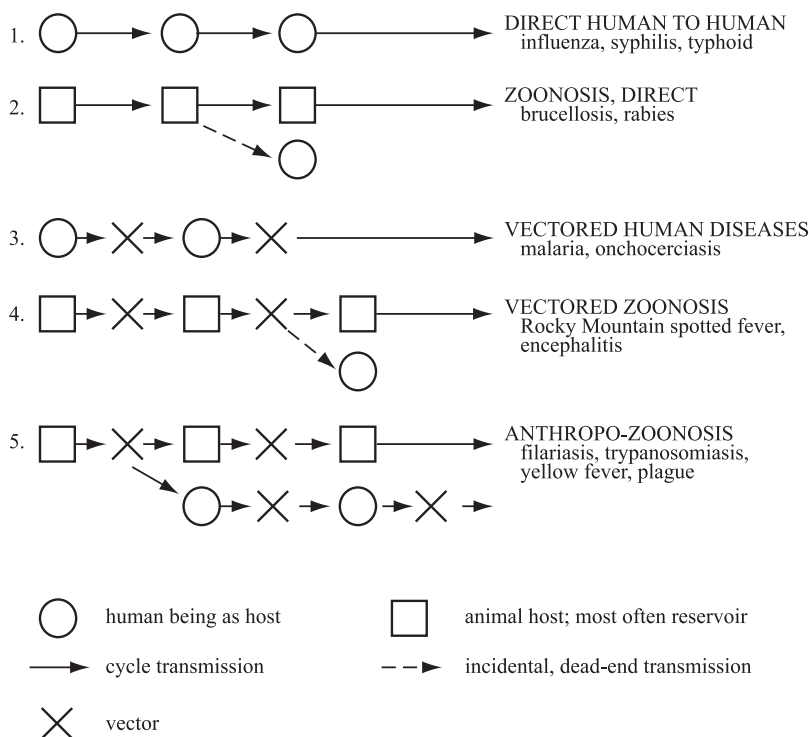


FIGURE 4.1. Chains of disease transmission.

Most vectored diseases involve transmission of agents among animals (chain 4). Many of these cycles are so ancient that agent and host have mutually adapted, and no disease symptoms appear in the infected hosts. Sometimes people are accidentally infected when they intrude on the animal habitat. Usually they develop no disease at all, as they are not compatible hosts. Occasionally they are so susceptible to the rare infection that case mortality rates are high. Whether they don't even get sick or they get so sick they die from the infection, people are usually *dead-end hosts* for zoonotic agents. This means that they cannot give the agent to a vector for transmission to another host. The amount of virus (or bacteria or any agent) circulating in the blood is known as *viremia*. When people are incidentally infected with viral encephalitis by a mosquito that has previously fed on an infectious chicken, for example, they do not develop viremia sufficient to serve as a source of infection for another mosquito imbibing a little of their blood. Encephalitis cannot, therefore, be spread among people by mosquitoes. Even when clusters of cases occur, they result from exposure to the zoonotic chain.

The final chain in the diagram represents the main focus of this chapter: anthro-po-zoonoses, which can be transmitted between animals or between people by vectors, and which sometimes cross between the two systems. The first disease for which this complicated system was discovered was yellow fever.

Yellow Fever

Until the 20th century, yellow fever was one of the most feared diseases in the Western Hemisphere. The step-by-step discovery of its transmission chain (Figure 4.2), with the subsequent recognition of the implications for many other diseases, ranks as one of the milestones of medical science.

Pandemics (international epidemics) of yellow fever swept the Americas in the 18th century. The pandemic of 1793–1804, for example, started in Grenada and spread through the Lesser Antilles and Venezuela, finally striking ports from New Orleans to Boston. During this Napoleonic period, the French suffered losses of more than 30,000 men in their garrisons in Guadeloupe, Martinique, and Santo Domingo. By the late 19th century, yellow fever was endemic from Brazil to Yucatan and throughout the Caribbean; it was seasonally epidemic along the Atlantic seaboard. Then the U.S. Yellow Fever Commission determined that the infectious agent (a virus, now considered the prototype virus of the *Flavivirus* genus, *Flaviviridae*—a family of RNA viruses mostly transmitted by arthropods; it includes dengue) was transmitted by the mosquito *Aedes aegypti*. Using this first-ever proof of arthropod transmission of human disease agents, General Gorgas attacked the mosquito in Cuba, thereby suppressing the endemic focus. Soon after, suppression of yellow fever allowed construction of the Panama Canal. After the 1905 pandemic was stopped in New Orleans by anti-*Ae. aegypti* operations, the newly established Rockefeller

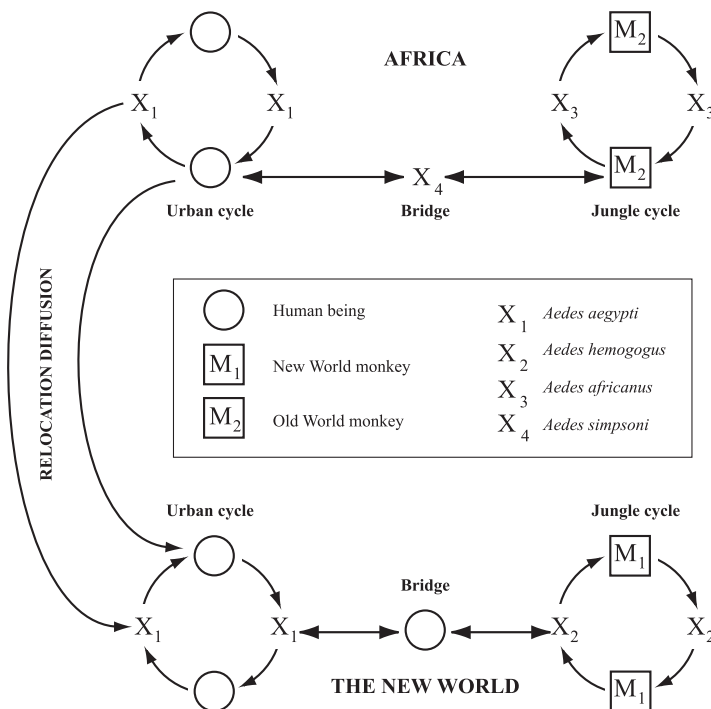


FIGURE 4.2. Yellow fever transmission chain.

Foundation undertook the task of eradicating *Ae. aegypti* throughout the Caribbean. Partly because the mosquito was an introduced species that was only superficially established, the operation was a success. Yellow fever appeared to have been eliminated.

Then came isolated reports of yellow fever—not epidemics—and investigation discovered that the agent was transmitted among monkeys by mosquitoes, especially *Ae. hemagogus* (see Figure 4.2). This astonishing finding was dubbed the “jungle cycle.” Woodcutters and settlers, exposed by their behavior to the monkey-biting jungle mosquito, occasionally contracted the disease; however, as long as *Ae. aegypti* was absent, there was no urban cycle and no epidemic.

In West Africa, whites had been dying of yellow fever in “the white man’s graveyard” for centuries. It was thought, however, that Africans did not get the disease, which was epidemic only in coastal European settlements. Further study revealed that Africans had greater resistance (lower case mortality) to the disease, but that in many parts of Africa the infection was in fact almost universal. That is, studies of blood (serology) showed that antibodies to the virus were very common. The mosquito *Ae. aegypti*, which evolved in Africa, was demonstrated to be transmitting the virus among people in the African towns and villages (so that the “urban cycle” was also the “village cycle”). Further search demonstrated that there was also a jungle cycle: *Ae. africanus* and related mosquitoes were transmitting the virus among monkeys. That yellow fever originated in Africa is indicated by the resistance of the human population, by the ubiquity of *Ae. aegypti* there, by the genetic variation of the virus, and by a further complication to the system. There is a mosquito, *Ae. simpsoni*, that breeds in the curled leaves of banana plants and other vegetation on the fringe of human settlements. It bites people in the villages and monkeys in the forest. Because *Ae. aegypti* is indigenous and exists in many strains fully adapted to various niches (e.g., breeding in tree holes), its eradication seems impossible. Because of *Ae. simpsoni*, which does not have an equivalent in the New World system, the infection that is enzootic in the monkeys (which serve as a reservoir for the disease) can be easily reintroduced to the villages and towns even when the urban transmission chain is broken. The jungle cycle is intimately linked to the village cycle, and yellow fever continues to occur.

The major defense in Africa is the potent vaccine for yellow fever. In tropical America, however, vaccine is used mainly for those occupationally exposed to the jungle cycle. Efforts to prevent the disease have been concentrated on preventing *Ae. aegypti* from becoming reestablished in the cities, which recently seems again to be a losing battle. Because yellow fever does not occur today in the United States, there is little incentive for this country to spend billions of dollars and massive efforts to eradicate *Ae. aegypti*, which is widespread and now deeply ensconced. As a consequence, the mosquito is repeatedly exported, via shipped goods, from the United States to Venezuela and other places in Latin America, where it has been repeatedly eradicated.

Once the intricacies of this vectored disease were understood, the puzzles of many other vectored diseases became solvable. Scientists looked for other arthropods capable of transmitting agents and for other animal reservoirs.

Vector Ecology

Most disease agents require specific hosts in which they live. They are even more restrictive about the cold-blooded arthropods in which they reproduce. Because an agent usually goes through changes in life stages and multiplies within a vector, most arthropods cannot be biological vectors. An inappropriate arthropod finds its throat blocked, its abdomen too heavy for flight, and its life expectancy so shortened that both arthropod and agent die before transmission. Such poor adaptation of a vector to an agent implies that the relationship is of recent origin. For example, part of the evidence that typhus is a relatively recent disease in humans is that the human louse itself dies from the infection. Transmission comes through a bite or another abrasion contaminated with louse feces or crushed-louse fluids, instead of from repeated blood meals, which would occur in a well-adapted system. (Fleas sometimes die, too, from ingesting large amounts of bacteria such as *Yersinia pestis*, which multiply and cause the blood to clot; clot chunks then enter the wound with the next desperate blood meal of the hapless flea.)

A full range of arthropods is involved in disease transmission almost everywhere, but one may identify broad patterns that generally coincide with biomes (see Vignette 6.3). At high altitudes and latitudes, ticks are especially important vectors (Table 4.1). This is because a tick is capable of transmitting rickettsia and viruses *transovarially* (into its eggs), so that the thousands of tick larvae that hatch from an infected tick's eggs are equally infected, through generations. In this way the disease agent is transmitted despite freezing temperatures, which otherwise stop transmission by killing, for example, adult mosquitoes and the agents they carry.

Mosquito-transmitted diseases are most important in warm, humid lands where these insects can live long and reproduce frequently (Table 4.2). Reproduction in mosquitoes is temperature-dependent and occurs more quickly when the temperature is warm. The female mosquito takes blood to get protein to lay her eggs (male mosquitoes do not bite people). If a mosquito transmitting malaria in Georgia were to become infected from her first blood meal after emerging as an adult, she might infect 11–13 people in her life. The same species in New York would be unlikely to infect more than 4 or 5. As temperatures get colder, the metabolism of mosquitoes slows down. Below freezing it stops entirely, and death follows. A few mosquitoes, such as the vector of La Crosse encephalitis, sometimes pass the infection *transovarially*, but this is rare (see Table 4.2). Thus the disease cycle is interrupted annually in cold climates. A disease such as yellow fever can be introduced only as an epidemic in the middle latitudes and does not become endemic there, as it did in the Caribbean. In the wet–dry tropics, some mosquitoes have the ability to *estivate* (or to hide and dehydrate during the dry season) and survive. The infected adult mosquitoes can renew the transmission chain when the rains come.

There is growing appreciation of yet another biological strategy. Many viruses transmitted by mosquitoes, including several kinds of encephalitis found in the United States, are actually zoonoses of birds. Winter kills mosquitoes and breaks transmission over most of the United States, yet early every summer there have been puzzling new cases of Eastern equine encephalitis in Virginia. The viruses, it seems,

TABLE 4.1. Some Diseases Vecteded by Ticks and Mites

Disease	Agent	Major vector	Endemic locale	Reservoir/ comments
<i>Viral</i>				
Encephalitis, Central European	Togavirus, group B	Tick: <i>Ixodes</i> spp.	Europe	R: rodents
Encephalitis, Russian spring—summer	Togavirus, group B	Tick	Europe, North Asia	R: rodents; tick transovarial transmission
Louping ill	Togavirus, group B	Tick	Great Britain	R: sheep
<i>Rickettsial</i>				
Rocky Mountain spotted fever (similar to boutonneuse fever, South African tick typhus, India tick typhus, etc.)	<i>Rickettsia rickettsii</i>	Tick: <i>Ixodes</i> spp.	North America, especially Southeast piedmont	R: small mammals; tick transovarial transmission
Scrub typhus (tsutsugamushi disease)	<i>Rickettsia tsutsugamushi</i>	Larval trombiculid mites: <i>Leptotrombidum akamushi</i> and <i>deliens</i>	Asia: Siberia to Indonesia	R: rodents; mite transovarial transmission
<i>Bacterial</i>				
Tularemia	<i>Francisella tularensis</i>	Ticks: dog, rabbit, Lone Star	North America, continental Europe	Zoonosis of rabbits and rodents
Relapsing fever (endemic)	<i>Borrelia recurrentis</i>	Ticks: <i>Ornithodoros</i> spp.	Tropical Africa, Mediterranean, Asia Minor, Americas	R: rodents; tick transovarial transmission
<i>Protozoan</i>				
Babesiosis	<i>Babesia microti</i>	Tick	Nantucket Island, United States	R: rodents; rare in humans

are annually reintroduced along the great bird flyways during avian migration and arrive just in time for the new spring mosquitoes to become infected. The virus simply winters in warmer climes. When West Nile virus (WNV), a bird zoonosis that causes encephalitis in people, was first introduced into New York City, health officials struggled to contain the arbovirus. They wanted to prevent it from migrating south to Florida for the winter, to survive, spread, and return the next year. They targeted the main breeding sites in the stagnant swimming pools that had been created by water conservation measures that summer. Cold winter appeared to have killed the mosquitoes and ended transmission. The arbovirus's reemergence early the following spring was baffling, until it was noticed that swarms of mosquitoes emerged when the cover of a sewer line was removed. The vector of WNV likes

TABLE 4.2. Some Diseases Vecteded by Mosquitoes

Disease	Agent	Major vector	Endemic locale	Reservoir/ comments
<i>Viral</i>				
Chikungunya	Togavirus, group A	<i>Aedes</i> spp.	Africa, Southeast Asia	R: monkeys
Encephalitis, Eastern equine	Togavirus, group A	<i>Aedes</i> spp.	Americas	R: birds
Encephalitis, Western equine	Togavirus, group A	<i>Culax tarsalis</i>	Americas	R: small mammals, reptiles?
Encephalitis, Japanese B	Togavirus, group B	<i>Culex</i> <i>tritaeniorhynchus</i> , <i>C. gelidus</i>	East Asia, Pacific islands	Pig amplifying host
Encephalitis, St. Louis	Togavirus, group B	<i>Culex</i> spp.	Americas	R: birds
Encephalitis, Murray Valley	Togavirus, group B	<i>Culex</i> <i>annulirostris</i>	Australia, New Guinea	R: birds
Encephalitis, La Crosse	Bunyavirus, group C	<i>Anopheles</i> <i>triseriatus</i>	Central United States	Transovarial transmission
Dengue fever	Togavirus, group B	<i>Aedes aegypti</i>	World tropics; hemorrhagic form, Southeast Asia	R: monkeys, Southeast Asia
Yellow fever	Togavirus, group B	<i>Aedes aegypti</i>	Tropical Africa, tropical forest South America	R: monkeys
<i>Protozoan</i>				
Malaria	<i>Plasmodium</i> : <i>vivax</i> (tertian) <i>falciparum</i> (malignant) <i>malariae</i> (quartan) <i>ovale</i> (W. African)	<i>Anopheles</i> spp.	Tropics and subtropics	Humans are source
<i>Helminthic</i>				
Filariasis (elephantiasis)	Nematode: <i>Wuchereria</i> <i>bancrofti</i>	<i>Culex fatigans</i> , <i>Aedes</i> spp.	Tropics and subtropics	No reservoir: infective microfilaria, night or day periodic, depending on local vector; zoonosis of wild and domestic animals in Malaysia
Filariasis (elephantiasis), Malayan	Nematode: <i>Brugia</i> <i>malayi</i>	<i>Mansonia</i> spp.	Southeast Asia, India, China	

organic pollution, and found warmth underground in New York City. Perhaps aided by the vehicles of people migrating to keep warm in Florida, WNV did spread to the Gulf of Mexico, and managed to cross to the Mississippi and West Coast flyways also. Birds were hard hit. As is the case with the native Eastern equine encephalitis, however, infection of people, who are dead-end hosts, is incidental.

Analyzing the efficacy of a vector requires detailed knowledge of its particular biology. The following characteristics, for example, affect mosquitoes' usefulness as vectors: flight range, altitude range, sex ratio, breeding habits, preferred breeding sites, life expectancy, age structure of local mosquito population, alternative (other than blood) food, activity time, host preference, resting habits, biting habits, and tolerance for specific agents (viruses, microfilariae, protozoa, and others). Working with entomologists, epidemiologists and geographers can attach probabilities to each step, and each contact, each stage of an epidemic, and can model its course or monitor an intervention. Even without such detailed information from costly field research, knowing the time of biting, flight range, and type of habitat for breeding helps geographers understand viable interventions and cultural buffers.

Some mosquito species require fresh water with no pollution for the larvae; others flourish on organic pollution such as sewage. Some breed in containers, tree holes, or leaf tendrils; others need open expanses of water. A few like brackish conditions. Some mate for life when they emerge as adults from the water; others hover around blood sources, hoping to find a mate. Some are weak fliers; others are easily carried long distances by wind; and yet others are strong fliers, even upwind. Most mosquitoes prefer to bite other animals, including snakes, frogs, and birds, rather than humans—but a few species prefer people, and others tolerate them when hungry enough. Some mosquitoes readily enter human dwellings and after feeding rest on the walls; others never enter dwellings and bite only outside. Some are active during morning or evening; others are active only at night. The detailed ecological requirements for specific mosquitoes are inherent in the landscape, and the distribution of relevant species determines where various disease agents can be spread.

Aedes aegypti is an especially dangerous vector, as is evident from its ecology. It prefers to bite people; it is active during times of day, especially early evening, when people are active. *Ae. aegypti* breeds in containers and has adapted well to opportunities offered by human dwellings, such as rain gutters, flowerpots, jars of bathwater, toilet bowls, and backyard trash containers. It has readily spread around the world with human help, much as the rat has. This is due to its habit of laying eggs just above the water line in containers. The eggs can thus sail across the ocean and hatch when the water jars are filled in the new port. Aside from the human louse, *Ae. aegypti* is probably our most commensal (domiciliated) arthropod vector. It transmits several of the most dangerous viruses (see Table 4.2).

Flies inhabit a wide range of biomes (see Table 4.3). In general, they are most important in arid places. The flies of Australia, of the wet-dry tropics, and of arid grazing lands are infamous. A few types of these flies are bloodsucking vectors of human disease. The most notorious of these is the tsetse fly, which transmits trypanosomiasis (African sleeping sickness). This disease mainly affects animals. In one form, *Trypanosoma rhodesiense*, it is an anthroppo-zoonosis that has a reservoir in

TABLE 4.3. Some Diseases Vectored by Flies and Reduviid Bugs

Disease	Agent	Major vector	Endemic locale	Reservoir/ comments
<i>Viral</i>				
Sand fly fever	Bunyavirus, group C	Midge, sand fly: <i>Phlebotomus</i> <i>papatasi</i>	Mediterranean climate areas, Asia Minor, Central and South America	Human—fly complex crucial
<i>Bacterial</i>				
Tularemia	<i>Francisella</i> <i>tularensis</i>	Deer fly	North America, Europe, Japan	Zoonosis of rabbits and rodents Humans are source
Bartonellosis (Oroya fever, Carrion's disease)	<i>Bartonella</i> <i>bacilliformis</i>	Sand fly: <i>Phlebotomus</i> spp.	Andean valleys in Peru, Ecuador, Columbia	Humans are source for <i>gambiense</i> R: wild game and cattle for <i>rhodesiense</i>
<i>Protozoan</i>				
African trypanosomiasis (sleeping sickness)	<i>Trypanosoma</i> <i>gambiense</i> , <i>T. rhodesiense</i>	Tsetse fly: <i>Glossina</i> spp.	Tropical Africa	R: wild and domestic animals
American trypanosomiasis (chagas)	<i>Trypanosoma</i> <i>cruzi</i>	Reduviid bugs	Tropical America	R: dogs, cats, rodents in different places Fatal untreated
Leishmaniasis, visceral (kala-azar)	<i>Leishmania</i> <i>donovani</i>	Sand fly: <i>Phlebotomus</i> spp.	Tropics and subtropics, Mediterranean	R: dogs, rodents
Leishmaniasis, cutaneous (Oriental sore espundia, uta)	<i>Leishmania</i> <i>tropica</i> , <i>L.</i> <i>brasiliensis</i> , <i>L. mexicana</i>	Sand fly: <i>Phlebotomus</i> spp.	Arid margins of Asia, Africa, Mediterranean, South and Central America	Humans are source
<i>Helminthic</i>				
Onchocerciasis	Nematode: <i>Onchocerca</i> <i>volvulus</i>	Blackfly: <i>Simulium</i> spp.	Tropical Africa, Central America to Amazon	Humans are source
Loaiasis	Nematode: <i>Loa loa</i>	Mangrove fly: <i>Chrysops</i> spp.	West and Central Africa	

African wildlife and is virulent in humans. In another, presumably older form, *T. gambiense*, it is milder, is slower to kill, and has its reservoir mainly among people, with animals being incidental. There are still other forms that affect only animals. The disease has had a major impact on human population distribution and nutrition by making the raising of livestock for protein and draft power virtually impossible across large areas of Africa. Even this fly, however, needs water to breed and becomes less active when humidity is very low. Another vector, the biting blackfly, *Simulium damnosum*, is the major vector of onchocerciasis (river blindness). It lays

its eggs in oxygenated, fast-flowing water by anchoring them to underwater stones. Exposure to the disease cycle, as the name suggests, is associated with use of the fertile soils near the streams and rivers or with use of the rivers for laundry, fording, or fishing. Hunter (1980) first described the complexity of the disease system and the multisided, integrated attack that is needed to control it. The various forms of leishmaniasis are transmitted by sand flies. Although characteristic of rather arid terrains, they require some humidity and are usually found at specific altitudes or ecological niches where the relative humidity is high enough. It finds shelter from the sun and low humidity in the plaster and adobe walls of human dwellings and the burrows of rodents.

One of the lesser-known and most unusual vectors of deadly disease is the cone-nosed “kissing bug,” or the reduviid bug. These bugs (that’s their name!) transmit the protozoal agent, a trypanosome, of American trypanosomiasis. Much better known as Chagas disease, it infects people in most countries in South and Central America, as there are many species of reduviid adapted to the different habitats. There are hundreds of reservoir animals in the forest, from giant sloths to dogs, but the disease has been most intense among people living in thatch-roofed housing, especially on the forest edge or frontier. The bugs, which resemble cockroaches in appearance and living habits, differ in the important way that they suck blood: They like to suck blood from lips and exposed membranes. After eating they defecate, and a person’s hand then rubs the bite, infecting it with thousands of trypanosomes in the bug’s feces. After initial, usually mild sickness, the trypanosomes multiply for years until eventually the person dies from complications of an enlarged heart (and related effects). There is more discussion of Chagas disease in Chapter 5, in connection with population mobility and landscape change in the Amazon.

THE LANDSCAPE EPIDEMIOLOGY APPROACH

Landscape epidemiology is a geographic delimitation of the territory of a transmitted disease in order to identify cultural pathways for disease control. Working with the American Geographical Society, Jacques May mapped mosquito vectors, and rickettsial and helminthic diseases, among others (see Chapter 1 and Plate 1). Several holistic approaches to the study of arthropod-borne or naturally occurring diseases have been developed in Europe. In Germany, for example, a holistic ecological approach was developed under the leadership of Helmut J. Jusatz. He produced the first world atlas of epidemic diseases (Rodenwaldt & Jusatz, 1952–1961), which not only portrayed disease distribution but also analyzed its association with climate, topography, hydrology, and flora/fauna. This work was expanded into a series of monographs on the medical geography of such countries as Kenya, Kuwait, and Korea (Jusatz, 1968–1980). This holistic approach and focus on ecological associations (*landschaftökologische*, in German) includes settlement and cultural patterns. The modeling approach of Pavlovsky provides a framework to do this. It needs to be noted that in U.S. medical geography in general, the study of the natural ecology

of disease, even when enhanced with May's influential cultural processes, has been weak.

May's Multifactor Zones and Pavlovsky's Nidus

May (1958a) described areas where transmissible diseases were present as regions where multiple factors had to occur and overlap for disease systems to occur. He wrote of *three-factor* and *four-factor* diseases. Each factor has its own spatial distribution, and all factors had to coincide in space to create the disease conditions. In the example that follows, plague is a four-factor disease: The agent is one factor; the rat host (reservoir) is a second; the flea vector is a third; and the human host is a fourth. May called a place where the three factors of agent, vector, and reservoir coincide, but disease is not known because no humans are present, a *silent zone* of disease. People build houses, roads, or campgrounds, or otherwise intrude into a silent zone, and are surprised to develop encephalitis, yellow fever, or plague. When people such as pastoralists in the grasslands or woodcutters in the forests penetrate into the silent zone, or farmers settle it, disease can result. In the former Soviet Union, this was a common occurrence in the settlement of Siberian mining and industrial towns. It was Pavlovsky's hypothesis that the potential for specific diseases could be identified from the simple existence of certain environmental conditions, so that planners could design housing, protective clothing, or work scheduling to shield the human population from exposure to the zoonosis. In the United States, there has been limited work on the biogeography of zoonoses. Instead, the silent-zone concept is being extended to nonbiotic environmental hazards, such as toxic land.

As noted at the start of this chapter, the concept of *landscape epidemiology* was developed by a Russian parasitologist and geographer, Eugene N. Pavlovsky (1966). Landscape and connections with Russian culture were a central concern for Russian geographers, not unlike the French geographers of the time. Pavlovsky was one of the leading parasitologists of the former Soviet Union; he was director of the Zoological Institute of the Academy of Sciences and president of the Geographical and Entomological Societies of the USSR. Because of the importance of the landscape epidemiology approach he developed, the Institute of Medical Geography became a separate entity in the Soviet Academy of Sciences. He developed the *doctrine of natural nidality*, or the natural focus of disease. In a *natural nidus*, infection is maintained among wild animals and arthropod vectors. These zoonoses, which Pavlovsky determined and established through fieldwork, occur in particular kinds of terrain. There is a biogeography to the life cycles of the arthropods and various other animals involved. Their food sources, soil, climate, slope, exposure, and other ecological parameters determine local distribution and possible occurrence of the disease cycle. By knowing the conditions necessary for specific diseases, scientists can use the landscape to identify disease hazards. Landscape modification can create, or be used to prevent, the establishment of disease cycles. Perhaps the easiest way to understand the concepts of landscape epidemiology is to consider one of Pavlovsky's best-known examples, bubonic plague.

Bubonic Plague

“Ring around the rosy” (the rash around the flea bite), “a pocket full of posies” (flowers [or, archaic, verses—of the Bible?] used as amulets to ward off evil spirits and disease), “ashes, ashes” (for the sign of the cross made on the foreheads and house doors of sick people—or, in a German variant, “achoo, achoo,” the sneeze as lethal harbinger producing the response *Gesundheit*), “we all fall down” (and die). So deep is the apocalyptic experience of the bubonic plague in the European soul that 600 years after the Black Death, their cultural descendants still teach their children the bubonic plague song.*

Bubonic plague is a natural zoonotic disease that has spread to parallel ecologic niches in many parts of the world, and has created new resident urban ecologies as well. The infection drains from the flea bite into the lymph nodes, which become swollen *buboes*. When they rupture, infection of the blood (*septicemia*) spreads the bacteria throughout the body. Endotoxins cause shock, blood coagulation, and hemorrhaging into the skin vessels, often turning the body black. Untreated bubonic plague today has a case mortality rate of 50–60% (Benenson, 1995). A second form of plague occurs under certain circumstances of climate, crowding, and intensity of infection. Secondary infection of the lungs then causes contagion to other people as a *pneumonic plague*, infecting them through the respiratory system. Untreated pneumonic plague is usually fatal. The great plagues causing massive deaths have involved the highly contagious, lethal pneumonic form. Wills (1996) gives an excellent discussion of the evolution and natural history of bubonic plague.

The bacteria that are the agents of bubonic plague may have been cycling among rodents in the grasslands of Central Asia since prehistory (although McNeill [1976] thought that the original focus was in the Indian Himalayan foothills and that the infection spread to Eurasian steppes in the second millennium). There is also an ancient focus of plague among rodents in the Great Lakes area of Africa. It is not clear whether the bubonic plague that was the Plague of Justinian (A.D. 542–543), and came to Constantinople from Egypt, originated in India or Africa. The tropical black rat, *Rattus rattus*, had managed to cross from the Indian Ocean to the Mediterranean Sea ports. Newly introduced to Europe, it did not yet occur in northern Europe, and thus that devastating epidemic remained limited to the eastern Mediterranean. In 1346, bubonic plague did spread out of the steppes to the coastal ports of the Black and Mediterranean Seas and across to Europe. By the time of the Black Death, 1348, the black rat had spread throughout the cities of Europe. Population pressure was also severe. Europe was primed for a pandemic that would reduce its population by more than a third.

Bubonic plague is a rodent zoonosis transmitted by fleas (Table 4.4). The rodents in the central Asian grasslands, the natural nidus of plague, show no symptoms; that is, they are quite resistant to their ancient infection. Consider for a moment what makes up the living community in which the agent of bubonic plague, the bac-

*With such a beautiful example of deep culture, it seems a shame to note that these origins are disputed, as no one has found a record of the verse before the 19th century. But why else would we teach our children such nonsense generation after generation?

TABLE 4.4. Some Diseases Vectored by Fleas and Lice

Disease	Agent	Major vector	Endemic locale	Reservoir/ comments
<i>Rickettsial</i>				
Typhus fever, epidemic	<i>Rickettsia prowazekii</i>	Fluids of body louse, <i>Pediculus humanus</i>	Cold climes, especially mountains in South America, Himalayas, Balkans	Source is people
Typhus fever, murine, endemic	<i>Rickettsia typhi</i>	Flea, especially <i>Xenopsylla cheopis</i>	Worldwide	Rats
<i>Bacterial</i>				
Plague, bubonic	<i>Yersinia pestis</i>	Flea, especially <i>Xenopsylla cheopis</i>	Widespread natural foci; western United States, Vietnam, South Africa, South America, Central Asia	R: rodents
Relapsing fever (epidemic)	<i>Borrelia recurrentis</i>	Louse, <i>Pediculus humanus</i>	Local areas in Asia Minor, North and East Africa, South America	Source is people; in outbreaks from nidal, tick-borne

terium *Yersinia pestis*, circulates continually (Figure 4.3). The rodents live in burrows deep enough to provide insulation from the severe heat and cold of Central Asia. Under the grass, the horizon of organic soil is deep. The soil is drained well enough that the burrows are unlikely to be flooded, even when the snow melts in spring. The degree of flooding or drainage depends on local slope, as well as on the amount of rainfall and the solar radiation received. In the burrow are nesting materials of an organic nature, grass roots, and a variety of arthropods (flies, roaches, and others) that take shelter there. Among the worms and arthropods, which might be predators, flea eggs hatch into larvae, which live off the organic matter. Because rats breed frequently and have large litters, there are susceptible baby rats for adult flea vectors to infect when they eat. Therefore, actively infected young rats are often present to feed the rat fleas when they first emerge from the larval stage. In vacated burrows, adult fleas can maintain infection for months while waiting for new hosts.

The landscape must have the right kind of vegetation for the rats to eat, and a certain mix of predators and competitors to regulate the population. There are also micrometeorological limits to this vegetation and fauna. Sometimes fire serves to scatter the rodents and introduce infected fleas to new burrows. At the margins of the territory, where winter cold or summer droughts are so severe that the rat population is unstable, it is difficult for the bacteria to circulate continually to new hosts. These secondary nidi often collapse autonomously. Experts in the former Soviet Union became quite precise in identifying the slope, exposure, soil type, acid-base balance, vegetation associations, and so forth that mark areas where the transmission chain is most likely to be established in primary nidi. They concentrated on fumigating burrows in that landscape and managed to reduce greatly the natural focus of bubonic plague in Central Asia.

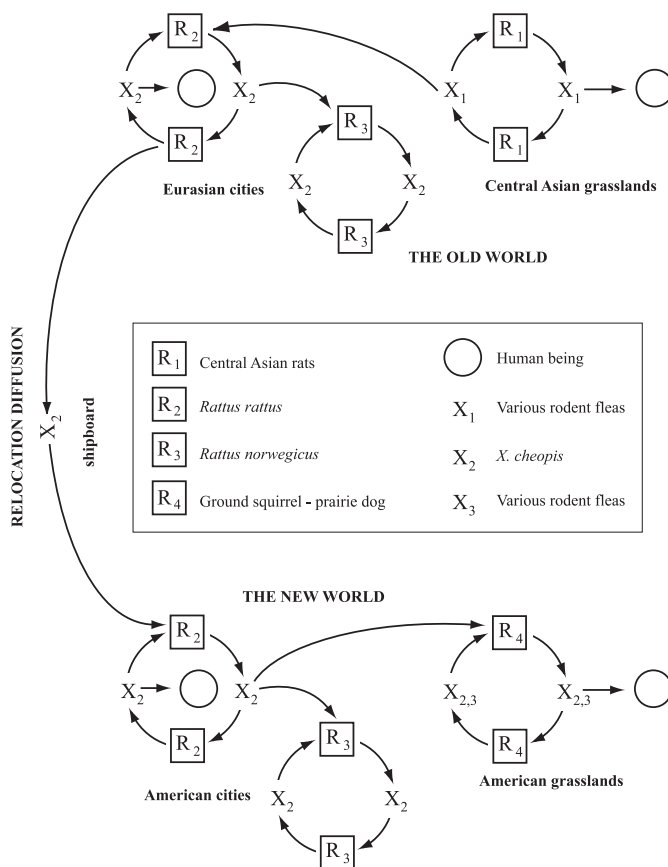


FIGURE 4.3. Bubonic plague transmission chain.

Occasionally nomads and hunters must have camped in this natural nidus of the disease, and individuals must have been bitten by fleas and died, perhaps scores of miles away. When infected fleas were carried by caravans in bedding or on bodies from this natural nidus to a city such as Constantinople (today's Istanbul), a *transference*, or *relocation diffusion*, of the disease cycle occurred. The hungry fleas bit urban rats, which were quite susceptible to this new infection and died in droves. As the urban rats died and turned cold, their fleas sought new hosts and spread the epizootic. Urban rat populations were, and are, enormous. As the rodents died in the millions, many millions of fleas looking for hosts encountered human beings. One type of flea, *Xenopsylla cheopis*, made an especially dangerous vector because of its willingness to bite humans, who are avoided by most flea species.

The cultural ecology of European medieval cities was ideal for epidemics of plague, as described in Chapter 2. Housing was very crowded; water was of limited availability, especially in the quarters of the poor; hygiene was terrible. Soap was a luxury; hot water was virtually unknown; food was stored in and near dwellings, which rats knew. The 13th-century population boom had led to massive defor-

estation and overexpansion of areas of agricultural development. This landscape change had caused erosion and pushed the transportation system's ability to supply foodstuffs to the growing cities beyond its limits. The climate had worsened and grown colder and wetter, and crop failure was more frequent. Inflation, common malnutrition or even starvation, filth, and disruption were rampant. Europe was primed for "the plague." As people fled the epidemic, or as goods were transported with their stowaway fleas and rodents, the disease agent reached new towns and even the new rat populations of northern Europe. Bubonic plague swept from east to west across Europe, and then returned from west to east through a new generation of susceptible children. Some people developed a terminal pneumonia that enabled the bacteria to spread by direct means also (through respiratory contagion). As noted earlier, the first wave of the Black Death killed a third of the population of Europe; in some places more than half the people died. Each successive wave over the next couple of centuries, however, had lower case mortality rates, which dropped from nearly 90% to 60% and perhaps less in places.

As the brown rat, *Rattus norvegicus*, spread into Europe in succeeding centuries and outclimbed, outfought, and displaced the black rat, *R. rattus*, the ecology of the disease changed. The brown rat was not as sociable. The black rat was originally a wild, tropical rat before it became domicilitated. It preferred to hang around in people's houses, storehouses, and barns, where it was warm. *Rattus norvegicus* was from the temperate latitudes of Asia. It preferred to live in sewers and other places where it could avoid people. The parameters of contact among dead rats, fleas, and people changed. Many of the housing and nutritional conditions changed, too, under new population and economic conditions. The great pandemics abated. Bubonic plague has continued, however, to break out when war or other severe disturbances have destroyed the ecological balance.

Meanwhile, plague had been carried by ship to the cities of the New World. In the late 19th century, for example, an epidemic in San Francisco led citizens of European origin there to burn down crowded Chinatown, identifying the "filthy Chinese" as the source. This conflagration effectively dispersed the rats. From the urban rat epizootics, wild rodents must have become infected. Today bubonic plague is enzootic among the burrowing rodents of the Southwest. In the grasslands of the U.S. Southwest, prairie dogs and ground squirrels live in burrows in which most of the nidal conditions of Central Asia are duplicated.

A few people die every year in the southwestern United States from bubonic plague. Frequently, later investigation determines that a cat brought home an infected ground squirrel or that some similar event occurred (cats are so susceptible to bubonic plague that they even develop buboes). From time to time an epizootic will become evident among ground squirrels in a western state park, and the park will have to be closed. A public health threat exists that campers or other travelers in the arid grasslands will pick up some infected fleas in their baggage (as Asian travelers before them did) and transport them to a city like Chicago, where there is a large rat population that has become quite tolerant of our worst rat poisons. Rats infected with the plague have been caught in monitoring traps in Golden Gate Park in San Francisco. Few in the United States, however, understand anything about the

biogeography of the natural nidus of bubonic plague in their country. Unlike their (former) Soviet counterparts, U.S. geographers have not been involved in any effort at control or eradication.

Plague outbreaks have occurred in most of the world's great ports. Plague has become enzootic among the gerbils of South Africa, the field rats of India and of Java, and the marmots of Manchuria (the northeastern territories of China, where in the early years of the 20th century a great plague epidemic killed more soldiers in the Russo-Japanese War than all the combined armaments). Worldwide, more than 220 species of rodents have been shown to harbor plague. Although the incidence of plague continues to decline, the war in Vietnam showed again that when the cultural-biological interface is disturbed, plague can flare into epidemics quickly. Tens of thousands of cases of bubonic plague occurred in refugee-swollen Saigon and the disturbed countryside of South Vietnam. Outbreaks of the deadly pneumonic form followed, but were contained in the cities by antibiotic therapy and a short-lived vaccine. In 1994, in Surat, India, plague broke out in the squatter slums and panicked the nation. Monsoon failures had compounded the rural-urban migration, bringing rural migrants into the crowded housing. Food had been imported to the region and stored. Rats had been observed dying in nearby villages, but government officials were initially afraid of inciting panic. Later, although hundreds of people died, a major epidemic was prevented from spreading by meeting travelers at train, bus, and airplane stations. Some countries quarantined passengers arriving from India. It would seem that the containment strategy worked; what is clear is that the underlying ecology has not been addressed.

In 2003, 2,118 cases of bubonic plague (an internationally reportable disease) were reported to the WHO, with 182 deaths. As the WHO puts it, "today the distribution of plague coincides with the geographical distribution of its natural foci." The propensity to outbreak when the ecology is disturbed remains.

THE CULTURAL DIMENSION OF WATER-BASED DISEASE TRANSMISSION

The importance of the human role in habitat modification, of cultural buffering systems, and of human mobility for disease systems has been discussed in Chapter 2. These factors are illustrated here with three diseases: malaria, dengue hemorrhagic fever, and schistosomiasis. These three diseases are chosen not only because of their importance, but also because they are water-based in different ways. They also serve as examples of how activities that promote economic development, especially in the tropics, can modify nidal conditions and affect disease patterns. Table 4.5 shows the mortality from nine of the worst vectored diseases. The regionalization of some can clearly be seen. The table also shows the *disability-adjusted life years* (DALYs) for the same disease. When a disease like schistosomiasis has low deaths (16,000) but high DALYs (almost 2 million years lost), it is clear that the mortality distribution does not begin to show the economic and human importance of this chronic, debilitating disease.

TABLE 4.5. Some Major Vectored Diseases: Deaths and DALYs (in Thousands), 2002

Disease (Vector)	Total	Africa	Eastern Mediterranean	South and SE Asia	Western Pacific	Americas	Europe
<u>Water-based</u>							
Malaria (<i>Anopheles</i>)	1,272 <i>46,486</i>	1,116 <i>40,855</i>	59 <i>2,250</i>	65 <i>2,777</i>	11 <i>441</i>	1 <i>111</i>	0 <i>21</i>
Dengue fever (<i>Aedes aegypti</i>)	19 <i>616</i>	0 <i>5</i>	1 <i>30</i>	12 <i>381</i>	4 <i>131</i>	2 <i>69</i>	0 <i>0</i>
Japanese encephalitis (<i>Culex</i>)	14 <i>709</i>	0 <i>0</i>	2 <i>83</i>	8 <i>306</i>	3 <i>320</i>	0 <i>0</i>	0 <i>0</i>
Lymphatic filariasis (<i>Culex</i> , <i>Ae.</i> , <i>An.</i>)	0 <i>5,777</i>	0 <i>2,011</i>	0 <i>122</i>	0 <i>3,219</i>	0 <i>411</i>	0 <i>10</i>	0 <i>0</i>
Onchocerciasis (<i>Blackfly Simulium</i>)	0 <i>484</i>	0 <i>470</i>	0 <i>10</i>	0 <i>0</i>	0 <i>0</i>	0 <i>2</i>	0 <i>0</i>
Schistosomiasis (Snails intermediate)	16 <i>1,702</i>	2 <i>1,334</i>	9 <i>226</i>	0 <i>7</i>	4 <i>55</i>	1 <i>74</i>	0 <i>0</i>
<u>Other vectored</u>							
Leishmaniasis (Sand fly)	51 <i>2,090</i>	8 <i>383</i>	5 <i>348</i>	36 <i>1,358</i>	2 <i>50</i>	0 <i>45</i>	0 <i>6</i>
Trypanosomiasis (Tsetse fly)	48 <i>1,525</i>	47 <i>1,484</i>	1 <i>39</i>	0 <i>0</i>	0 <i>0</i>	0 <i>0</i>	0 <i>0</i>
Chagas disease (Reduviid bug)	15 <i>667</i>	0 <i>0</i>	0 <i>0</i>	0 <i>0</i>	0 <i>0</i>	15 <i>655</i>	0 <i>0</i>

Note. Data from WHO (2004). For each disease in each region, deaths (in thousands) are given first, followed by DALYs (in thousands) in italics. Burden of disease in DALYs by cause, sex, and mortality stratum.

Malaria

Malaria is by far the most serious vectored disease in the world. It is one of the three diseases targeted by the WHO with the Global Fund. In the 1950s it was possible to dream about eradicating malaria because it has no reservoir. In the 1960s eradication or control over large areas of Asia and South America was in fact achieved. In the 1990s, however, malaria reclaimed land and again grew to infect more than half a billion people and threaten many more (see Plate 6). Malaria kills 1–2 million people a year (over 1 million in tropical Africa), totally incapacitates tens of millions, and causes hundreds of millions to struggle in exhaustion with chronically severe anemia (see Table 4.5). About 40% of the world's people, in more than 100 countries, live in regions where malaria is endemic today. The health improvement in the 1960s in countries in which malaria was almost eradicated was eye-opening. Infant mortality rates plummeted. Fertility rates increased as miscarriages and cases of infertility that were due to malaria disappeared.

The terrible pressure malaria has exerted on the human population over the ages is demonstrated in the population vertex of the triangle of human ecology by the presence of high-mortality genetic diseases, most famously sickle cell anemia. Yet the case mortality rate today for malaria is low compared to that of most serious

diseases. Malaria is generally, depending on type, a chronic, debilitating disease. Despite decades in which scientific armament has been marshaled against malaria, it is resurgent in many countries where previously it was virtually eradicated. Cases in Sri Lanka, for example, went from 1 million a year in the 1950s to less than 10, but then rose to 38,500 in 1982 (4% caused by *Plasmodium falciparum*—see below), to 676,000 in 1987 (27% *P. falciparum*), and to 1 million a year in the 1990s. It is still estimated that about 80% of all clinical cases and 90% of parasite carriers live in tropical Africa. More than 1.7 billion people live in once-endemic areas in which transmission had been greatly reduced but is now reinstated, and in which the ecology is unstable and the situation is deteriorating. Today malaria kills one child every 30 seconds. The mosquito vectors are resistant to all the major insecticides; the agent is resistant to all the major drugs; and the ancient scourge is upon us again.

The disease cycle consists of direct transmission of the agent between humans by various species of *Anopheles* mosquitoes. There is no animal reservoir, although monkeys, birds, and other animals have their own forms of malaria. The agent is a protozoan, of a genus called *Plasmodium*, occurring in four species. *P. vivax* causes benign tertian malaria, historically prevalent in Europe. It has a liver-dormant stage (hypnozoites) that allows it to survive even long winters and to reactivate in time to infect the spring mosquitoes. Its case mortality is low enough for it to have been nicknamed “benign tertian.” It does not occur in West Africa, where people genetically lack the required antigens in their blood cells and seem immune to it. *P. ovale*, the least common malaria, occurs only in West Africa—perhaps as a niche replacement. *P. malariae* causes quartan malaria. It is so old and well adapted as a human parasite that science has yet to discover how it hides and relapses in the blood. The most dangerous form of malaria is caused by *P. falciparum*. It has more than a 10% case mortality rate and often has unusual symptoms of disorientation, coma, shock, and renal failure. Apparently it has most recently transferred from birds. It spreads more slowly because it has not yet fully adapted to *Anopheles*, which are not bird feeders. It lacks the ability to go dormant in the human liver and so must reproduce in enormous numbers—which so block, explode, and shatter the circulatory system that it can kill in hours—in order to infect mosquitoes and stay in hypercirculation and continual reinfection.

All forms of malaria have complicated life cycles within the body, although the cycles differ in detail. Protozoa have much more genetic information, and are thus much more complex, adaptable, and cagy, than viruses and bacteria. The sexual forms, gametocytes, emerge from hiding inside the red blood cells within the gut of the *Anopheles* mosquito after she takes a blood meal. They mate in her gut wall, and reproduce large numbers of sporozoites, which migrate to the salivary glands. These get inoculated into people when, after depositing her eggs, she takes another blood meal. They swim to the liver, within which they grow and mature over a week or two into the blood forms, merozoites, which enter red blood cells and commence gobbling hemoglobin and multiplying by factors of 10. They consume the red blood cells and then break out of the empty husks into the blood stream in a synchronized manner (hence the alternation of fever and chills). The merozoites inside the cells produce the male and female gametocytes, which do not circulate in the blood (and

so are not killed by drugs), but which also do not reproduce until they are taken up by a mosquito to complete the cycle, in about 2 weeks. The complexity of this cycle is such that the hematozoites dormant in the liver were not discovered until 1980, after a century of intense microscopic observation. Each stage has different surface proteins, different antigens, different nutritional needs, and different behavior. The multiple, changing symptoms, forms of hiding, and details of human immunological response are still of such baffling complexity and mystery that malaria has defied decades of intense effort to develop a vaccine.

The cultural ecology of malaria is also one of the most studied. May (1958a) described a village in the highlands of Vietnam that had low-grade, endemic malaria. The disease was a drain on the population's energy, but it was not the worst of these highland villagers' problems. The people lived in houses on stilts and tied their water buffalo and cattle under their houses. They cooked in the houses. These customs were quite different from those of the lowland Vietnamese, who moved into the highland area one year. The lowlanders built their houses on the ground and kept their animals in proper barns. They also had cooking sheds separate from the houses. They were rapidly driven out of the highlands by epidemic malaria. The local species of *Anopheles* that transmitted the agent was not a strong flier. The height of a house on stilts deterred her. She preferred to bite animals and seldom went past the tethered animals under the house. She, like all *Anopheles*, was active at night and especially in the evening, when the house was full of smoke. The cultural practices of the upland people thus constituted a rather successful protective buffer. Changing nothing in the system except the cultural behavior of the people was enough to render the land uninhabitable.

The experience of the United States illustrates some of the complex forms of cultural interaction with the biotic disease system. Originally malaria was an Old World disease. The British brought *P. vivax* to North America, and the slaves they imported brought *P. falciparum* from Africa. The early, glowing reports of how healthy the colonies were faded as malaria spread. The nuclear settlements of the plantations were especially good foci of disease transmission. The coastal plantations became so deadly that the owners spent their summers in resorts (the foundation of several beach and mountain resort towns) and cities such as Savannah, Georgia. The greater resistance of Africans to the disease became a common explanation of why the institution of slavery was "that which is necessary."

People did not know the cause of the disease, but they were keen observers. They knew that as the forest around Savannah was removed and land was converted into rice fields, malaria increased. They thought that the cause was a miasma (*mal aria* means "bad air"), which the trees had protected them from. In what was probably the first environmental legislation for public health, Savannah demanded that all land within a mile of the city be used only for dry agriculture. It compensated farmers out of tax revenues for the loss they incurred by not being able to grow rice or indigo. Malaria thus decreased in Savannah for a few decades, until the disturbances of the Civil War.

Malaria was *the* American disease of the late 19th century. The confluence of the Ohio and Mississippi was so deadly that many would travel there only in win-

ter. The census of 1890 undertook a survey of causes of death. Although this census was not up to today's statistical standards of reporting, the rates it recorded for malaria—over 7,000 deaths per 100,000 people across the South, and more than 1,000 in such states as Michigan, Illinois, and California—are impressive (Figure 4.4). An average case mortality rate of 5% means that almost the entire population was infected over large areas. Yet by 1930 malaria had disappeared from the North and West, and caused fewer than 25 deaths per 100,000 people in the South, except for a few counties. What had happened?

The United States has one important malaria vector, *Anopheles quadrimaculatus*. It breeds in water where it can anchor its eggs on vegetation that intersects the surface in open sunlight, and where the currents are not too strong. When the forest was removed across the country, sunlight was let into marshes and into the poorly drained glaciated land of northern states. Developers of rice plantations planted lots of intersecting vegetation, and much occurred naturally elsewhere. Slow travel by rafts and barges turned canals and rivers into linear transmission channels. Susceptible migrants picked up the infection along the waterways, and travelers who were infected spread it to new places. Houses were of poor quality, and their glassless windows and slatted walls offered little obstruction to mosquitoes. People commonly had poor nutrition and had few animals to divert the mosquitoes.

At the turn of the 20th century, however, the poorly drained land of the glacial till country was deliberately drained for agriculture. Houses were more often con-

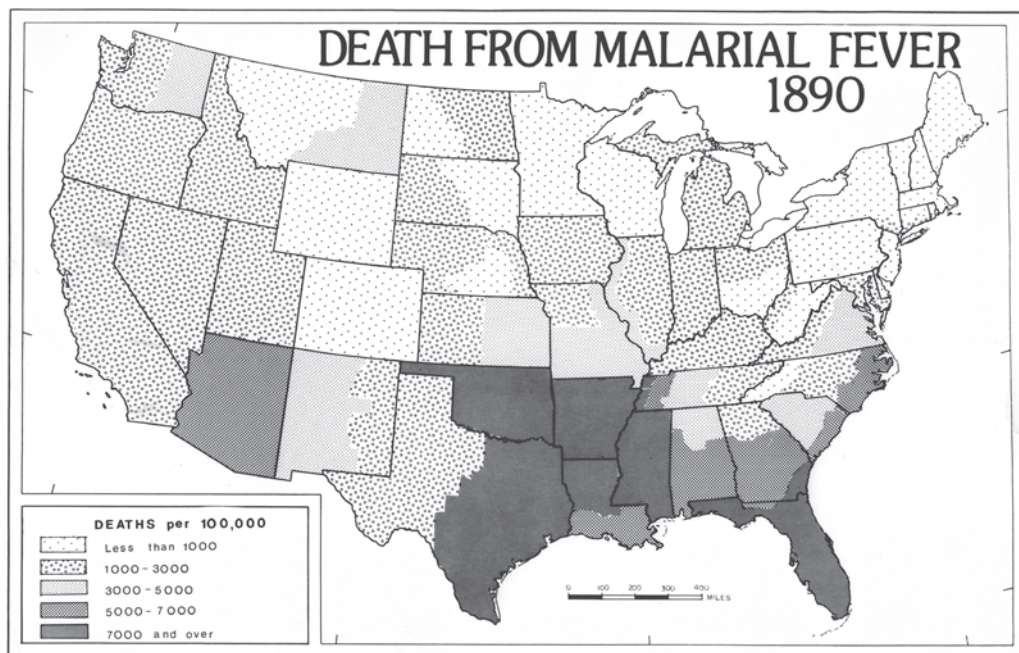


FIGURE 4.4. Malarial deaths in the United States, 1890. Adapted from U.S. Census Bureau (1890, map 17).

structed of brick or other good material and had glass windows and screens (mainly to keep flies out). Transportation shifted from the rivers and canals to the railroads; settlements as well as travelers shifted to the dry upland. The numbers of animals increased. The nutrition of the people improved. Quinine, an antimalarial drug, became available at a more reasonable price when the tariff against its importation (from Java, mainly) was finally dropped. The slower breeding of the mosquito in the cold climate of the North led to the disappearance of Northern malaria. In the South the incidence was greatly reduced as the population became more settled (and therefore less mobile), but some counties near sinkholes or the oxbow lakes of the Mississippi remained highly malarious.

Eradication Campaigns

In the late 19th century, Laveran discovered the agent of malaria, and Ross proved its transmission by *Anopheles* mosquitoes. In the United States, a public health official named Carter (1919) was amazed at how little use had been made of this knowledge to prevent malaria. The disease was so common in the South as to be taken for granted. He had to demonstrate to state and other officials the tremendous economic costs of lost workdays and debilitated farmers. The new Rockefeller Foundation became interested. It supported the training and staffing of entomologists and other experts, as well as the creation of public health departments at the county level. Carter himself determined the flight range of the U.S. vector (1 mile), its water preferences, and the effects of water impoundment for mills. Mills were commonly closed on Sunday, and on Monday the elevated mill ponds were drawn down. Carter determined that this was good because it stranded the flitage that anchored larvae.

Carter's findings became relevant when the Tennessee Valley Authority (TVA) proposed to construct dams that would impound large bodies of water in the malarious South. The economic value of developing cheap electricity was obvious, but what would the cost to the public health be if malaria again became epidemic? The Army Corps of Engineers worked closely with health officials. All people near the proposed reservoirs were checked for malaria and treated, and all dwellings within a mile of each reservoir were screened. All vegetation was cleared within 10 feet above maximum and below minimum water levels to prevent intersecting vegetation. The water level had to be fluctuated, in spite of navigational needs, at fixed times of the year in order to strand flitage. It all worked, and malaria almost disappeared from the TVA area, instead of becoming epidemic.

Similar measures were adopted elsewhere, and levels of malaria were reduced. When threatened by a resurgence of malaria brought by soldiers returning after World War II from fighting in malarious countries, the United States mobilized for eradication of the disease. The military office that had tried to eradicate malaria around military bases where soldiers were training was located in Georgia. It was transformed to head the mobilization against malaria. The weapon was the new insecticide, DDT, which could be sprayed on houses and would for months afterward kill the mosquito vector when it rested on the walls after taking its blood

meal. In a massive effort, workers sprayed millions of houses across the South and detected and treated thousands of cases. Malaria was eradicated in the United States. The institution that had coordinated the attack became the CDC, located in Atlanta, Georgia (a city in the previously malarial South), rather than Washington, D.C.

When soldiers returning decades later from the war in Vietnam brought in relapsing malaria cases by the thousands, public health concerns were again raised. The plasmodia, however, found the environment so modified that they could no longer get established in the United States. Few people now sleep where mosquitoes can fly into their bedrooms and bite them; detection is swift, and treatment is intensive. The control of malaria in the United States has been so complete that many people forget why reservoir landscapes have exposed banks and fluctuating water levels. Uses for recreation, livestock, and fish cultivation are coming into conflict with water management laws that were enacted in every state in the South for control of malaria. Public health workers have grown complacent and government officials ignorant. Yet today, because of population mobility, more than 1,000 cases of malaria a year occur. Some cases have been fatal when physicians failed to recognize the disease. There have, after a gap of 30 years, been a few cases of indigenous transmission in Texas, California, Michigan, Florida, New Jersey, and even New York City. Homeless people sleeping without shelter are again exposed.

Beginning in the late 1950s, the WHO attempted to eradicate malaria worldwide. The effort achieved notable successes, especially in Southeast Asia and Central America. By the 1970s, however, health officials talked about control instead of eradication, and in the 1980s they lost control. In most places the eradication campaign had made little progress at all, especially in tropical Africa, where *P. falciparum* predominates and prevalence remains at the high traditional levels that induced a lethal genetic adaptation. Many reasons for the failure of eradication have been offered by authorities. In some cases maintenance and surveillance failed when, for budgetary reasons, the special malaria teams were combined with multipurpose health programs. The international availability and costs of insecticides have changed. Some (e.g., Farid, 1980) believe that when the reduction in malaria led to faster population growth, the capitalist countries cut back their financial support; some even suggest that support was intentionally withdrawn from the eradication effort because of a desire to keep Third World countries poor, sick, and dependent. Whatever faults there were in administration and in financing, however, the original hopes of eradication were probably unrealistic because they failed to appreciate the complexities of tropical ecosystems. Tropical ecology is very complex, and eradication was modeled on the simple U.S. system described above.

In temperate countries, such as the United States, there are few vector species. The United States had only one. Sheets of open, fresh water with intersecting vegetation were the breeding areas. Spraying walls of houses with a residual insecticide interrupted transmission because the vector bit at night in the houses and then rested on the wall. Figure 4.5 illustrates the contrasting, very complex vector ecology in tropical Malaysia. *Anopheles sundaicus* breeds in brackish water, such as mangrove swamps along the coast. *A. campestris* breeds in rice fields, while *A. letifer* likes coconut, rubber, and swamp trees at low elevations. *A. umbrosus* is the vector of the

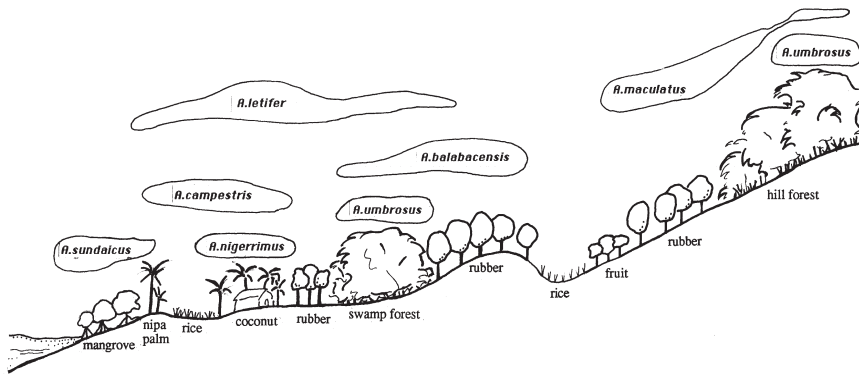


FIGURE 4.5. Malaria vectors in Malaysia. The size of the “clouds” indicates the relative importance in each habitat, or land use zone, of various species of *Anopheles* mosquitoes in peninsular Malaysia. Reprinted by permission from Sandosham (1970).

forest. *A. balabacensis* will breed in the shaded streams of the natural forest or of the rubber tree replacement forests. All these vectors must compete with many other species of mosquitoes. Each of the vector species occurs in relatively small numbers and prefers to bite animals. Some species bite people only outdoors in the early evening and never enter houses. Others enter houses, but fly out the window afterward and rest under vegetation. Most of the *Anopheles* species in Figure 4.5 are normally the vectors of endemic, chronic malaria at low levels. *A. maculatus*, however, breeds in streams exposed to sunlight and readily bites humans. It rapidly increases to very large numbers when given the simple, open habitat it prefers. Thus, when the forest is cut down for land development schemes or plantations and the forest streams are exposed to sunlight (see Plate 9), *A. maculatus* rapidly increases in numbers; as a vector of epidemic malaria, it is one of the most dangerous in the world.

The objective of the eradication campaigns was to interrupt transmission, using some of the known intervention points (Figure 4.6). Residual insecticides sprayed on houses were used to kill the adults after they had taken a blood meal. Humans were examined, and often chloroquine was given as a general prophylaxis to the entire population, to kill the plasmodia in their blood. It was recognized that the mosquitoes could not be suppressed forever. The hope was that when they came back, the agent of the disease would no longer exist in that place. This strategy had worked in temperate lands, but the complex ecology of the tropics was frustrating. Although some vectors bit people in their bedrooms at night, rested on the wall, and died, for example, others bit people only outdoors or flew out the window after biting people and never touched the insecticide. Broader attempts to interrupt breeding did not work. Oiling water bodies or stocking them with tilapia fish, which eat mosquito larvae, did not affect mosquitoes breeding in tree holes or hoof prints. People cutting firewood in the mangrove swamps continued to be bitten by *A. sundaicus*. For many such reasons, when the wall-resting, common vector became resistant to the insecticides and rebounded, some malaria was still present in the

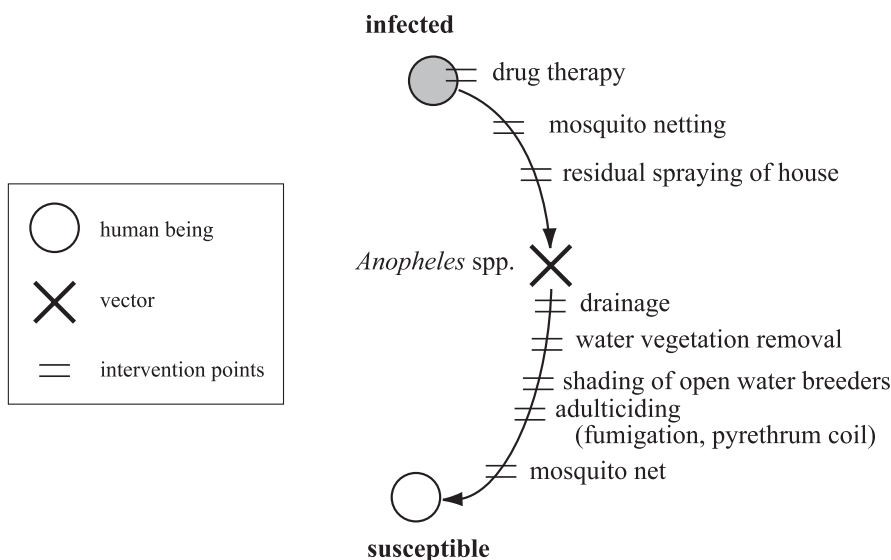


FIGURE 4.6. Malaria intervention. Malaria is vectored by various species of *Anopheles* mosquitoes, but human malaria only infects humans.

community, and the transmission chain resumed. Often the population had lost any partial immunity, however, and epidemics, higher infant mortality, and intensified health problems resulted.

Human mobility also played an important part in eradication failure. Laborers clearing the forest for planting would catch malaria from the *A. maculatus*-borne epidemics and return to their villages. People from areas in which the plasmodia had become resistant to chloroquine prophylaxis would migrate to other areas and introduce the resistant strain. As we will discuss in Chapter 5, Prothero (1961) had earlier demonstrated in Africa the importance of *labor-sheds* (vast areas within which people migrate to work) and pastoral mobility patterns for reintroducing an infection to cleared areas. There was an international failure to analyze the problem and coordinate efforts at the appropriate scale. Refugee movements confounded what efforts were made.

Nevertheless, real progress is occurring in tropical Africa today as the WHO Global Fund focuses on its millennial goal of decreasing malaria. The WHO recently announced on its website that 27 countries have achieved a reduction of up to 50% on malaria cases since 1990. Much of this is due to the drive, funded in large part by the Gates Foundation, to get insecticide-impregnated bed nets (alias mosquito nets) out of African ports where they have languished because of corruption and lack of infrastructure, and to millions of children in even difficult-to-reach villages. In Tanzania, where chloroquine had not been allowed even over the counter for decades, hospital personnel were delighted to discover in 2008 that it worked again! The toll of almost a million children dying annually in Africa has become intolerable and is likely to be reduced. In some places there is recurrent interest in permanently

altering the environment (e.g., through drainage and housing) to defeat in detail the ecology of the local disease system rather than relying on the magic bullets of chemistry. A new drug, *Artemisia annua* (*qinghaosu*) from the traditional Chinese pharmacopoeia is proving useful for multiply resistant strains of plasmodia, but no one doubts that it is already being too commonly (mis)used on the Thai–Cambodia border and that resistance will soon develop. The long-promised DNA vaccine has not yet appeared. Despite the recent progress, no one today even dreams of eradicating this ancient scourge.

Dengue Hemorrhagic Fever

In the 1950s a dangerous new disease began spreading in Southeast Asia. It was first noticed in Manila, but within a few years it was occurring in Thailand. At first it primarily struck Asian children, with a case mortality of over 10%, although this figure has fallen to about 3% with better therapy. Investigators considered many factors, including diet, racial susceptibility, and complications from other diseases, without finding the disease agent. Virologists finally isolated the agent of this new disease, and it was the same as that of dengue fever; thus the new disease was called *dengue hemorrhagic fever* (DHF). Like many arboviruses, including the agent of yellow fever, it is a flavivirus (at the time classed as a togavirus antigenically in group B). It occurs in four types, simply called 1, 2, 3, and 4.

Dengue fever is a monkey zoonosis transmitted by *Aedes albopictus* in the forests of Southeast Asia. All four serotypes occur in that jungle cycle, where dengue probably evolved (it may have been introduced millennia ago from East Africa, where it also occurs and where it acquired its common name). The Asian *Ae. albopictus*, alias “the tiger mosquito” because of its striped legs, will bite people and does breed in containers around human habitation, but it likes vegetation and generally prefers other animals. Dengue fever has been a ubiquitous childhood disease among the rural population of Southeast Asia. With urbanization during the last century, the disease at first disappeared from the urban population because the vector was not common in towns. With the arrival of *Ae. aegypti* from Africa, however, dengue became epidemic in the cities, where people had not developed the usual adult immunity. *Ae. aegypti* is a supervector. In Southeast Asia it breeds in clear water, in containers ranging from empty tin cans to large urns that store bathwater for the dry season. Dengue became endemic in the urban population, affecting mainly children and giving adults immunity to the local type of virus. It was in this milieu that DHF first emerged.

Dengue fever had been notorious for the aches (its American alias is “break-bone fever”) and the rash it causes, but it had never been a killing disease. What was probably its first pandemic spread appeared in Benjamin Rush’s description of it in Philadelphia in 1780, where it was undoubtedly a visitor from the triangular slave–rum–sugar trade of the Caribbean. There were sporadic epidemics in the 19th and early 20th centuries until mosquito control and the eradication of *Ae. aegypti* from the Caribbean and South America brought decades of repose (even as it ended

epidemics of yellow fever). In recent decades, however, *Ae. aegypti* has again recolonized the Caribbean and much of South America (Gubler & Trent, 1993).

DHF kills more than 15,000 people a year (most of them children), 10,000 of them in India. It usually puts infected people in the hospital when abdominal pain and high fever occur, 2–3 days after infection. There is bleeding into the mucous membranes of the abdomen and into the skin, as well as vomiting. This stage may be followed by a dengue shock syndrome (DSS), which was responsible for most of the early deaths but is better treated today. Intense research has focused on DHF/DSS, and some aspects of the mystery are now clear.

There are several hypotheses about the etiology of DHF. One early one was that while thousands of rhesus monkeys were caged on the docks of Singapore and Manila, en route to U.S. laboratories for development of the poliomyelitis vaccine, wild jungle strains were introduced to the city. Another hypothesis was that the virus had mutated. All four types of the virus were isolated from dying patients, however, and it seemed unlikely that all four types would mutate at the same time. The modified hypothesis is that *Ae. aegypti* is such a supervector that all types of dengue have been able to become more virulent: Transmission opportunities are so excellent that they can chance endangering their hosts more (see the discussion of Ewald's ideas in Chapter 2). The new disease is indubitably associated with *Ae. aegypti*. As the mosquito has spread among coastal towns and along the railroads and roads, DHF has spread with it. In the urban systems of Southeast Asia, squatter slums and people living on boats have greatly increased the supervector's breeding habitat. As rain gutters, storage jars for roof runoff, ant traps around table legs, flowerpots, and garbage heaps have proliferated, so has the mosquito.

The most widely supported hypothesis for two decades has been that DHF is caused by the immune system's overresponding to repetitive infection with dengue. The troop movements of World War II in the Pacific, amidst all armies, refugees, and prisoners, moved strains vigorously among Australia, New Guinea, the Philippines, Japan, Indonesia, and Thailand; this movement ensured that all four serotypes were widespread. As people have migrated to the city from different regions, they have brought all four types of dengue fever into simultaneous circulation. Population mobility and the supervector have combined to increase the chance that a child will be infected with two different types of dengue virus within a relatively short span of time. Some researchers (e.g., Halstead, 1980) believe that the immune system of some children reacts so strongly to a second infection by a related, but different, dengue virus that DSS results. This has defied attempts to develop a vaccine because one may only make an infection worse.

DHF continues to increase and to spread. The initial biennial frequency of the disease has largely disappeared. It occurs throughout the year, but increases during the peak mosquito bite periods associated with the monsoon. Although better treatment has brought down the death rate, tens of thousands of children continue to be hospitalized each year in all the countries of tropical Asia. Following the reestablishment of *Ae. aegypti* in the Caribbean in the 1970s, dengue virus spread there epidemically in the 1980s. In the 1990s DHF/DSS became endemic from Cuba to Venezuela and Brazil. At first DSS was not common, but it has emerged since the

arrival of a Southeast Asian biotype of dengue. One thing not yet understood is the difference between dengue viruses that cause DHF/DSS in the presence of preexisting dengue antibody and those that do not.

Ae. aegypti is common in the U.S. South. There is public health concern that dengue will not only launch some epidemics but become established in Texas or Florida. The introduction of *Ae. albopictus* to Texas in 1985 from a boatload of tires has heightened concern. *Ae. aegypti* is a much more dangerous, anthropophilic vector, but it is essentially a tropical mosquito. *Ae. albopictus* poses a new threat because it survives winter better. In Asia the tiger mosquito occurs from India through China and Japan. In the United States it has already spread from Texas to Pennsylvania and Illinois, as well as throughout the Southeast. It is not as happy in paved cities as *Ae. aegypti*, but it seems to love the suburbs for their vegetation and in many areas has become more common than *Ae. aegypti*. DHF may eventually occur over much of the United States.

Dengue fever has been closely tied to international population mobility from tourism and business. Travelers' information about dengue as well as malaria prophylaxis and other health information is kept current on the CDC website. Only a small fraction of dengue fever infections even today result in DSS. Even without it, however, the substantial loss to work and quality of life is indicated in the DALYs of Table 4.5.

Schistosomiasis

Schistosomiasis, an endemic helminthic disease in at least 73 countries, is the most rapidly spreading serious parasitic disease in the world. More than 200 million people are infected by the schistosome, but the low direct mortality belies the enormous disability, lost work and energy, and increased susceptibility to death from other causes (Table 4.5). Schistosomiasis (sometimes called bilharzia in Africa) is not ordinarily fatal, just debilitating. It is not, strictly speaking, a vectored disease, but it definitely has a natural nidus. A species of snail serves as an *intermediate host*, biologically necessary to its life cycle.

The parasite is a schistosome, a kind of blood fluke (i.e., a flatworm or helminth) that lives as adults in the veins (see Table 4.5, Table 4.6, and Table 1 in Vignette 2.1). Although there are five species, three account for most human disease. *Schistosoma japonicum* and *S. mansoni* inhabit the veins around the intestine; *S. hematobium* inhabits the veins around the bladder. *S. mansoni* occurs in Africa, the Arabian peninsula, and Latin America from the Caribbean to northeastern Brazil. *S. hematobium* occurs in Africa, the Middle East, and small foci in India. There is no significant reservoir for these transmissible diseases except people. *S. japonicum* has mainly occurred in East Asia; today it has been largely eradicated in Japan, but continues to infest the Yangzi River lowlands and lake country, where international concerns about its spread as a result of the Three Gorges Dam continue to be much studied. Schistosomiasis has been found in several pockets along the Mekong River (to which it was introduced by Japanese soldiers), where it has recently been associated with a new species, *S. mekongi*. It also spread during World War II to limited areas in

TABLE 4.6. Some Semi-Vectored Diseases with Intermediate Hosts

Disease	Agent	Main intermediate host(s)	Endemic locale	Comments
<i>Viral</i>				
Rabies	Rhabdovirus	Canidae (foxes, dogs, coyotes, jackals)	Worldwide, except some islands	Also bites of skunks, bats, raccoons: wildlife rabies increasing in United States
<i>Helminthic</i>				
Clonorchiasis (liver fluke)	Fluke: <i>Clonorchis sinensis</i>	Freshwater fish, snails: Amnicolidae	East and Southeast Asia	Human infection from larvae in cysts in raw fish, eggs passed in feces
Hydatidosis	Dog tapeworm: <i>Echinococcus granulosus</i>	Definitive hosts dogs, wolves; intermediate hosts herbivores, dog—sheep—dog cycle most important; also dog—kangaroo—dog, wolf—moose—wolf, dog—cattle—dog	Where dogs are used for herding: Middle East, Australia, Argentina; also Kenya	Human infection from ingestion of eggs from dog fecal contamination
Schistosomiasis (bilharzia)	Fluke: <i>Schistosoma mansoni</i> , <i>S. haematobium</i> , <i>S. japonicum</i>	Snails: <i>Biomphalaria</i> spp. for mansoni, <i>Bulinus</i> for other types	<i>S. japonica</i> in East and Southeast Asia; <i>S. mansoni</i> in Africa, Northeast South America and Caribbean; <i>S. haematobium</i> in Africa, Middle East to India; <i>S. intercalatum</i> in West Africa	Humans are main source; domestic animals involved with <i>S. japonicum</i> ; infection by skin contact with water
Trichinosis	Intestinal roundworm: <i>Trichinella spiralis</i>	Swine; bears, dog, cats	Variable worldwide, depending on pork habits	Human infection from eating raw or undercooked meat

the Philippines and Sulawesi (in Indonesia). In addition to humans, dogs, cats, pigs, water buffalo, field mice, and rats can host the agent; collectively they constitute a persistent reservoir.

Schistosomiasis is a chronic disease that, in cases of intense infection, can be disabling. It weakens people and causes them to die from other things. In places where it occurs, such as the Nile Valley, it frequently infects more than half of the population. The symptoms of the chronic infection result mostly from the passage of enormous numbers of eggs. The male schistosome lives in a groove in the female, where it can continually fertilize her prodigious egg production. Individual worms

produce eggs for several years and have been known to do so for more than 20 years. Schistosome eggs have a “hook” or “tooth” on them to penetrate to the intestines or bladder for egress. Millions of eggs may block blood vessels. Some eggs do not get out quickly and are carried in the blood throughout the body, and so tear at veins, liver, and other tissues. Some worms wander from the veins they normally inhabit and go astray into other organs. In cases of heavy infection (many thousands of worms have been recovered), the passage of the infesting parasite through the liver can injure it. Common symptoms of the chronic disease include fever, headache, pain, and bleeding into the urinary tract or intestine.

From the veins around the intestines, eggs are passed in the feces; from those around the bladder, they are passed in the urine. Hooks on the eggs tear holes to get out of blood vessels and result in bloody urine and other blood loss. Most human waste is deposited into or washed into water. When the schistosome eggs enter water, a form called *miracidium* hatches and swims downward, searching for a snail of the proper genus. The presence of suitable snails has allowed schistosomes to spread to Central and South America; they have thus far failed to find suitable snails in the United States. In China, the usual amphibious nature of the intermediate host snails, which support *S. japonica*, allowed a type of intervention (burying and drowning snails) not possible for the aquatic snails in Africa.

After the parasite multiplies in the snail, it breaks out in a microscopic form called *cercaria* and swims to the surface looking for people. When it encounters bathers, it penetrates their skin and follows a complicated internal path through metamorphosis in the liver to its eventual home in a blood vessel as a sexually reproducing adult trematode. The temperature of the water, the speed of the current, and the hours of sunshine all become critical factors in how long the cercaria survives in the water without a host.

The cultural side of the equation is the water contact behavior of the population. To contract schistosomiasis, people do not need to eat or drink anything contaminated. They merely need to swim, bathe, or do their laundry. In semiarid lands or in dry times of the year, human activity often centers on water sources. If the appropriate snail is present and its habitat is favorable, a traveler or immigrant can introduce the schistosome eggs and begin a cycle. The more clustered the population is, the more intense the cycle of infection can become. As the floods of the great rivers are controlled, as irrigation is extended into arid lands or converted from seasonal basin irrigation systems to perennial systems, and as fertilizer and fish ponds enrich the water, so the habitat of the snails that are intermediate hosts is increased, and the disease spreads in Africa and in South America.

The age structure and activity patterns of the population affect the disease's epidemiology. Farmers in the field do not use latrines. Children cannot be kept from swimming in the canals. Women must do laundry. The time of day and the place of contact in the river or canal expose people to different activity levels of the cercariae and to different levels of snail and water contamination foci. Deep wells with hand pumps to provide water for laundry and bathing have often reduced incidence in women and children. The expansion of irrigated agriculture, on the other hand, has increased occupational exposure in southern and eastern Africa. Incidental foci of

exposure may be things as ordinary as the water basins used for ablutions by Muslims before prayers at a mosque.

Figure 4.7 illustrates intervention points in the life cycle of schistosomiasis. Chemoprophylaxis of the population can kill the adult worms. The drugs are powerful, however, and have many side effects. Usually hospitalization is necessary, and few peasants can afford the time and lost work. The U.S. tourists who get infected can be successfully treated and cured at home, but where a large local population is infected and subject to reinfection, there is no feasible treatment. If all urine and feces were disposed of in a safe manner, contamination of the water could not occur. Given the nature of agricultural work and the activities of rural children, however, complete sanitary disposal is often not feasible. In some places, such as lower Egypt, the water table is so high that latrines provide little separation of waste and water. Fast water, drainage, and concrete linings of ditches that are regularly cleaned can decimate a snail population. Special spillways have been engineered into some African dams to cause such focused floods, but riverine agricultural (recessional)

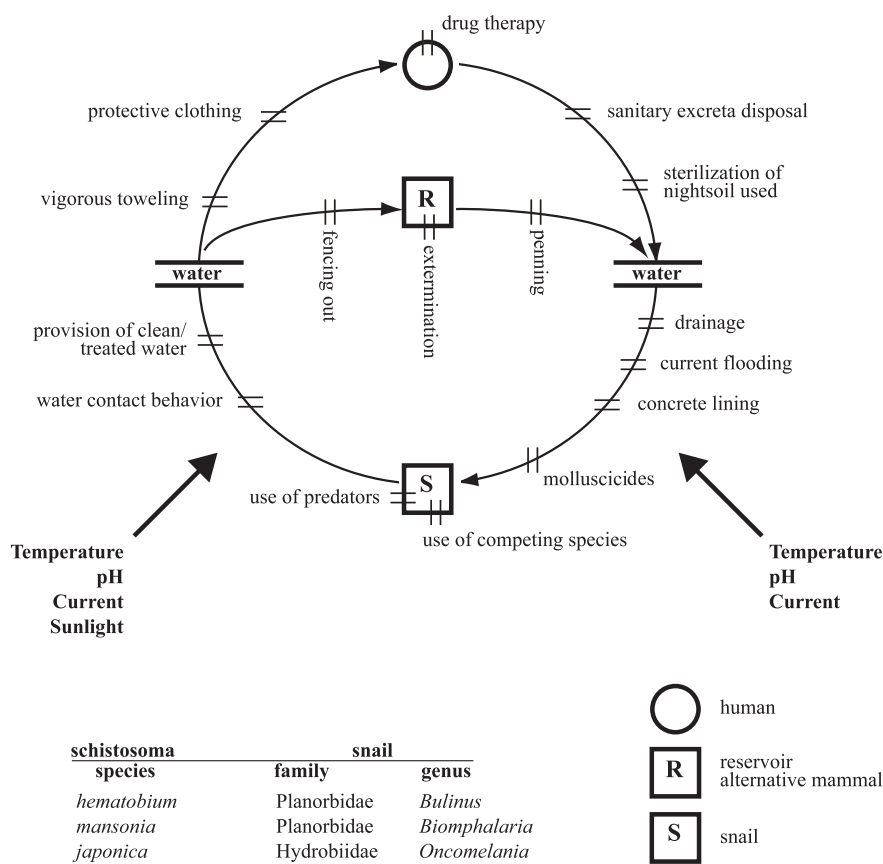


FIGURE 4.7. Schistosomiasis intervention. Each kind of schistosomiasis has a different genus of snail as an intermediate host. Environmental characteristics of bodies of water are important at two points in the schistosome's life.

plantings many miles downstream in another country can be destroyed, too. Molluscicides have been used to protect certain settlements or urban populations. They are expensive, as they continually wash away downstream, where they may accumulate and result in fish kills and other harmful effects.

There has been considerable interest in biological controls. Many animals have forms of schistosomiasis that are not infectious for humans (such as those resulting in “swimmer’s itch” in the United States), but that may infect suitable snails and outcompete the human parasite. Some success with biological competition has been achieved in the laboratory, but dominance seems to be temperature-dependent and transfers poorly to the field. Finally, the water contact behavior of a human population can be modified to some extent if enough is known about the disease hazards. For example, the time of day when certain contact activities occur, such as washing the laundry, can be altered so as to take place when cercariae are less active. Clean, safe water can be supplied for domestic use from tube wells.

The spread of and increase in schistosomiasis are so serious that many kinds of international research are focused on it. Scientists are working on vaccines, on new drug therapies, and on biological competition and predation. Health educators and social scientists are working on modification of population behavior and the design of protective systems. The complex ecology of the disease differs by type of schistosome, by local snail ecology, by local economy, and by local customs. Population mobilization and a labor-intensive war on snails were said to have decreased the disease in Mao’s China. The changes in agricultural modernization and irrigation improvement, with new water impoundments for hydroelectricity, economic liberalization, and growing population mobility (see Chapter 5), seem to work largely against its control. Schistosomiasis is spreading in the different ecologies of the coffee plantations and new settlements of Brazil, as well as in new irrigation works in eastern, western, and southern Africa. The microgeography of the disease can guide local intervention efforts, but aside from the work of Kloos in Ethiopia and the Nile tributaries, little geographic research has been done.

THE CULTURAL ECOLOGY OF TICK-BORNE AND OTHER TRANSMISSIBLE DISEASES

Tick-Borne Diseases

Ticks vector an array of viral and rickettsial diseases, from Russian spring–summer encephalitis to boutonneuse fever in India and South Africa (see Table 4.1). Each has its own characteristic landscape ecology. The landscape epidemiology of central European encephalitis, for example, was studied by Wellmer and Jusat (1981). That encephalitis is caused by a togavirus vectored mainly by a common tick, *Ixodes ricinus*. Wellmer and Jusat delimited the temperature, humidity, precipitation cycles, vegetation, animal associations, and other environmental conditions needed for its maintenance. They found that lower mountain elevations with dense and diversified vegetation are required by the disease system, which is limited by such factors as an annual isotherm of 46°F (8°C), homogeneous vegetation such as occurs in

pine forest biomes, and light soils with humus for sufficient ground moisture. They determined that the foci are limited to a few square kilometers each, and that a density of only 1 virus-infected tick per 1,000 ticks was required to maintain a natural nidus. The German geographers called this a *standortraum* (multifactor location space). They mapped and analyzed the geoecological conditions for the repeated infection of specific regions by infected ticks and the recurrent infections of people who entered the natural nidus. (See Vignette 4.1, on field mapping.) American studies have not been so specific, although there are two excellent candidate diseases to research: Rocky Mountain spotted fever (RMSF) and Lyme disease. Both of these tick-borne serious diseases are increasing in prevalence in the United States, yet their distribution pattern is disjunct. Lyme disease occurs, but mysteriously has not become prevalent in the heartland of RMSF. What makes them ecologically distinct? Is it only a matter of time until they merge?

Tick vectors pose a four-stage ecological puzzle. RMSF is generally vectored by the “dog tick,” *Dermacentor variabilis*, and Lyme disease by the “deer tick,” *Ixodes scapularis* (now considered to include *I. dammini* in the South). There are important differences between the two disease systems, but they share basic characteristics. Three separate blood meals are required for the life cycle. The adult female lays up to 10,000 eggs, which hatch into six-legged larvae vernacularly called “seed ticks.” These must feed for a week on an animal. These animals are usually small and close to the ground; mice are favorites, as well as birds and reptiles. Then the larvae metamorphose into eight-legged nymphs. The nymphs climb grass and bushes and “quest” vigorously, waving their legs at vibration, heat, or carbon dioxide, to latch onto an animal host and get a second blood meal. After feeding, they drop off and change into adult ticks. According to climate and species, larvae and nymphs as well as adults may crawl under leaves or other cover to spend the winter. In spring they emerge hungry for another meal and, as they feed, infect new rodents before new tick larvae hatch. Adult ticks need the third blood meal to reproduce. They may also feed on rodents, rabbits, opossums, or even birds (useful for dissemination), but they tend to quest higher in vegetation and to be more discriminating in what they latch on to. They prefer raccoons, foxes, deer, or other large mammals such as dogs, which can disperse them over a wider territory. Each of three blood meals provides an opportunity for the tick to become infected and in turn infect its next host. When a large animal is infected, it can become an amplifying host, able with high viremia to infect many arthropods that feed on it.

Here RMSF and Lyme disease diverge, however, in an important way. The agent of Lyme is a spirochete, a bacteria, *Borrelia burgdorferi* (see Vignette 2.1, Table 1). Each larva hatches out “innocent” and must begin the cycle of infection again with a blood meal from an infected host. If the host it feeds on is not infected, then the nymph stage is not infective. If the second blood host is not infected, then the adult tick seeking its third meal is not infective. The enzootic disease is maintained by 25–50% Lyme infection rates of nymphs. In great contrast, the rickettsial agent of RMSF, *Rickettsia rickettsii*, can pass transovarially in ticks to the new generation. Larvae are born infected and able to infect a host. Even infection rates lower than

10% among ticks seem adequate to maintain the disease system because the tick itself through its generations is a reservoir.

RMSF, sometimes called tick-borne typhus, was first identified in Montana and Idaho (hence its name), but it is most common today in the Piedmont region of the U.S. Southeast. It has been isolated from the chipmunk, meadow vole, pine vole, white-footed mouse, cotton rat, cottontail rabbit, opossum, and snowshoe hare, and serological (antibody) evidence for infection has been found in many other mammals (Burgdorfer, 1980). Better publicity has enabled RMSF to be diagnosed earlier and treated with antibiotics, sometimes even before diagnosis. Although this complicates statistical reporting, it has brought the case mortality down considerably. The onset of the disease involves fever, chills, aches, nausea, and a rash that spreads from the wrists and ankles. Many of these symptoms are common to other diseases and are easily misdiagnosed without antibody identification. Many, probably most, cases are so mild they are asymptomatic. Of those who do develop clinical symptoms and are not promptly treated, however, 6–15% may die.

RMSF has been increasing in the Southeast, but it is hard to determine how much of the increase is due to greater incidence and how much to better reporting. The increase and focus of the disease is attributable to the rapid extension of suburbs into the wooded, open-field succession habitat of the dog tick. The Piedmont is one of the most rapidly growing regions within the southeastern Sunbelt. The dispersed settlements of new suburban houses, along with parks, jogging paths, and family dog sojourns, are *in* the natural nidus of RMSF. When human residential land use and recreation are extended into the woods, infected ticks can be brought home.

The first U.S. cases of Lyme disease were identified in 1962 in Lyme, Connecticut. It is not certain whether Lyme disease was relocated from Europe, where the spirochete has been well established, or had an indigenous variant here. It causes initial symptoms of the generic headache/bodyache/fatigue variety. If not treated, serious symptoms of crippling arthritis and heart abnormalities or neurological complications can develop months or years later. Lyme disease already has spread to the Pacific Coast and the Gulf Coast. It has established three main foci: New England and the surrounding area, the upper Midwest, and the Pacific Northwest (Herrington, 1995). In the East, the reforestation that has occurred since midcentury has brought the white-tailed deer and other forest animals back to levels not seen in two centuries. Deer, raccoon, and other animals favored by the deer tick (which is somewhat misnamed, as there are ticks that spend their entire life on deer, whereas *I. scapularis* needs three separate hosts) are those most quickly domiciliating to suburban residential areas. The forest fringe–suburban grass intersection can support an intense nidus, with more than half the adult deer ticks infected. Diffusion has occurred along deer paths in river corridors and electric pylon corridors into suburban areas. The infection has, in ticks, probably ridden birds to jump over large areas and spread across the country. The West Coast nidus is not so dangerous for people. The main tick involved there, *I. neotomae*, is the most infected but feeds chiefly on woodland rats, not people. A second tick—*I. pacificus*, which feeds on many differ-

ent hosts, including rats and people—is needed to connect the nidus to humans. Because of its diverse diet, as it were, the prevalence of infection in this linking tick is low, however, and so is the risk to people.

Why has Lyme disease not become highly prevalent in tick heaven, the Piedmont where RMSF thrives? The ecology of transmission is in the details. The southern deer tick, *I. dammini*, in its larval stage also commonly eats lizards and snakes, and they do not carry the bacterial infection. Therefore, while a nymph may occasionally bite an infected host, perhaps a migrant, and become infected, and Lyme transmission does occur now; nevertheless the nymphs do not reach the population infection level necessary to maintain a natural nidus of the disease. Since the dog tick is born infected with the RMSF rickettsia, it does. (Where would they diverge on the transmission diagram depicted in Figure 4.8?)

The cultural interventions and buffers for the two systems are similar. (Where would you add them to Figure 4.8?) Both diseases can be treated successfully with antibiotics if caught early. Theoretically, reservoir hosts can be destroyed; even if people wanted to exterminate the local deer, however, taking on the white-footed mouse would seem impossible. Almost uniquely among vectors, ticks can serve as reservoirs that survive winters and even pass infections generationally. Residential efforts to interrupt disease transmission usually focus on the tick. Because ticks get dehydrated on their questing roosts in the summer sun, by late afternoon they usually have climbed down to the cooler and humid ground. Therefore, walking into the nidus at the end of the day is protective. The different stages of the tick can be killed with insecticide, which can be sprayed in the yard, or put on the dog in collars. Around the house, keeping the lawn mowed short and brush out of the dog's fenced yard helps keep the ticks away. For a human, body buffers while walking into the nidus include long white pants (for easy viewing) tucked, even taped, into socks and boots; long sleeves; and repellent, perhaps. Because it takes hours of feeding for the tick to pass the infection, the best single body buffer is careful examination of children and dogs, as well as oneself, a couple of times a day. For the

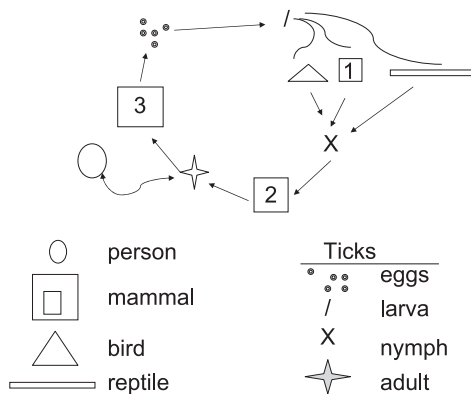


FIGURE 4.8. Tick-borne disease transmission cycles.

pinhead-size deer tick nymph, this is, of course, easier said than done. In a study in Maryland, the best landscape predictor of Lyme disease risk has been identified in a remote sensing/GIS study (Jackson, Hilborn, & Thomas, 2006) as the proportion of land cover that is forest–grass edge. Landscape planning guidelines have then suggested that “casual peridomestic contact with tick and host habitat” could be reduced by clustered forest and herbaceous cover, as opposed to current high forest–herbaceous interspersion. This landscape change would minimize Lyme disease risk in low-density residential areas.

The landscape epidemiology of both diseases rests on the tick’s habitat, abundance of small mammals, winter shelter, dehydrating sun, bird predators, and such. RMSF in the past has seemed to involve the intrusion of people into the natural nidus of the disease (May’s *silent zone*). People and their dogs walked into the old field habitats and hunted on the forest’s edge. Lyme disease, and today RMSF, can be said to result from the expansion of the nidus to include people, as deer have moved into suburbia. (This distinction, if it ever were true, is no longer clear, as houses are pushed into the forest’s edge; displaced deer *are* then in suburbia.)

There are both great opportunities and a great need for geographers to delimit the natural nidi of tick-borne diseases in North America by using GIS-integrated remote sensing and human data. In a study about predicting the risk of Lyme disease in Wisconsin, for example, habitat suitability for *I. scapularis* was determined by using maps of soil orders (determined years ago) to get at such characteristics as soil acidity, coniferous–deciduous forest cover, excessive moisture content, fungi adverse to ticks, abundance of winter leaf litter, and the like, to estimate densities of ticks in specific habitat (Guerra et al., 2002). Statistically significant factors were then used in a GIS to generate a risk map predicting the presence of the tick vector. The extensive area of suitable habitat in western Wisconsin corresponded to areas of increased incidence of human Lyme disease. It explained the disease’s axis of rapid expansion. Locating human settlement would further have identified the risk of expansion into new areas. Pavlovsky would have been appreciative.

There are other, almost unstudied, serious tick-borne diseases in the U.S. and around the world. Relapsing fever, for example, has been an internationally notifiable disease for a century because of its epidemic, louse-borne manifestation. Its endemic form, from the Mediterranean coast to Central Asia to the Andean chain to the U.S. and Canadian West Coast, is transmitted by argasid ticks. These *Ornithodoros* ticks live up to 20 years and may have more than one nymph stage. The infection produces many carrier states. Babesiosis, a disease that resembles malaria but is transmitted by ticks, has more than 100 variants around the world. Mostly associated with animals, especially rodents, in the United States it has been identified in dogs—and in people who have no spleens! The Lone Star tick (*Amblyomma americanum*) has been known to occur over much of the U.S. South. It has been identified as the vector of tularemia, and also of the recently identified human disease ehrlichiosis, especially in the southeastern U.S. The disease agent also causes Senetsu fever in Japan. So little is known about the ecology and distribution of so many vectored diseases—and there is so much work to do.

Other Transmissible Diseases

It is not possible in one chapter to discuss many of even the most serious transmissible diseases (see Table 4.5). A few comments have been made on several basic disease ecologies earlier in this chapter. Onchocerciasis, or river blindness, is a filarial disease that has for centuries denied some of the most fertile alluvial land to agricultural settlement in West Africa and parts of Central America. The WHO's intensive multinational assault on the West African nidus of onchocerciasis is one of the success stories of the past few decades, and it is possible that this disease will be eradicated. The development of the antihelminthic prophylactic ivermectin, originally to protect the American dog from mosquito-borne heartworms, had been a big part of the successful program. The WHO maintained for years on its home website an explication of the GIS that was used to locate villages, individuals who needed treatment, habitat, and interventions and direct treatments. Lymphatic filariasis (elephantiasis) occurs throughout tropical Asia and Africa, and in India and elsewhere it is heading to the city with the *Culex* mosquito. Trypanosomiasis, or African sleeping sickness, affects all efforts at hydroelectric, agricultural, and transportation development in Africa, to which it is confined by its uniquely occurring vector, the tsetse fly. Chagas disease, or American trypanosomiasis, has spread more in South America with every settlement that pushes into the Amazon (see Plate 10). It has also spread to the cities with migrants, living in the cracks of poor housing, and to the United States. Although its vector, the reduviid bug, occurs to a very limited extent in the U.S. Southwest, the disease is a new national risk because it threatens the blood supply. Innumerable arboviruses occur whose implications for human health are not even known. New arboviruses, like West Nile encephalitis, have been introduced (relocation diffusion). The diseases identified by the tables in this chapter are all suitable subjects for landscape epidemiology analysis.

REGIONALIZATION

At the beginning of this chapter, we have stated that the study of landscape epidemiology involves four types of regions whose dynamics and locations impart patterns to the distribution of diseases. Let us review some aspects of nidal diseases in these terms.

Despite the transportation and mobility transformations of the last few centuries, many nidal diseases are still limited to their realm of evolution. Trypanosomiasis occurs only in Africa because the tsetse fly occurs only there. Schistosomiasis and onchocerciasis must have been introduced to the United States thousands of times, but because the necessary species of snails and blackflies are absent, the diseases cannot establish themselves (although suitable vector substitutes were found in tropical America). Plague, on the other hand, has found enough ecological parallelism in various species of burrowing rodents to have become thoroughly established (by relocation diffusion) in all the grassland biomes around the world.

Biomes, encompassing as they do biotic expression of climate, topography, and soils, broadly categorize the conditions under which a wide variety of nidal diseases flourish, survive, or occur periodically. Mosquito-borne diseases are much more complexly established and intensively transmitted in rain forest and subtropical rice lands than in deciduous forest; they cannot be endemic in coniferous regions unless bird migrations or other mediums connect biomes. The disappearance of a disease at a certain elevation in a country's highlands is often the first clue that the disease is transmitted by mosquitoes. Trypanosomiasis disappears as one journeys from the wet savanna to the dry. Mountains and high-latitude lands are the favored biomes for flea-, louse-, and tick-borne diseases.

The culture realm determines many of the ways in which a habitat is modified and a population is exposed to or protected from disease. In May's (1958a) example, the lowland Vietnamese who built their houses on the ground were part of the Chinese culture realm, not the Southeast Asian realm in which houses are built on stilts. The dispersed settlement form helped prevent malarial transmission in parts of the United States—an advantage the nuclear European villages did not have during many centuries of endemic malaria. Types of clothes and construction materials are cultural characteristics that have profoundly affected the importance of disease transmitted by sand flies, reduviid bugs, and lice.

Realm of evolution, biome, and culture realm overlap to delimit broad regions of disease occurrence, but it is at the micro scale that one can determine the limiting factors of disease transmission and specify intervention points. A transmissible disease does not occur uniformly across the earth's surface. The dynamics of its occurrence from season to season and year to year are affected foot by foot by the immediate weather, soils, topography, land use, and flora and fauna. People can destroy a nidus of disease by exterminating the reservoir or plowing up the vector's habitat and not know what they are doing. They can also bring in water or clear vegetation and create conditions conducive to intense disease transmission without planning for those consequences. Mapping the exact location and conditions where disease transmission is marginal or most intense is of great significance for breaking transmission paths or protecting the population (Vignette 4.1).

Among the major developments in regionalization, analytical modeling, epidemic surveillance, and prediction of vectored diseases in this new century are the ongoing development and application of GIS technology and methodology (see the discussion in Chapter 3). Satellite imagery (remote sensing), field surveys of land and people, and epidemiological data can be combined to detect such habitat changes as extent of flooding (and mosquito outbreaks) or to specify the vegetation, slope, and topographic conditions of a targeted nidus (Mott, Nuttall, Desjeux, & Cattand, 1995). Efforts are being made worldwide to predict outbreaks of Rift Valley fever from water on the land, to give early warning of major crop failures and famine, to track locust breeding conditions and outbreak paths, and in general to monitor and/or model numerous habitat change and vectored disease relationships. At a different scale, almost every U.S. state government now has a GIS web-site providing information on the incidence in humans, animals, and birds of WNV,

rabies, and other diseases. There are serious scale issues that have not yet been worked out, such as how to use vegetation or slope information by the acre when tick density varies by the foot and population data are at a county scale. Nevertheless, remote sensing is becoming vital to landscape analysis and to international relief and intervention. Vectored diseases such as dengue and malaria have also been targets of much concern and speculation about the impact of global warming. Research efforts are being made to develop simulations using GIS. The WHO is now trying to use its website to encourage such applications and to make GIS software and analytical programs accessible to poor countries. The part of its website devoted to control of tropical disease connects to many studies as examples, such as WHO's planning and implementation of eradication efforts in West Africa against onchocerciasis and dracunculiasis (guinea worm). The WHO website presents examples of methods of integrating field survey; satellite usage; multiple land use covers of settlement, agriculture, streams, and topography; and soils, health service, and more, through use of a GIS. It also encourages the adoption of this methodology elsewhere.

ECOLOGICAL COMPLICATIONS

Vectored diseases with characteristic landscapes (i.e., nidal diseases) are especially sensitive to human manipulation of biomes and species. Simple ecosystems are especially prone to epidemic propensity, and they are what humans make.

Odum (1978) and others have described the characteristics (Table 4.7) of mature and youthful ecosystems. These characteristics require some elaboration. The most mature ecosystem, a tropical rain forest, has an enormous variety of species compet-

TABLE 4.7. Some Generalized Characteristics of Ecosystems

	Youthful	Mature
Number of species	Few	Many
Number of species individuals	Many	Few
Niche	Generalists	Very specialized
Growth	Rapid	Slow
Reproduction	Frequent, prolific	Infrequent, limited
Dissemination	Wide, fast	Local, slow
Size tendency	Small	Large
Information, feedback	Low	High
Food chains	Linear	Complex
Nutrient conservation	Low	High
Entropy	High	Low
Overall stability	Low	High
Prototype animals	Mice, sparrows	Elephant, chimpanzee
Prototype plants	Dandelion, pine	Mahogany tree

Note. Information derived, for the most part, from Odum (1978).

ing. Each species tends to be specialized; to be limited in the number of individuals who can live in its specialized niche; and to grow slowly, reproduce slowly, and disseminate slowly. The different life forms are regulated through feedback stemming from competition, predation, symbiosis, and commensalism. This feedback creates a resilient system that dampens fluctuations while maximizing the use of energy in life support. The system has taken millions of years to develop its complexities. It is, however, highly vulnerable to human intervention. Once large areas of rain forest are destroyed, as by bulldozers, reestablishment of the intricate relationships among its slow-reproducing, slow-growing, and poorly disseminating life forms will take much more than a human lifetime. In contrast, a young or simple ecosystem such as a recently colonized landslide has a few generalized species that are able to withstand extremes of weather and soil conditions. They must reproduce and grow quickly and disseminate easily. Although there are few species, each species has enormous numbers of individuals. The young ecosystems are subject to great fluctuations, as there are few sources of feedback to stabilize them. A flock of birds, a herd of caribou, or a snake passing through can be devastating for plants and rodents in a young ecosystem. Only the sheer numbers, rapid reproduction, and dissemination of species maintain such simple systems.

One of the most profound human influences on the earth has been *ecological simplification* (see Figure 4.9). People create the simplest possible ecosystem for agriculture. When farmers plant a crop, they do not want competition from any bird, insect, or animal; they want the food all for themselves. They want the crop to grow rapidly, to produce lots of seed, and to be removed by harvest. Because there are so few of the feedback balances that exist in complex ecosystems, rats, finches, or the wrong kind of rain can also be devastating.

The simple, young ecosystems that people create with their houses, road clearings, farms, meadows, and herds of domesticated animals are ideal for epidemics. When arthropods or disease agents inimical to human health are present in a simple ecosystem, they exist in large, explosively growing numbers with few controls. Consider a rain forest in which there are several species of *Anopheles* mosquitoes capable of transmitting malaria, but they are competing with hundreds of other mosquito species (Figure 4.5). Each of the vector species, one of many species that might bite a person passing by, exists in relatively small numbers; it has many species of animals to feed upon that do not carry human malaria. Malaria occasionally gets transmitted in such a landscape, but more often it does not, and no focus is developed. When trees are felled, exposing streams to sunlight, conditions similar to a young ecosystem are created. The species of *Anopheles* that breed in those conditions have large numbers of individuals, reproduce rapidly and so take blood meals frequently, have few controls, and create epidemic malaria. This has been and continues to be the outcome as people push to the agricultural frontiers and convert forest to agribusiness, as discussed at length in Chapter 5.

A medical anthropologist (Dunn, 1972) once directly demonstrated the perils of ecological simplification for human health. Forest-dwelling aborigines in Malaysia had many kinds of intestinal parasites: several varieties of roundworm, tapeworm, whipworm, flukes, and hookworm. People had only a light infection of any one

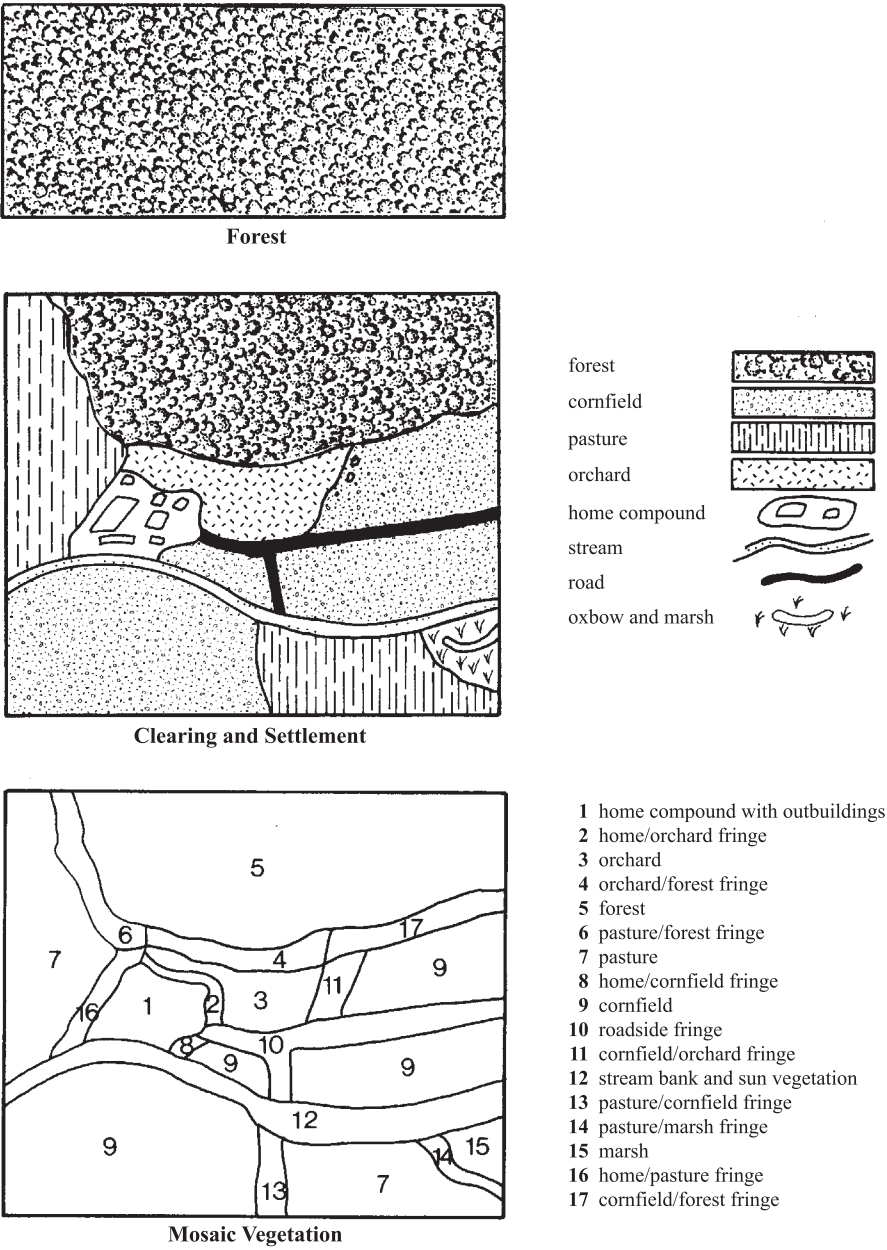


FIGURE 4.9. Ecosystem simplification and multiplication. Epidemic capacity is increased when a complex ecosystem, characterized by diversity of species with few individuals, is replaced by a mosaic of habitats, each characterized by a few species with many individuals.

parasite, however, because they move frequently and foci did not build up. Under these circumstances the parasitic infections seldom resulted in disease. When the people moved to the fringes of towns and became sedentary, however, the ecological conditions were simplified. Only a few species of parasitic worms could exist. An intense focus of their transmission was formed, resulting in high levels of infection involving many individuals of each of those few species. So many hookworms caused anemia; so many roundworms caused chronic diarrhea. The helminthic infections became diseases.

Humans have created intense transmission of and intimate human exposure to animal diseases by domesticating and herding animals. Animal intestinal worms and fleas, vectored viral diseases, and salmonella infections have become part of the human environment. Besides intentional domestications, humans have also domiciliated many animals and arthropods. Audy (1965) used the term *domiciliated* for creatures that have become closely adapted to living with people and depend on people for their great numbers and success, but were not deliberately domesticated and are not kept by people for any purpose. These domiciliated animals include house mice, brown and black rats, bedbugs, and *Aedes aegypti*. The human louse is so domiciliated that it would become extinct without humans. The intimacy of these species with humans is so great, and the house or barn ecosystem so simple, that when a disease that can be vectored or passed by herded animals or domiciliated arthropods is introduced, the intensity of the foci can produce incidence rates like those of the black death.

In Figure 4.9, a virgin forest is cleared by a farmer for cultivation and pasture. What had been a relatively complex ecosystem is replaced by a series of simple ecosystems. The risk of epidemic inherent in a simple ecosystem is multiplied by the number of such systems located in this small area. One additional complication has been introduced: The fringe between two habitats is itself a habitat conducive to passing infection between ecosystems. For example, birds and rodents may shelter in the edge of the woods and venture out to exploit the cornfield. Birds exposed to forest mosquitoes could pass the encephalitis virus to field mosquitoes, which in turn infect field hands. Mice, raccoons, and deer on the forest edge/grass fringe of suburban lots provide such an example in Lyme disease, discussed above. In the example of African yellow fever, discussed earlier in this chapter, *Aedes simpsoni* breed in the tendrils of banana trees that have been planted in the forest fringe. Because they bite both forest monkeys and village humans, they are able to pass the virus repeatedly through people to the settlement-dwelling *Ae. Aegypti*. Not only can such a focus be intense; it can be very difficult to eradicate permanently.

Development obviously changes the ecosystem in ways that affect disease transmission. The potential for disease introduction and human exposure is further altered by changes in human mobility. The following chapter discusses ecological simplification, mobility change, and some of their consequences for land development and population settlements. The landscape changes of the vectored disease systems described in this chapter are at the core of emerging diseases and new risks.

CONCLUSION

North American geographers are not generally familiar with the approach of landscape epidemiology. They are more used to analyzing the patterns within a city; using socioeconomic, ethnic, and demographic factors to explain the occurrence of phenomena such as homicide, tuberculosis, or heart attacks; and then regionalizing to determine the most efficient locations for treatment, given the incidence patterns (see later chapters). Biogeographers and agricultural geographers are familiar with the ecological processes and characterizations, but are generally unfamiliar with human disease systems. Asking many of the same geographic questions but using the approach of disease ecology, medical geographers can analyze the biogeography of nidal disease systems to specify location and system intervention points that will promote control or eradication, and to predict the risk created by landscape changes of settlement and development. Much more can be done to minimize transmission of nidal diseases and to maximize health education, vaccination, and regulatory protection of targeted groups.

REFERENCES

- Ackerknecht, E. H. (1945). *Malaria in the upper Mississippi Valley 1760–1900*. Baltimore: Johns Hopkins University Press.
- Audy, J. R. (1965). Types of human influence on natural foci of disease. In B. Rosicky & K. Heyberger (Eds.), *Theoretical questions of natural foci of disease: Proceedings of a symposium* (pp. 245–253). Prague: Czechoslovak Academy of Science.
- Benenson, A. S. (Ed.). (1995). *Control of communicable diseases in man* (16th ed.). Washington, DC: American Public Health Association.
- Burgdorfer, W. (1980). Spotted fever-group diseases. In J. H. Steele (Ed.), *Bacterial, rickettsial, and mycotic disease* (Vol. 2, Section A, pp. 279–302) (CRC Handbook Series in Zoonoses). Boca Raton, FL: CRC Press.
- Calhoun, J. B. (1979). *The ecology of the Norway rat*. Bethesda, MD: U.S. Department of Health, Education and Welfare.
- Carter, H. R. (1919). The malaria problem of the South. *Public Health Reports*, 34, 1927–1935.
- Childs, S. J. R. (1949). *Malaria and colonization in the Carolina low country, 1526–1696* (Johns Hopkins University Studies in Historical and Political Science, Series 58, No. 1). Baltimore: Johns Hopkins University Press.
- Daniels, T. J., & Falco, R. C. (1989). The Lyme disease invasion. *Natural History*, 7, 4–10.
- Drake, D. (1854). *A systematic treatise, historical, etiological and practical, on the principal diseases of the interior valley of North America*. New York: Lenox Hill.
- Dunn, F. L. (1972). Intestinal parasitism in Malayan aborigines (Orang Asli). *Bulletin of the World Health Organization*, 46, 99–113.
- Farid, M. A. (1980). Round table: The malaria programme—from euphoria to anarchy. *World Health Forum*, 1, 8–33.
- Gubler, D. J., & Trent, D. W. (1993). Emergence of epidemic dengue/dengue hemorrhagic fever as a public health problem in the Americas. *Infectious Agents and Disease*, 2, 383–393.
- Guerra, M., Walker, E., Jones, C., Paskewitz, S., Cortinas, M. R., Stancil, A., et al. (2002). Predicting the risk of Lyme disease: Habitat suitability for *Ixodes scapularis* in the north central United States. *Emerging Infectious Diseases*, 8(3), 289–298.
- Halstead, S. B. (1980). Dengue haemorrhagic fever: A public health problem and a field of research. *Bulletin of the World Health Organization*, 58, 1–21.

- Herrington, J. E. (1995). An update on Lyme disease. *Health and Environmental Digest*, 9, 29–32.
- Hunter, J. M. (1980). Strategies for the control of river blindness. In M. S. Meade (Ed.), *Conceptual and methodological issues in medical geography* (pp. 38–76). Chapel Hill: University of North Carolina, Department of Geography.
- Jackson, L. E., Hilborn, E. D., & Thomas, J. C. (2006). Toward landscape design guidelines for reducing Lyme disease risk. *International Journal of Epidemiology*, 35, 315–322.
- Jusatz, H. J. (Ed.). (1968–1980). *Medizinische landerkunde* (Vols. 1–6) (Geomedical Monograph Series). Berlin: Springer-Verlag.
- Kucheruk, V. V. (1965). Problems of palaeo-genesis of natural pest foci with reference to the history of rodent fauna. In A. N. Formasov (Ed.), *Fauna and ecology of the rodents*. Moscow: Moscow Society of Naturalists.
- Markovin, A. P. (1962). Historical sketch of the development of Soviet medical geography. *Soviet Geography: Review and Translation*, 3, 3–19.
- May, J. M. (1958a). *The ecology of human disease*. New York: MD Publications.
- May, J. M. (1958b). *Studies in disease ecology*. New York: Hafner.
- McNeill, W. H. (1976). *Plagues and peoples*. New York: Doubleday.
- Meade, M. S. (1976a). Land development and human health in west Malaysia. *Annals of the Association of American Geographers*, 66, 428–439.
- Meade, M. S. (1976b). A new disease in Southeast Asia: Man's creation of dengue haemorrhagic fever. *Pacific Viewpoint*, 17, 133–146.
- Meade, M. S. (1977). Medical geography as human ecology: The dimension of population movement. *Geographical Review*, 67, 379–393.
- Meade, M. S. (1980). The rise and demise of malaria: Some reflections on southern landscape. *Southeastern Geographer*, 20, 77–99.
- Mott, K. E., Nuttall, I., Desjeux, P., & Cattand, P. (1995). New geographical approaches to control of some parasitic zoonoses. *Bulletin of the World Health Organization*, 73, 247–257.
- Odum, E. P. (1978). *Ecology: The link between the natural and the social sciences* (2nd ed.). New York: Holt, Rinehart & Winston.
- Pavlovsky, E. N. (1966). *The natural nidity of transmissible disease* (N. D. Levine, Ed. & Trans.). Urbana: University of Illinois Press.
- Pollitzer, R. (1954). *Plague* (Monograph Series No. 22). Geneva: World Health Organization.
- Prothero, R. M. (1961). Population movements and problems of malaria eradication in Africa. *Bulletin of the World Health Organization*, 24, 405–425.
- Rodenwaldt, E., & Jusatz, H. J. (Eds.). (1952–1961). *Welt-Seuchen atlas*. Hamburg, Germany: Falk.
- Sandosham, A. A. (1970). Malaria in rural Malaya. *Medical Journal of Malaya*, 24, 221–226.
- Shell, E. R. (1997). Resurgence of a deadly disease. *Atlantic Monthly*, 280, 45–60.
- Stevens, J. (1977). American mobilization for the conquest of malaria in the United States. *Journal of the National Malaria Society*, 3, 7–10.
- U.S. Census Bureau. (1890). *1890 census of population*. Washington, DC: U.S. Government Printing Office.
- Weil, C., & Kvale, K. M. (1985). Current research on geographical aspects of schistosomiasis. *Geographical Review*, 75, 186–216.
- Wellmer, H., & Jusatz, H. J. (1981). Geoecological analysis of the spread of tick-borne encephalitis in Central Europe. *Social Science and Medicine*, 5D, 159–162.
- Wills, C. (1996). *Yellow fever, black goddess: The coevolution of people and plagues*. Reading, MA: Addison-Wesley.
- World Health Organization (WHO). (2004). *The world health report 2004*. Geneva: Author.

FURTHER READING

- Anderson, R. M., & May, R. M. (1991). *Infectious diseases of humans: Dynamics and control*. Oxford, UK: Oxford University Press.

- Bruce-Chwatt, L. J. (1985). *Essential malariology* (2nd ed.). New York: Wiley.
- Crosby, A. (1972). *The Columbian exchange: Biological consequences of 1492*. Westport, CT: Greenwood Press.
- Crosby, A. (1986). *Ecological imperialism: The biological expansion of Europe, 900–1900*. Cambridge, UK: Cambridge University Press.
- Diamond, J. (1997). *Guns, germs, and steel*. New York: Norton.
- Fonaroff, L. S. (1968). Man and malaria in Trinidad: Ecological perspectives on a changing health hazard. *Annals of the Association of American Geographers*, 58, 526–556.
- Garrett, L. (1994). *The coming plague: Newly emerging diseases in a world out of balance*. New York: Farrar, Straus & Giroux.
- Gottfried, R. S. (1983). *The black death*. New York: Free Press.
- Guerrant, R. L., Walker, D. H., & Weller, P. F. (Eds.). (2001). *Essentials of tropical infectious diseases*. New York: Churchill Livingstone.
- Haddock, K. C. (1981). Control of schistosomiasis: The Puerto Rican experience. *Social Science and Medicine*, 15D, 501–514.
- Harrison, G. (1978). *Mosquitoes, malaria and man: A history of the hostilities since 1880*. New York: Dutton.
- Harwood, R. F., & James, M. T. (1983). *Entomology in human and animal health* (7th ed.). New York: Macmillan.
- Houle, J. S. (1991). *Dermacentor variabilis in space*. Unpublished master's thesis, University of North Carolina at Chapel Hill.
- Hunter, J. M. (1990). Bot-fly maggot infestation in Latin America. *Geographical Review*, 80, 382–398.
- Hunter, J. M. (1993). Elephantiasis: A disease of development in Northeast Ghana. *Social Science and Medicine*, 35, 627–649.
- Knight, G. (1971). The ecology of African sleeping sickness. *Annals of the Association of American Geography*, 61, 23–44.
- Learmonth, A. (1988). *Disease ecology*. Oxford, UK: Blackwell.
- Learmonth, A. T. A. (1977). Malaria. In G. M. Howe (Ed.), *A world geography of human diseases* (pp. 61–108). New York: Academic Press.
- Lederberg, J., Shope, R. E., Stanley, C., & Oaks, J. (1992). *Emerging infections*. Washington, DC: National Academy Press.
- Maupin, G. O., Fish, D., Zultowski, J., Campus, E., & Piesman, J. (1991). Landscape ecology of Lyme disease in a residential area of Westchester, N.Y. *American Journal of Epidemiology*, 133, 1105–1113.
- May, R. M. (1994). Chagas disease in changing environments. In B. Cartledge (Ed.), *Health and the environment*. Oxford, UK: Oxford University Press.
- Moulton, F. R. (Ed.). (1941). *A symposium on human malaria*. Washington, DC: American Association for the Advancement of Science.
- Palka, E. J., & Crawford, T. W. (1994). North Carolina: Natural nidus for Rocky Mountain spotted fever. *North Carolina Geographer*, 3, 1–16.
- Prothero, R. M. (1963). Population mobility and trypanosomiasis in Africa. *Bulletin of the World Health Organization*, 28, 615–626.
- Pyle, G. F., & Cook, R. M. (1978). Environmental risk factors of California encephalitis in man. *Geographical Review*, 68, 157–170.
- Rahn, D. W., & Evans, J. (Eds.). (1998). *Lyme disease*. Philadelphia: American College of Physicians.
- Rodriguez, A. D., Rodriguez, M. H., Hernandez, J. E., Dister, S. W., Beck, L. A., Rejmankora, E., et al. (1996). Landscape surrounding human settlements and *Anopheles albimanus* (Diptera: Culicidae) abundance in southern Chiapas, Mexico. *Journal of Medical Entomology*, 33, 39–48.
- Rosicky, B., & Heyberger, K. (Eds.). (1965). *Theoretical questions of natural foci of disease: Proceedings of a symposium*. Prague: Czechoslovak Academy of Science.

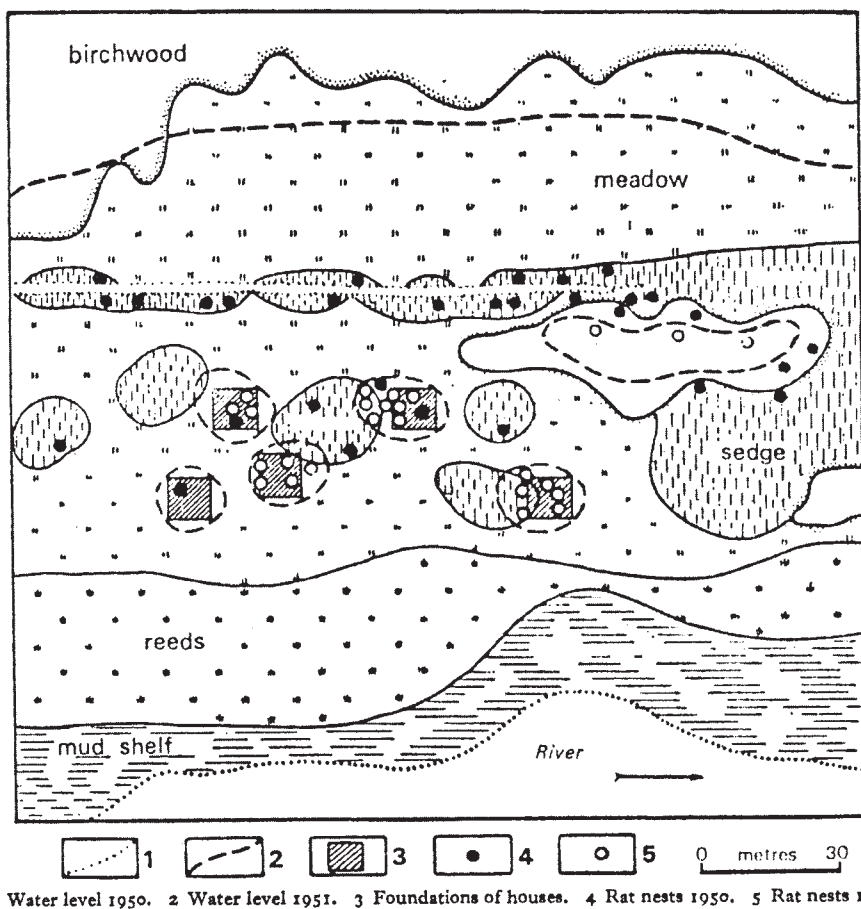
- Sauer, C. O. (1925). The morphology of landscape. *University of California Publications in Geography*, 2, 19–53.
- Schulze, T. L., Taylor, R. C., Taylor, G. C., & Bosler, E. M. (1991). Lyme disease: A proposed ecological index to assess areas of risk in the northeastern United States. *American Journal of Public Health*, 81, 714–718.
- Singhanetra, R. A. (1993). Malaria and mobility in Thailand. *Social Science and Medicine*, 37, 1147–1154.
- Vachon, M. (1993). Onchocerciasis in Chiapas, Mexico. *Geographical Review*, 83, 141–149.
- Wellmer, H. (1983). Some reflections on the ecology of dengue haemorrhagic fever in Thailand. In N. D. McGlashan & J. R. Blunden (Eds.), *Geographical aspects of health* (pp. 273–284). New York: Academic Press.
- White, G. F., Bradley, D. J., & White, A. (1972). *Drawers of water: Domestic water in East Africa*. Chicago: University of Chicago Press.

VIGNETTE 4.1. Field Mapping for Landscape Epidemiology

Broad patterns of the occurrence of transmissible diseases may be generalized from biome distribution, as depicted in Vignette 6.3. Only the largest-scale topographic maps, however, which show the smallest areas, have the necessary detail of slope, vegetation, and water occurrence/flow to be useful in landscape epidemiology. It is one thing to establish that malaria, onchocerciasis, or California encephalitis is endemic in a place. It is another to specify exactly where people are exposed or where factors in the natural nidus are most susceptible to intervention or avoidance.

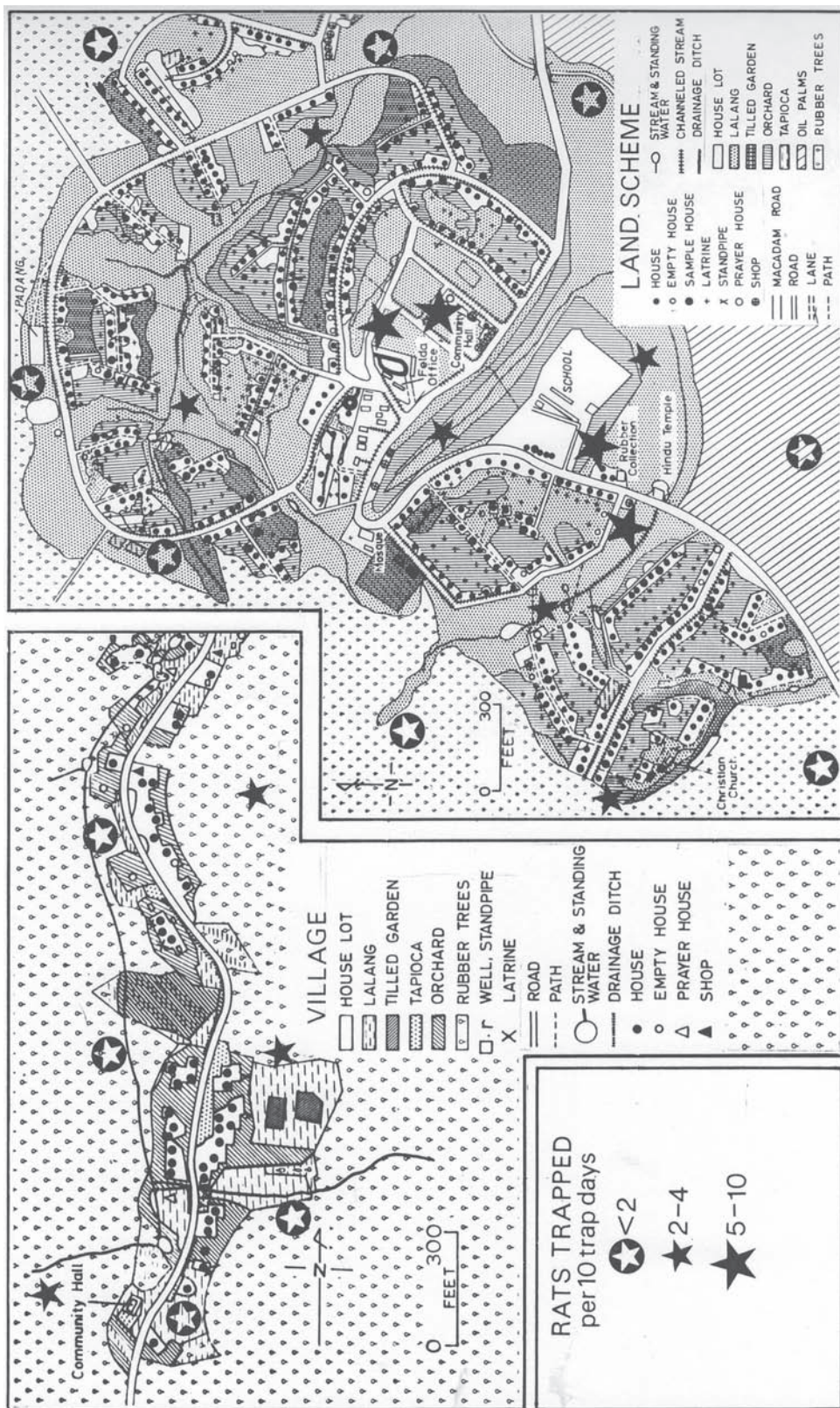
Field mapping is a valuable technique for such analysis. Often a base map can be constructed from air photos or satellite imagery. Elevation (and therefore slope) can be determined, and detailed contour maps can be constructed for the study area. Sometimes a topographic map of sufficiently large scale is already available. In the past a geographer engaged in fieldwork sometimes had to triangulate between flagged sticks with a compass and construct a base map from scratch. This can now be done much more exactly as well as more easily, using a global positioning system (GPS) off a satellite. Analysis of air photography, composite satellite imagery, and field techniques is taught in many geography curricula. Products such as topographic maps or land (tax) plats can now often be scanned into a computer system, although detail and complexity can still be problems requiring researcher generalization and categorization. As is commonly the case with GIS and remote sensing, “ground truth” for accuracy and interpretation still needs to be established through fieldwork.

Vignette 4.1, Figure 1, illustrates the importance of such detailed habitat mapping for understanding the transmission of disease. When the water level rose one year, the foundations of houses became islands of higher land. Rats left their preferred sedge nesting areas and moved into close proximity with humans. Thus a change in river level resulted in greatly increased opportunity for the transmission of flea-, mite-, and tick-borne diseases to the human population.



VIGNETTE 4.1, FIGURE 1. Micro-scale mapping of landscape. From Kucheruk (1965, p. 98). Copyright 1965 by the Moscow Society of Naturalists.

Meade's investigation of health consequences stemming from the resettlement of people in the government's land development program in Malaysia included a study of biotic changes and increase or decrease of transmissible disease risk (Meade, 1976a). A field map for a land settlement scheme and a traditional village was constructed, using a Bronson compass and branches stuck in cans of cement by which to triangulate (Vignette 4.1, Figure 2). In that way, paths, orchards and vegetable gardens, and areas that became marshy after storms could be mapped, as well as houses and particular buildings. Rat traps were set on a spatial sampling frame for fixed periods of time. After the rats were chloroformed, their ears were examined for the species of mite that transmits scrub typhus. Thus, although laboratory studies could not be done to identify antibodies or to culture the rickettsiae, the presence of all other elements necessary for the disease's transmission and their locational risk could be ascertained. The micromobility of the population could then be studied in relation to dangerous habitat.



VIGNETTE 4.1, FIGURE 2. Land use habitat for scrub typhus and the changes resulting from land development.

The land scheme represented a reversal from the traditional village habitat danger. In the village, most of the *lalang* (*Imperator cylindrica*), a coarse, invasive grass that was preferred rat habitat, occurred in open space associated with the rubber trees where the men tapping latex were occupationally exposed to rat mites. Villagers walked on the road in this linear settlement and had little exposure. The land scheme, however, had created a nuclear settlement containing considerable vacant space that was occupied by *lalang*. The houses had quarter-acre lots that often extended into the ravines, where settlers planted papaya trees and tapioca (which the field rats also consumed). Because of the form of the settlement, women and children going to the school, clinic, or store walked on paths through the ravines and fields of grass and were exposed to mites. The area near the school and latex collection station was especially heavily populated by rats and their mites. The men tapping the rubber trees, in contrast, had been disciplined to maintain clean cultivation and control the grass and weeds. Few rats were trapped in the rubber cropland. Much of the vacant land within the settlement had been planned for future development. The *lalang* always invades such vacant, sunny lots. Leaving the undeveloped land under trees, or planting rubber trees there until it was developed, might have reduced the hazard of scrub typhus.

CHAPTER 5

Transitions and Development

The human population of earth—now more than 6.5 billion—has tripled since World War II, in just over half a century. It will increase by half again before natural increase ceases late in the 21st century. After increasing for centuries, the world population growth rate peaked at a little over 2% a year in the late 1960s and then began to slow down. Although the present growth rate of 1.2% (2008) will continue to decrease, the base human population will get larger, adding more than 800 million in each of the next three decades (United Nations Population Division [UNPD], 2001). Its age structure will change during this time, as fewer children are born and a higher proportion becomes older. By itself, the change in age structure will alter causes and patterns of death. But population growth does not only change population composition; it leads to population redistribution. New population settlements and the explosive growth of old ones through migration are occurring already, impacting global ecology. The majority of earth's human population became urban in 2007. This is a new thing under the sun, and surely one of the most profound alterations ever in implications for human health.

The lives of people, their families, their communities, and their societies are being transformed everywhere. Human consumption is growing exponentially and, especially through energy use from fossil fuels, changing the climate and creating pollution (see Chapters 6 and 7). “Globalization” not only of trade and employment, but of food, information, travel, communication, ideas, and more, is making new connections faster than can be realized and tearing old ones asunder. There is no local disease condition, from sick Asian birds to resistant tuberculosis in fetid squatter slums, that can't be on your street next week. And countries in the developing world can have both a starvation problem and an obesity problem at the same time!

The purpose of this chapter is to explain some of the complexity and dynamics of the population–habitat–behavior interactions that constitute disease systems, and

the consequent need for careful evaluation of health consequences in developmental and other changes. The concepts developed in Chapters 2 and 4 on cultural ecology and landscape epidemiology are applied here to show how social and economic change affect disease occurrence. First the population transitions of demography, mobility, and nutrition are explained and the conditions of each stage in these transitions are discussed. The impact of mortality and fertility change upon age structure and disease causation, and the consequent methodological need to control the confounding of such change causes, is discussed. Then the mobility transition is used as the frame to examine the disease ecology of new settlements, new exposures of population to risk, and new health conditions created by rapid urbanization and development of new urban forms. Finally, some of the implications of globalization for health status and emerging diseases in the future are considered.

ECOLOGIES OF POPULATION CHANGE: MULTIPLE TRANSITIONS

Healthy living things and systems of life adapt and change. Human population change over the past two centuries has been transformative for life on earth, the major “driving force,” even as it has itself been transformed. The next half century seems certain to continue the transformation. The changes have been systematically described over recent decades as demographic transitions between preindustrial and postindustrial dynamic states of being (of what would be, in a different context with which readers may be familiar, self-organizing open systems containing their internal fluctuations). What we understand of these changes was originally modeled on the European experience, but there have been enough variations on this theme now to approach the universal. Here is a simple summary: High mortality rates (especially for infants and children) decreased greatly, causing rapid population growth until birth rates also decreased greatly. The fall in mortality was caused by/was associated with a shift in cause of death from infection to noncommunicable, degenerative causes. As first mortality and then fertility fell, the age structure of the population shifted from very young to older; nutritional status and its relationship to health and susceptibility changed; and population movement (or mobility) changed, and with it exposure to risks and opportunities for adaptability.

These population changes are described and explained in this section in terms of four interacting transitions: demographic, mortality (or “epidemiological”—Omran, 1977), nutritional (Popkin, 2003), and mobility (Zelinsky, 1971). The falling fertility is fundamental to changes in growth, mobility, and age structure, but this chapter treats the causes of fertility decline only as context. There is a large body of social science research on why human birth rates and fertility behavior have fallen globally and in each population, and there are many questions about the political economy of those changes that can be asked—but that is the subject of another book (about population geography). Medical geography is about health, disease, and life.

Change in Population on Earth

The world's population of 6.5 billion is today not only growing at a lower rate each year, but since 2000 it has increased by fewer people each year than the year before. About midcentury it is expected to reach a maximum of 8–10 billion. Then, if it continues to follow the pattern set by Europe, population may begin to decrease as fertility rates fall below mortality rates. These projections, of course, are based on assumptions that may not be valid decades into the future. The demographic impact of HIV was not foreseen, and newly emerging and reemerging diseases may drastically change the future as we can foresee it. Nevertheless, several processes have been started and demographic structures have been changed that transform the state of health. The age structure of the population has been transformed; and, as noted above, the majority of the human race has become urban, with profound implications.

The Demographic Transition

Mortality and fertility have both been high, fluctuating around each other, for most of human existence. The progression from this state of affairs to the present conditions of low mortality and fertility in the most economically developed countries is known as the *demographic transition*. It is composed of a transition in mortality and a transition in fertility (reproductive behavior), and a consequently transformed age structure.

The Swedish demographic experience as shown in Figure 5.1 classically illustrates the model. First, mortality declines. For a while fertility continues at its historically high level. During this early stage of the transition, *natural increase* (the difference between births and deaths) accelerates. At the peak of natural increase, there is an increase in emigration. Then, under conditions associated with urbanization and economic demands for a more educated work force, the birth rate begins to fall. In the classic European model, the change to a “small-family norm” happens partly because children, who add to the wealth of a family in a rural agrarian society, drain a family's wealth in an urban manufacturing society. There have been exceptions and differences among regions and countries in the speed of this idea's diffusion. For example, in some parts of rural agrarian 18th-century France, fertility actually declined sooner than in urban industrial England, so this relationship wasn't simple even in Europe. More attention has been paid recently to cultural attitudes about childbearing, to the status and the education (especially literacy) of women, and to diffusion of the very idea of birth control. Cultural values and societal mores about such critical behaviors as age at marriage, divorce and remarriage, women's working for wages, importance of sons, and/or costs of dowries matter a lot. Social-organizational changes from support by extended families to provision of health, family planning, educational, and child care services transform the social environment. The transition is completed when birth rates fall to meet low death rates—and even below death rates, as has happened in European countries and is happening now in several Asian countries. The *total fertility rate* (the average

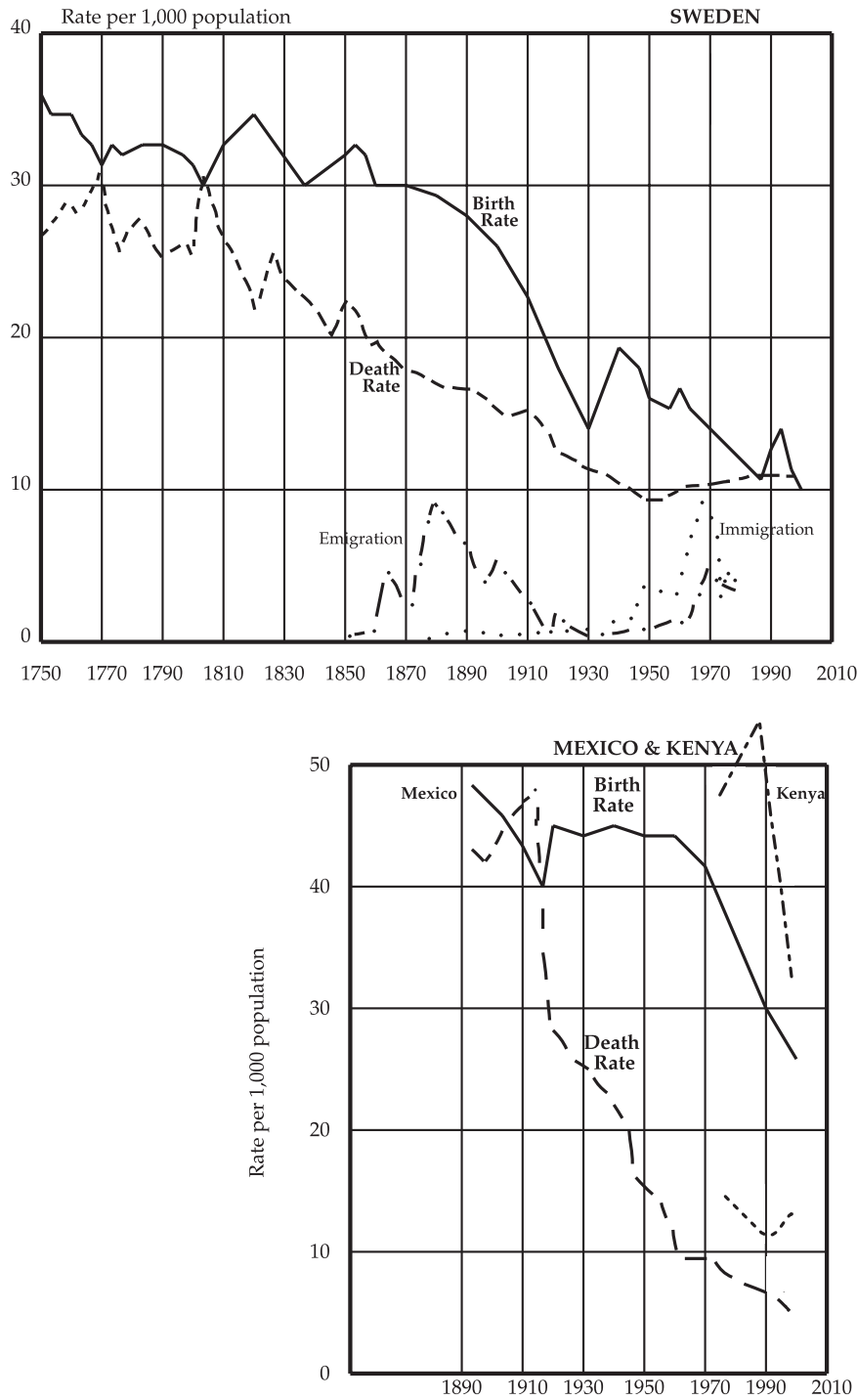


FIGURE 5.1. The demographic transition. Different periods of transition are reflected in the experiences of Sweden, Mexico, and Kenya. Original figure concept and data from Population Reference Bureau.

number of children a woman would have if she went through her life at today's age-specific birth rates) has fallen as low as 1.3 children in Japan, Germany, Spain, and Italy. The Japanese, German, Spanish, and Italian age structure that results from these low birth and death rates now has a higher proportion over 65 than under 15. These changes and those in the following paragraph are illustrated in Table 5.1.

The circumstances of the demographic transition have been different in several important ways for countries in the developing world. The magnitude and rate of population growth increased, sometimes reaching more than 4% a year. This created such youthful populations—as shown in Figure 5.1, Kenya reached in 1988 a total fertility rate of 8.2 children and an astounding 52% of population under age 15—that demands on the economy, social institutions, and local environments threatened to increase death rates again. Indeed, in the case of AIDS in Kenya and other countries in sub-Saharan Africa, they have. In most of Latin America, the decrease in infant mortality seemed to plateau for a decade. Social scientists hypothesized that the impact of modern medical technologies (vaccination, assisted delivery and care of newborns, vector control, nutritional supplementation, and other practices from outside the traditional system) had done all that these technologies could. Issues of internal distribution—of food, housing quality, crowding, literacy, and knowledge—had become critical. Then the fall of mortality and fertility resumed in Latin America. Fertility decline began in sub-Saharan Africa, as Kenya's fertility fell rapidly and population giant Nigeria's total fertility rate fell to 5.9.

The transition has varied in its pace and circumstances. Some countries have been “accelerated,” lowering their fertility before economic development and urbanization. Indonesia, for example—the world's largest Muslim country—has lowered its total fertility rate to 2.6 children, even though its world parity gross national income per person (\$GNI_{ppp}) is only \$3,460 and it is even now only 42% urban, having been 20% urban when it began lowering its fertility. More remarkably, Thailand has lowered its fertility below replacement level (1.7 children), even though it is still an overwhelmingly rural, low-income country and was much more so when it began to lower its fertility.

So each country in its own time and way—with cultural behavior and environmental pressures unique to each, with starts and plateaus and leaps—has been progressing through the relentless demographic transition (as studying Table 5.1 reveals). With the important exception of the impact of lowered fertility on maternal mortality, fertility is further considered in this text only as it affects the age structure of populations.

The Mobility Transition

Population movement in both time and space exists at various scales. The concept of *migration*, as used until recently, implies the intention of a permanent move. It is rarely measured unless it crosses an administrative, political border. Migration may be international in scale, or it may involve moving to another building after marriage. *Circulation* is movement that returns to the original point. This may involve minutes to go buy milk, days on a vacation, weeks for a pilgrimage, or years of absence for

TABLE 5.1. Transitions

Country	2005 pop. (mil.)	% deaths			% YLL		Std. mortality/ 100,000				% pop. age		% urban	TFR	% >5 underweight	% 15+	\$GNI ppp
		IMR	>5		Comm.	Noncomm.	CVD	Cancer	Injuries	>65+							
			Diarrhea >5	Malaria <5						% HIV+							
Sub-Saharan Africa	752																
Nigeria	131	101	15.7	24.1	5.4	83	10	452	157	132	43	3	44	5.9	43	6	930
South Africa	47	51	0.8	0	21.5	77	15	410	154	120	33	4	53	2.8	10	30	10,960
Kenya	34	78	16.5	13.6	6.7	81	11	401	137	95	43	2	36	4.9	17	6	1,050
Malawi	12	78	18.1	14.1	14.2	89	6	430	150	105	46	3	14	6.5	18	2	620
Niger	14	150	19.8	14.3	1.2	87	7	456	169	163	48	2	21	8.0	44	2	830
South Central Asia	1,615																
India	1,104	56	20.3	0.9	0.9	58	29	428	109	117	36	4	28	3.0	44	1	3,100
Iran	70	31	5.5	0.2	0.1	22	49	466	113	113	30	4	67	2.1	9	14	7,550
Bangladesh	144	54	20.0	0.7	—	60	28	428	111	101	35	3	23	3.0	43	—	1,980
Pakistan	162	80	14.0	0.7	0.1	70	21	425	107	99	42	4	34	4.8	31	—	2,160
Afghanistan	30	165	18.9	1.0	—	71	18	706	153	134	45	2	22	6.5	46	—	—
Middle East/North Africa	408																
Egypt	74	28	12.8	0.4	<	32	61	560	84	35	36	5	43	3.2	5	33	4,120
Morocco	31	36	12.1	0.4	0.1	44	44	411	67	48	30	5	57	2.5	10	22	4,100
Jordan	6	22	10.7	0.3	2.0	31	45	384	144	102	37	3	79	3.7	4	26	4,640
Saudi Arabia	25	21	6.2	0.2	—	22	53	405	109	72	37	3	86	4.5	—	44	14,010
Yemen	21	76	16.1	7.5	0.1	61	28	553	108	102	46	4	26	6.2	43	—	820
Southeast Asia/Pacific	590																
Indonesia	222	28	18.3	0.5	0.1	41	44	361	132	87	30	5	42	2.6	20	4	3,460

Thailand	65	18	16.2	0.3	1.5	43	40	199	129	74	23	7	31	1.7	16	7	8,020
Vietnam	83	16	10.4	0.4	0.4	40	44	318	123	72	29	7	31	2.2	27	—	2,700
Fiji	1	16	10.6	0	0.1	27	63	470	86	40	30	4	46	2.7	36	26	5,770
Papua New Guinea	6	54	15.3	0.8	—	64	25	442	118	106	40	2	13	2.7	25	18	2,300
East Asia	1,535																
China	1,304	23	11.8	0.4	0.1	23	56	291	148	79	22	8	37	1.6	6	3	5,330
Japan	128	3	0.4	0	2.0	8	76	166	119	39	14	20	79	1.3	—	3	30,040
Mongolia	3	39	14.5	1.0	<	37	47	488	306	96	31	4	57	2.5	5	13	2,020
Americas	888																
United States	297	7	0.1	0	0.6	9	75	188	134	47	21	12	79	2.0	1	33	39,710
Canada	32	5	0.2	0	0.3	6	80	141	138	34	18	13	79	1.5	—	14	30,660
Mexico	107	22	5.1	0	0.3	27	54	163	88	58	31	5	75	2.4	3	28	9,590
Brazil	184	28	12.0	0.5	0.7	30	50	341	142	81	29	6	81	2.4	4	28	8,020
Chile	16	8	0.5	0	0.3	17	64	165	137	50	24	7	87	2.0	1	25	10,500
Cuba	11	5	1.3	0	0.1	10	73	215	129	54	21	10	76	1.5	4	—	—
Haiti	8	83	16.5	0.7	5.6	84	15	402	112	38	42	3	36	4.7	14	8	1,680
Europe	730																
Russia	143	11	2.5	0	1.1	8	64	688	152	217	16	13	73	1.4	—	—	9,620
Sweden	9	3	0	0	0.1	4	85	176	116	30	18	16	84	1.7	—	10	29,770
Germany	83	4	0.2	0	0.1	5	86	211	141	29	15	18	88	1.3	—	12	27,950
United Kingdom	60	5	0.9	0	0.2	10	82	143	26	10	18	16	89	1.7	—	—	31,460
Italy	59	4	0	0	0.5	5	87	174	134	29	14	19	90	1.3	—	9	27,860

Note. Data from Population Reference Bureau (2008) for population, % HIV-positive, % population age under 15 and age 65+, % urban, total fertility rate (TFR), gross national income purchasing power parity (\$GNI_{PPP}). Data are for 2005 except HIV-positive (2003–2004) < means less than 0.05%; — means no data.

Data from the World Health Organization Statistical Information Service (WHOSIS) database (*www.who.int/tuberculosis/en/index.html*) for infant mortality rate (IMR—per 1,000 born); % of deaths under age 5 years due to diarrhea or malaria; % of years of life lost (YLL) to communicable and to noncommunicable diseases; age-standardized mortality rates per 100,000 for cardiovascular disease (CVD), cancer, and injuries; % children under age 5 underweight and % women age 15 or over obese. Data are for 2005, except diarrhea and malaria for 2000 and other causes of death for 2002.

long-term study or labor. Mobility affects exposure, introduction, and other aspects of disease risk, and has long been studied by population geographers.

Zelinsky (1971) suggested a *mobility transition* that adds the dynamic relationship of development to mobility. In a premodern society, with high birth rates and death rates, there is little residential mobility except for that following marriage (see Table 5.2). Circulation is limited to travel for customary religious observances, agricultural needs, local commerce, and warfare. As the demographic transition begins and death rates fall, a massive movement to three general destinations is generated by population growth: People migrate (1) to new agricultural frontiers, (2) from the countryside to cities, and (3) to foreign countries. Both social and labor-related circulation increase. Later in the transition, fertility falls, movement to cities and frontiers slackens, and the emigration virtually ceases; however, circulation still increases in intensity and in structural complexity. Later still, when fertility and mortality have stabilized at low levels, migration is primarily between cities, although there is significant immigration of unskilled workers. Circulation is vigor-

TABLE 5.2. Stages in Transitions

Stage	Demographic	Mortality	Mobility	Nutritional
Historic stable	Birth rates and death rates fluctuating at >35 per 1,000; population increase imperceptible; >45% population age <15; <5% age 65+	Life expectancy <35 years; infant mortality >180 per 1,000 births; death results mainly from infectious diseases	Individual migration almost unknown except for marriage; circulation local for basic agricultural, commercial, and religious needs	Seasonal malnutrition or staple crop deficiency; periodic famine catastrophic
Early in transition	Death rates fall; birth rates remain high; population grows rapidly by natural increase	Infectious diseases are controlled by sanitation, vaccination, suppression of vectors, and medical treatment; young population	Migration from rural areas to cities, agricultural frontiers, and labor internationally begins and grows; circulation increases socially	Growing shortages and vulnerability; heavy labor
Late in transition	Birth rates fall as death rates stabilize; population growth slows and population ages	Degenerative diseases become major causes of death; life expectancy extends to 65	Migration to cities and agricultural frontier slackens; emigration decreases; circulation increases in structural complexity and intensity	Income growth increases meat and sugar consumption; great inequality in deficiency and abundance
Future stable	Birth rates and death rates fluctuating at <20 per 1,000; population increase imperceptible or population declining; <20% age <15, >15% age 65+	Life expectancy >70 years; infant mortality <10 per 1,000 births; death results mainly from degenerative diseases	Multipurpose and vigorous circulation/migration, mainly between cities and for retirement; immigration of labor prevalent	High consumption of abundant fat and sweet calories; low labor; malnutrition: obesity

ous and accelerating; it includes pleasure-oriented trips as well as moves for social and economic purposes. Zelinsky presciently suggested that in the future the technology of communications and delivery systems may decrease migration and some forms of circulation.

The Nutritional Transition

As conceptualized by Popkin (2003), the *nutritional transition* involves a shift in the behavior of people in lower- and middle-income countries from traditional diets with their associated malnutrition and shortage to convergence on the so-called “Western” diet of overabundance and its associated malnutrition. As income per capita increases, a dietary preference (likely to be population based—i.e., deep in our human heritage) for increased animal protein and sweet foods leads to intake of more animal fat and higher energy density from sugar and alcohol. Sweeteners shift (technologically enabled) from mostly disaccharides like sucrose to monosaccharides like fructose and glucose—in other words, from honey and maple syrup to high-fructose corn syrup. At the same time, the hard physical labors of traditional life are replaced with more sedentary occupations and entertainments.

The following description of four stages in the demographic, mobility, and nutritional transitions interrelates them and their significance for the ecology of health and disease.

Four Stages in the Three Transitions

Before the Change: Pretransition Conditions

For thousands of years—at least since the Agricultural Revolution that domesticated plants and animals laid the basis for population growth in permanent settlements and the creation of cities, which in turn led to the creation of most of our contagious diseases—birth and death have fluctuated around each other at what seem to us today incredibly high levels. This can be called the *age of pandemics and famine*. Overall death rates of more than 40 per 1,000 population reflected infant mortality rates of over 200 infants dying in their first year for every 1,000 babies born alive, and child mortality of 50% before age 6. Life expectancy at birth was less than 35 years. The children died mostly from acute infectious diseases (transmission described in Chapter 2)—especially those that caused repeated and prolonged dehydration from diarrhea, or virulent contagious fevers (such as smallpox, whooping cough, diphtheria, typhoid, or pneumonia). People were made more susceptible to dying from these infections because of at least seasonal malnutrition, and sometimes starvation from crop failure. As discussed in Chapter 2, regional chronic deficiencies resulted from specific crops, soil conditions, and domesticated plants.

Matching this mortality, people had seven or more children (total fertility rate). Women were married at puberty, and complications of pregnancy and childbirth were major causes of death. Population growth was very slow, mostly on the order of 1% a century! The high birth rates, of over 40 per 1,000, combined with the mor-

tality of children, meant that the age structure was very young. Almost half the population was under age 15. People usually died within 5 miles of where they were born, having migrated only at marriage according to mother- or father-localizing custom. There was little circulation except for marketing, and perhaps religious observance.

Fifty years ago, these conditions still prevailed in much of sub-Saharan Africa and occurred widely in Asia and Latin America as well. More than 20 countries still had infant mortality rates over 200 per 1,000 births. Colombia, Egypt, and India were among those countries with infant mortality rates over 150; Italy the rate was 74 and in France still 56 (United Nations, 1950). Most of the world's population was rural, and over 40% of population were children under age 15. Today there are fewer than 25 countries in the world with infant mortality over 100. As shown by the rates in Table 5.1, the region of sub-Saharan Africa is still earliest (i.e., closest to the beginning) in the transitions. The death rates in most countries there did not begin to decline until after independence. Throughout the table, notice the relations to age structure (>40% of population under age 15); the high percentage of deaths due to communicable disease, especially diarrhea; and the low GNI_{PPP}. The variation within as well as between regions can be seen in this table as the relationships of income, urbanization, fertility, age structure, cause of mortality, and finally proportions of underweight small children and of overweight adults, shift. This complex table can be referred to throughout the following stages.

Early-Transition Stage

In the *age of receding pandemics and famines*, deaths from acute infectious diseases decreased in Europe, largely as a result of improved standards of living and public health. Early in the 19th century, England especially had experienced the horrible conditions of the early Industrial Revolution. Cities were death traps that took 5–10 years of life expectancy from those who moved into them. Orphans abounded, with the worst conditions of child labor and poorhouses imaginable. These were the conditions that Thomas Malthus and Karl Marx wrote about.

Over the course of the 19th century and the first half of the 20th, conditions of the cultural ecology of disease changed dramatically. Increased prosperity brought improved housing, increases in agricultural production and in imported food supplies, better clothing, improved education, knowledge of germs, and better government. It brought private philanthropy, the “garden city” movement, public sanitation, chlorination of water, pasteurization of milk, electricity and refrigeration, sterilization and other leaps in medical treatment, and many other changes that improved health conditions. Vaccination for smallpox became routine; in the 20th century other vaccines followed for rabies, typhoid, diphtheria, tetanus, and other deadly diseases. The previous luxury of soap became universal, and the habits of personal hygiene were transformed. Rene Dubos (1965) famously noted that even before sulfa drugs and (after World War II) antibiotics were used, infectious diseases such as leprosy and scarlet fever had all but disappeared and many others,

like tuberculosis, had greatly decreased. With public understanding of germs, new vaccines, antibiotics, X-ray and scanning machinery, and new biomedical treatments, deaths and serious illness from infectious disease almost disappeared (see the entries for Europe in Table 5.1). During this stage of the demographic transition, European populations grew as much as 1% a year. As population pressure in rural lands increased, people migrated to the newly industrial cities and to other countries.

The decline of mortality in most of the developing world proceeded in an accelerated manner. By the late 20th century “outside” or international intervention could help with vaccination programs, treatment for previously fatal childhood diseases, pre- and postnatal care, and safe hospital delivery. Mosquito control reduced and in some places even eliminated malaria and yellow fever. Smallpox became the first, and still the only, infectious disease eradicated from earth. Refugee and emergency food programs prevented starvation. For a while in Latin America, as noted earlier, the decline in infant mortality seemed to plateau. Social scientists concerned with population growth thought mortality might not come down further until there were internal structural changes in crowding, social class distribution of wealth, household distribution of food, and changes in power by class and gender. In some regions mosquitoes and microbes developed resistance to control measures: The effort to eradicate or at least control malaria failed (see Chapter 4), and the disease again increased in parts of South Asia and Central America from where it had been cleared. Mortality decline resumed, however, and life expectancy at birth in almost all countries outside Africa is now in the high 60s or 70s.

During this maximum-rate-of-growth stage (when mortality has fallen but births are still high), the mobility transition posits three kinds of migration destinations to relieve the growing rural population pressure, as noted earlier. First, people push to new agricultural frontiers as farmers who want to farm convert forest and marshland to new cropland. Second, they start migrating to cities in search of jobs and a means of support. Even as the world adds its last 2–3 billion people to developing countries, these and currently rural people will be added to cities. Today’s 50% urban is composed of more-developed countries at 75% and less-developed countries at 44% (Population Reference Bureau, 2008). In 20 years the world will approach 60% urban, and the less-developed countries are projected to be 56% urban (UNPD, 2002). Some health consequences of this are considered below and in future chapters on development and diffusion. Third, people living to grow up move internationally in search of jobs—a process that is already transforming nations.

Late-Transition Stage

The late stage of the demographic transition is entered when birth rates begin to fall. Once this starts, fertility continues to decline until the now low birth rates are met, and the transition is completed. During this stage the population stops growing by natural increase and, without migration, it can even decrease. This has happened to counties, provinces even, when accompanied by out-migration to cities and other

countries, as described in regard to the mobility transition. Fertility rates may have fallen even to replacement levels, and the base of the age-structure pyramid narrowed, but those born 20 years ago are now looking for jobs. In some countries, like the Philippines, the remittances sent home from those working internationally can even become the major source of foreign exchange. Migration out of rural areas and into cities compounds the lowered fertility to change age structure drastically. The broad base of youth disappears from the age structure pyramids. For a couple of decades, a country can enter the “demographic window” of development when the proportion of child dependents (under age 15 or 16) is greatly reduced and they are healthy; the proportion of elderly dependents (over age 65) is still tiny, and most of the population is composed of healthy, working-age adults. This variation in age structure that follows the transition can be seen in the country pyramids of Figure 5.2.

As the population urbanizes and ages, it also increases its income and consumption. As discussed in later chapters, people become exposed to the air and water pollution so far associated with heavy industrialization. Populations also seem to increase their cigarette smoking. Their diets increase in animal protein, fresh fruit, sweetened breakfast cereals, soft drinks, alcohol, and fast foods. The main causes of disease become noncommunicable and degenerative, such as cardiovascular disease and cancer (Table 5.1).

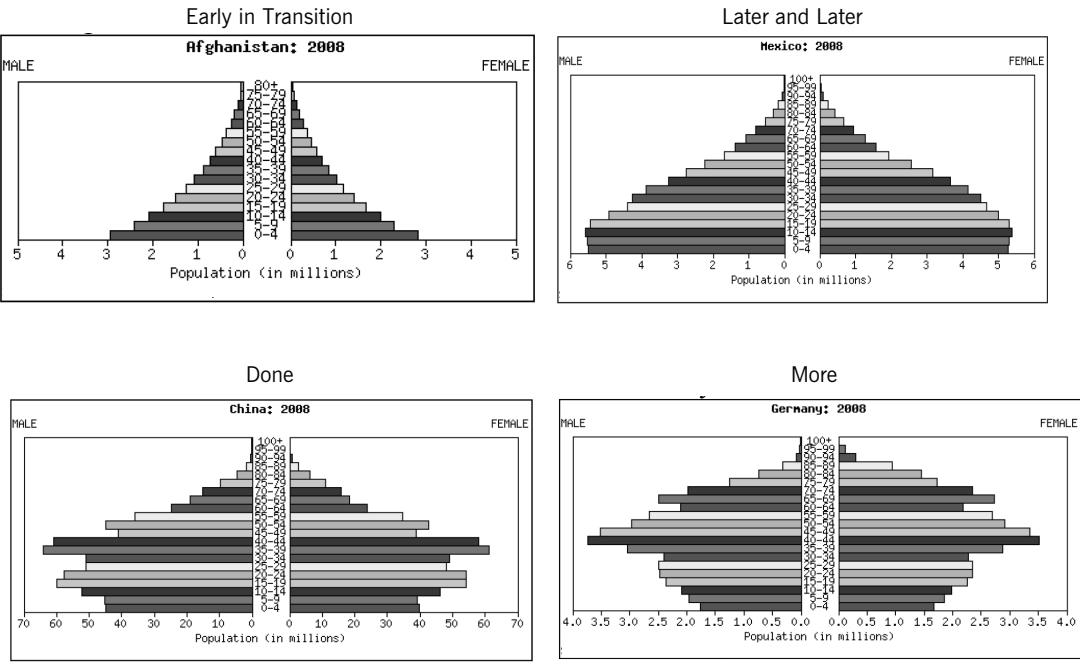


FIGURE 5.2. Population pyramids of age structure. Data from U.S. Census Bureau (2008).

After the Transitions Are Completed

There have been two major surprises in countries past “the end” of the demographic and other transitions. First, fertility so low that population decreases by attrition can characterize regions and countries (Western Europe, Russia, and soon Japan) that have completed the demographic transition. This was the big demographic surprise of the last decade. The United States has escaped this dynamic (see Figure 5.2) largely because of the immigration of so many millions, most of whom are working adults having on average 2.8 children instead of 1.8. The rural South was of course last through the transition and still has some of the highest infant mortality rates. There has also been cultural resistance to lowering fertility below two children: American fertility levels are highest among the Mormons of Utah, for example, and in the rural mountain West (and Alaska) generally.

The second biggest demographic—and major health—surprise was how fast a major, industrialized, late-demographic-transition country could deteriorate in health. The fragility of the interactions that supported good health and their vulnerability to deterioration in economic and social organization was shocking. When the Soviet Union collapsed in the early 1990s, the turmoil in economy and society was quickly reflected in soaring mortality. Within a year, Russia had epidemics of diphtheria and whooping cough even in Moscow, which affected mostly small children at startlingly high rates. Death from cardiovascular disease has soared (Table 5.1). It is ascribed largely to the impact of alcohol, as is a huge increase in infant deaths from fetal alcohol syndrome. The increase in infectious disease was variously attributed to poor quality control for vaccines and failure of delivery systems; deterioration of nutrition, housing, and heat; failure to pay medical personnel; and failure to produce and supply medicine or blood products. Suicide became epidemic, especially among those with their pensions and their life’s work destroyed. In more recent years, HIV infection and drug addiction have become social plagues.

At completion of the demographic transition, rural-to-urban migration is complete, and migration occurs largely between cities. Circulation is constant and intense. The causes of death overwhelmingly become degenerative diseases: heart disease, stroke, cancer, kidney failure, diabetes. Motor vehicle accidents also become a major cause of death, as well as adding a great burden due to *years of life lost* (YLL).*

In the last few years, fertility has increased a couple of tenths of a percent in a few northern European countries, perhaps responding to programs of parental leave and child care intended to lessen the burden on parents and interference with professional lives of having children. All low-fertility countries, with at least the legal exception of Japan, have become targets of international labor migrants. Many of these migrants intend to return home, and so are technically circulating—meaning that the movement of infections or ideas between source and destination

*The YLL measures the burden of disease by counting not only the number of deaths, but the years (and earning power) lost from those lives. Thus the YLL of AIDS or motor vehicle accidents is more than that of stroke or colon cancer.

goes both ways. The main demographic and health development, however, is one of disease ecology. Old diseases are becoming resistant to medicines and insecticides and are even adapting to new behaviors. Led by HIV, new diseases are emerging. The next stage of the demographic transition may yet become the *age of returning pandemics*.

MAJOR IMPACTS OF POPULATION CHANGE

Perhaps the most profound health impact of the progression through the demographic transition is the change in the age structure, or composition, of the population. There are living people who in their own lifetimes have gone from a community population that was 45% under 15 years old to one that is 15% children, and now is also 15% over 65. This transformation is a change for the lives and health of the individuals, especially women, as well as for the structures, norms, supports, and dependencies of society. When we explain the demographic transition, we like to use the “cousin rule” to illustrate the change in experience. In a society in which everyone had six children, including your parents and their siblings, you would have 60 cousins. Many would not live to grow up, but they would be born. In a society in which the majority of parents have two children, you have 4 cousins—and, because of mobility, they live in other states! The social relations that help people survive sickness, raise children, support the elderly, or recover from flood are fundamentally altered.

Who knew that such change was humanly possible? A woman who used to get married by 18 would probably, if fortunate enough to remain healthy and alive, have been pregnant or nursing for most of 20 years and sporadically for another 10. She might have had more than 10 children born alive (although half would have died before age 5), but there would have been several other stillbirths and miscarriages. She would have been suffering from malnutrition and severe iron deficiency even in good times, and the burden on her body would have made her susceptible to infectious diseases. Anemic from intestinal worms, schistosomes, malaria, and other infections, she would have faced obstructed births, hemorrhage, and sepsis from pregnancy and childbirth (and perhaps desperate, unsafe abortion to save her own life). Even today, a woman in Niger, early in the demographic transition, has a one-in-seven lifetime risk of dying from reproduction (World Health Organization [WHO], 2005). When fertility rates fall from a total fertility rate of seven children to five to three, many more women live to grow older.

Age Structure and Its Implications

As people live to grow old and reproduction falls below replacement, the age structure shifts from predominantly youthful to predominantly working-age, and then the population ages. The causes of death as a population goes through the mortality part of the demographic transition shift from acute infectious diseases that kill children, to low death rates of healthy adults, to degenerative and other noncom-

municable diseases of older age. Just looking at the pyramids in Figure 5.2 can tell you where deaths from diarrhea, from stomach cancer, from violence, or from heart attacks are occurring. To know the total fertility rate of a country is to have a good idea of its major causes of death. Notice also in the pyramids of those latest in the transition the growing problem of who will provide medical care and support in regions and countries with large proportions of elderly people. In the U.S. states of North and South Dakota, for example, some registered nurses have daily drives of more than 200 miles to visit their patients. Chinese civilization has held an ideal of “four generations under one roof.” Under the mortality rates that used to prevail, it was usually a dream. Today’s “one child” however, will see a society that is almost 30% over 65. If that child is a girl, she will provide care for her child, her parents, her husband’s parents, and surviving grandparents.

Look again at the changes in age structure that are happening as the world goes through the demographic transition. Public health needs to grapple with these changes of structure and the changes in health conditions and need for services that they portend. They are often approached in epidemiology and sociology as *stages of life*.

Stages of Life

Infancy. The greatest risk of dying before age 60 occurs in the first year of life, and within that in the first month after birth (neonatal) or the first month both before and after birth (perinatal). The chief components of this risk in developed countries are respiratory conditions and disorders related to short gestation, low birthweight, and congenital abnormalities. Sudden infant death syndrome, accidents, and pneumonia complete the top five causes of death. Infections pose the greatest risk early in the demographic transition, as discussed earlier.

Childhood. When fatal infectious diseases have been largely eliminated by public health systems, vaccination, and antibiotics, the major cause of death in childhood is accidents, and even these occur rarely. Leukemia does occur, and genetically caused diseases and congenital abnormalities still take a toll, especially in heart disease. In the United States, murder and suicide have become more common than death from pneumonia and meningitis.

Young Adulthood. Violence in the United States—accidents, homicide, and suicide—accounts for more than three-quarters of deaths among those ages 15–24. The overall death rate of barely 1 in 1,000 conceals great gender differences in violent death (e.g., white male accidents are 2.7 times white female accidents; black male homicide is 8.4 times black female homicide).

Maturity. During the years between ages 25 and 64, death rates increase nine-fold; degenerative diseases rise to dominance; and gender, racial, and ethnic differences increase. Although chronic obstructive pulmonary disease (COPD: mostly emphysema) is relatively more important for whites, liver disease for Hispanics,

and AIDS for blacks, the top 10 causes remain the same for all groups—except that suicide continues to be more important for whites and murder for blacks and Hispanics.

Seniority. As mortality increases steeply between ages 65 and 74, heart disease, cancer, and stroke come to account for more than 75% of all deaths. Kidney disease, liver cirrhosis, and diabetes as well as emphysema become major causes of death, and functional disability increases. Although three-quarters of seniors have no trouble with such personal care activities as walking, dressing, or bathing, the need for home help becomes more common. Blacks continue to have death rates half again as high as those of whites, and male deaths continue to exceed female.

Old Age. In old age, the causes of death in seniority escalate, and pneumonia becomes a common threat. Morbidity increases greatly. More than half the population suffers from arthritis, and a third from serious impairment of vision and/or of hearing. Mental illness also increases steeply. In this group there are high rates of institutionalization, lower proportions of minorities, and a uniquely low sex ratio. The population over 85 is the most rapidly growing group in the population structure of developed countries.

Confounding and the Need for Age Standardization

As conditions of poverty and education, behavioral norms, and environmental exposures change in the future, these stage-of-life associations will be altered in both developed and developing countries. The ability to “decompose” risks of mortality from specific disease by sex, ethnicity, region, education, occupation, and other groupings has made it clear that there are strong *cohort effects* (i.e., effects due to the shared experiences of being born at the same time and progressing through life’s experiences together) for mortality and morbidity. One has only to look at the age structure of China, the tobacco being smoked by young adults, and the air pollution being breathed by children to foresee the coming epidemics of lung cancer and heart disease in that country. Through both biology and experience, age structure pervades the experience of health and disease.

A basic four-part age division for four states is presented in Table 5.3. These states have different incidences of heart attacks, accidents, ear infections, pregnancy, mental illness, AIDs, and every other condition and health service need. Consider how the need for heart specialists or obstetricians must differ between Utah and Florida, or how you might try to determine whether North Carolina’s historically high exposure to tobacco has affected lung and heart illnesses, when the age structure indicating cumulative exposure is also so different. How then can mortality be compared between countries when they have such different age structures, as shown in the population pyramids of Figure 5.2? Are the millennial goals of the United Nations being more successfully implemented in one country or region than in another? How can “before” and “after” conditions be compared when age structure has changed too?

TABLE 5.3. Age Structure of Several States and the Nation as Standard, 2006

State	Florida	Utah	California	North Carolina	United States
Population (millions)	18.0	2.5	36.5	8.9	299.4
% 65+	16.8	8.8	10.6	12.2	12.4
% 18–64	61.0	60.2	63.3	63.5	63.0
% 5–17	16.0	21.3	18.8	17.4	17.8
% <5	6.2	9.7	7.3	6.9	6.8

Note. Data from U.S. Census Bureau (2006 estimates, FactFinder).

The etiology or treatment success of almost anything cannot be studied or addressed through any comparison of places without controlling for those differences in age structure. The ubiquitous presence of age as a confounding influence (age is related to nearly everything, from immune status to cumulative carcinogenic burden to behaviors involving exposure, activity, and knowledge) is routinely handled by *standardization*. In Table 5.3, the population age structure of the United States—the total of the parts, in this case—stands as a basis for comparison. Florida’s population of 18 million is 16.8% age 65 or older; this means that 3.0 million people are in this age group. The older populations of Utah are 220 thousand, California 3.8 million, and North Carolina 1.1 million. Of course there are going to be more deaths from any cause in that age group in California, since it is not only our most populous state but has the largest number of older people. Dividing the number of deaths from heart disease that occurs in Utah, California, or North Carolina, by the population of each state, respectively, would give crude death rates that are intrinsically not comparable. The “age-specific” death rates (in this example not year by year, but for the category of those age 65 or older) of the four states can be applied to the total U.S. population by age, yielding a hypothetical number of deaths that would occur if the United States had each death rate in turn. When these hypothetical deaths are divided by the standard U.S. population, then, a directly standardized rate is created that can be used to compare the mortality experience of the states because they are all being compared on the deaths that would result *if* each state had the age structure of the standard U.S. population. This process of age standardization is further explained and developed in Vignette 5.1.

NUTRITION IN TRANSITION

Technology and development affect nutrition in many ways. Roads, bridges, electricity, and refrigeration can have tremendously beneficial results. Food can be delivered to areas where crops fail. Varieties of food can be supplied year-round rather than seasonally, ending the seasonal malnutritions that occurred traditionally. Intestinal diseases can be curbed so that children can absorb their food. Modern processing and storing can deny rodents and vermin their free meals and save more food for people. (As discussed in Chapter 2, however, such food processing

can also spread beriberi.) Information on weather, farming practices, or new varieties of food crops can be disseminated quickly. Just a few years of education, or basic literacy, can increase nutritional knowledge and improve weaning foods. New foods and fish cultivation can be introduced; poultry and livestock diseases can be controlled to permit the production of more food.

Many people assume that most of the malnutrition in the world today is a consequence of poverty, and that economic development that increases people's income will improve their nutritional status. Certainly for those suffering from absolute deficiency of calories and protein, the economic ability to "demand" food can mean the difference between life and death. For malnutrition, however, the evidence is much more mixed.

Adverse effects of development are deeply intertwined with societal changes. Women who are drawn into economic activities outside the house have fewer children. Usually they also stop breastfeeding and start bottle feeding. Other women do so simply because bottle feeding becomes associated with higher status. Being too expensive for women earning less than U.S. \$1 a day, powdered milk is often overdiluted with unsafe water in bottles that cannot be boiled because of fuelwood shortage (the cost of fuelwood often exceeds half their earnings). The consequences of these practices can be epidemics of diarrhea and high infant mortality. As population growth results in land fragmentation, farmers with inadequate land or who become landless laborers not only cannot demand food; they may, with no means of income, lose their entitlement to it.

Development also plays an important role in food preferences. People growing cash crops or moving to local towns choose higher-status foods such as white bread or rice, and no longer have access to the more nutritious but coarser homegrown fare. Plantations and agribusiness have often been criticized for demanding that all of their laborers' attention be devoted to the cash crop while providing no land for vegetables, fruit, or staple foods (the same situation that used to exist with cotton sharecropped in the U.S. South). As one travels in developing areas, however, one sees many smallholders (farmers) who devote all their land to a remunerative crop. In the Peruvian altiplano, the land where potatoes were domesticated, several varieties are neglected today as people leave for seasonal labor opportunities and cannot cultivate those varieties needing labor at that time; as their vales remain untilled, those potato varieties (and unique forms of nutrition) move closer to extinction. On the other hand, globalized marketing has filled shelves with SPAM and opened supermarkets throughout Latin America and Asia. Coca-Cola, McDonald's, Kentucky Fried Chicken, and the like represent the United States almost everywhere, and locals with new money crowd them.

Food preferences play a crucial role in nutrition. Except for those people who are vegetarians for religious reasons, however, the evidence is that people whose diets have been based on vegetable protein and grain are eager to increase their meat consumption at the first economic opportunity. Japan used to be the most obvious example, but the preference for meat when accessible is universal today. China displaced the United States as the world's leading meat consumer in 2003,

and may now consume twice as much despite water problems and grain production/importation (Brown, 2004). Other changes in preference are more problematic. Consumption of sweet calories increases, from soft drinks especially, and dental cavities consequently increase in places where dental care is available only to the rich. Increasingly the sweetness is not from sugar (sucrose), but from fructose (corn syrup in the United States and its globalized products), which is metabolized differently and causes higher glycemic shock. Part of modernization in the developing world seems to be adoption of Western-style diets high in saturated fats, sugar, and refined foods; low in fiber; and accompanied by low levels of activity in work (less physical labor) and leisure (from transportation to entertainment). The excess calories and fat and the consequent obesity and serum cholesterol are considered a new kind of malnutrition. Popkin (2003, p. 581) says that as incomes rise, changes in diet and activity patterns lead to emergence of new disease problems and increased disability; in the last stage, however, behavior "begins to reverse the negative tendencies" and makes possible a "successful ageing."

The biggest emerging problems by far are diabetes and obesity. Both today are prevalent in both developing and developed countries; they affect children and adolescents as well as adults. Among the more affluent, late in the transition and after their completion, diets, occupations, leisure activities, and means of transportation "modernize" to parallel Western lifestyles. Diets high in fats and sugars, and lives low in activity, result in obesity. Popkin (2002a, 2002b, 2003; Popkin & Gordon-Larson, 2004) has found that this shift in the burden of the malnutrition of obesity to the poor occurs when countries reach a gross national product per capita of about U.S. \$1,700. Obesity in many countries, from Egypt to the United States, becomes more prevalent among the urban poor. In developing countries, too, the urban poor have less access to fresh fruit and vegetables, eat more fat and sugar, live in more crowded and unhealthy conditions, and are stressed by their marginal situation. Not only is it possible (as Figure 5.3 shows), but it is becoming increasingly common, for developing countries to have both a high percentage of children extremely underweight and malnourished, and at the same time a high percentage of children very overweight and malnourished. Obesity is a major risk factor for serious noncommunicable diseases, such as cardiovascular disease, hypertension, stroke, diabetes mellitus, and various forms of cancer. Diabetes mellitus itself is one of the top 10 causes of mortality in the United States and is a rapidly increasing problem.

Obesity was, as of 2007, prevalent among all groups of Americans. One in four (26%) American adults were obese, which was defined as having a body mass index (ratio of weight to height) of 30 or more. As recently as 1995, that was only 15%. Not one U.S. state met the new 15% target established by the Centers for Disease Control and Prevention's (CDC's) Healthy People 2010 initiative (Colorado was closest, at <20%). Three states had obesity rates exceeding 30% of the population (CDC, 2008). Even as U.S. public health agencies struggle to change the conditions promoting this epidemic, as described later in this text, it is spreading throughout the world.

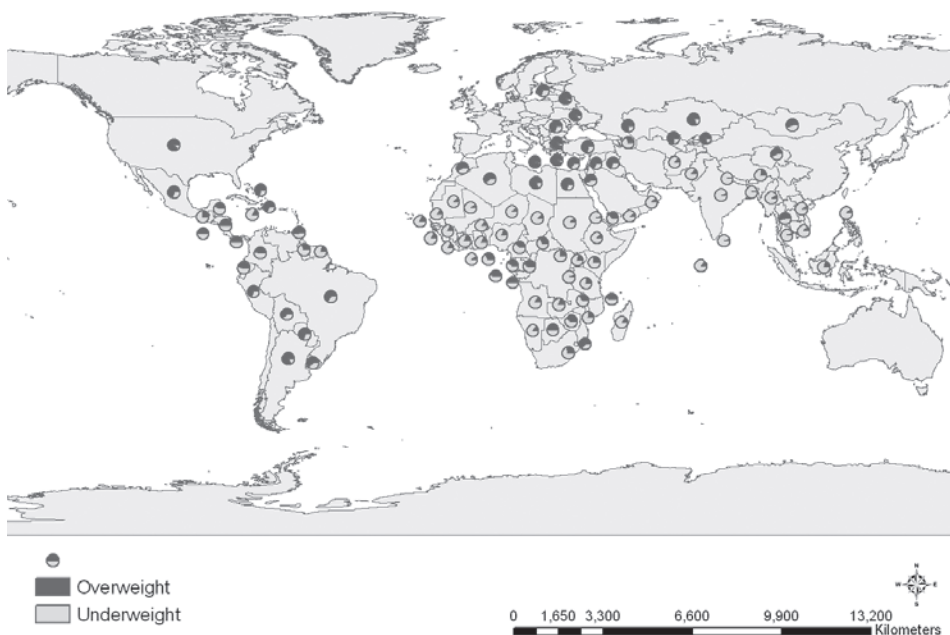


FIGURE 5.3. Proportion of underweight and overweight children under age 5.

THE MOBILITY TRANSITION AND TIME-SPACE GEOGRAPHY

During the middle of the demographic transition, when rural populations are growing rapidly by natural increase, the fragmented farmland and the traditional social structures and political systems face not only population pressure, depletion, and erosion, but today also new global markets and the new financial structures and entourages of nongovernmental organizations (NGOs). The alterations in ecology that result from mobility to Zelinsky's three destinations are considered at length soon. First the scales, importance, and measurement of population movement in time and space need to be considered

Analyzing Mobility in Space and Time

As noted earlier in this chapter, *circulation* is population movement that returns to its place of origin. Such movements occur in a continuum of time and space. People may move from room to room in a house in minutes; they may go out to their garden, walk the dog, visit a neighbor, or go to a playground in hours. They also journey to work, go shopping for food, and pursue other activities (e.g., trips to church, sports, schools, dentists, and friends and relatives). Young adults may go away for college or job training, may work locally, or may move to another state for employment. College students today often study abroad, and people of all kinds travel internationally for tourism, business, research, and other work. (In the first decade of the 21st century, for example, hundreds of thousands of Americans circulated

through Iraq or Afghanistan.) Circulation for college takes years, and commonly students do not in fact return to their home towns, and sometimes not even to their home states; in other words, long-term circulation can become migration (permanent). Leaving one's country for wage labor has become a common practice in the developing world. In the community of "illegal immigrants" in the United States, it has become impossible to estimate how many people leave Mexico, for example, with the intention of living permanently in the United States, and how many would circulate home to family and friends if the border were easy to cross. The old definition of *migration*—crossing political boundaries (or there would be no knowing that it had occurred) and moving with the intention of permanence—is becoming not only difficult to identify but, when circulation extends over years, irrelevant.

Traditionally, early in the mobility transition, agricultural people seldom migrated except for marriage. Sometimes "forced migration" occurred due to natural disasters, or war actions, government labor conscription for building roads or walls, or the like. A lucky few Muslims might earn the esteemed title *hadji* for making the long, arduous pilgrimage to Mecca and back home—a journey that might take over a year. People moved around their home compounds, their village establishments and markets; they circulated in larger areas of regional markets and of religious and business needs.

Such patterns still exist in a few parts of the Third World, but are rapidly being supplanted. As roads and bridges get built, people's activity areas increase. As they extend the area for work in plantations, cities, new agricultural lands, or international education or work, they also intensify the circulation with visits home and visits by village friends and neighbors to them. Villages become connected to the agricultural frontier, as well as to the big city (and thus to the world urban system). Each of these movements in time and space brings exposures to different insults and raises the potential of introducing old disease agents to new places or being exposed to new ones. As illustrated in Figure 5.4, people may move out of an area where a disease is endemic and introduce it to their destination area; if the agent is very contagious or the transit time between origin and destination is long (as in a refugee camp), the transit route may disseminate the disease. (This was the common case described in Chapter 4 when malaria moved down rivers and canals with migrants, going to their destinations, but also infecting riverine mosquitoes to infect future travelers and local residents.) Alternatively, people may leave an area in which the disease agent does not occur, only to be susceptible to it when they reach their destination. Or, finally, people may leave a safe area and move to one, but be susceptible to the hazards—cholera, rape and infection, malnutrition—in camps and stops along the way. The various points in time and space of exposure of susceptible individuals to infections identify different locations and different prevention or containment strategies for public health interventions.

Prothero (1961, 1965) was the first geographer to realize the importance of population mobility for disease diffusion and control. He conceptualized a typology of time-space mobility that proved useful in the control of malaria in East Africa. He mapped animal-herding territories and routes, pilgrimage routes, and other forms of culture-specific population movement that were vital to this effort. Prothero also

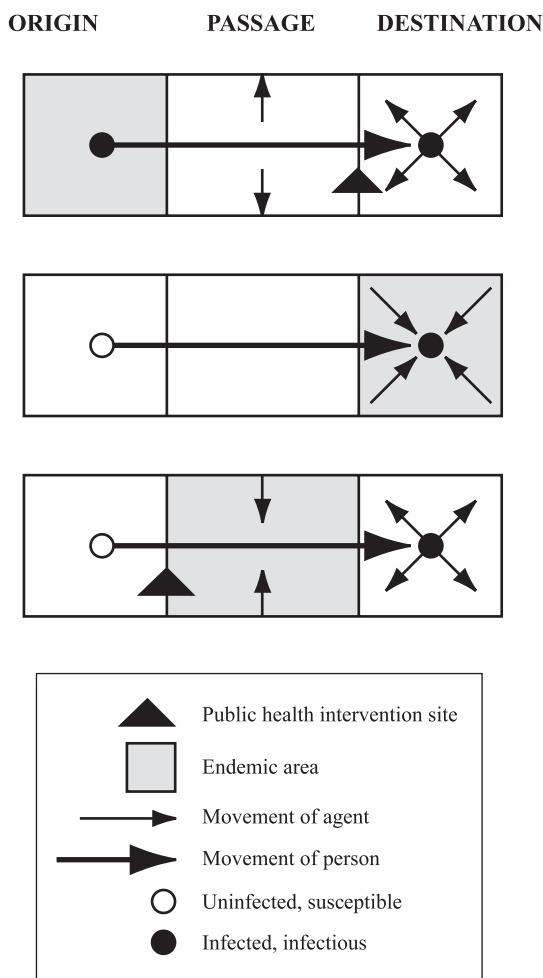


FIGURE 5.4. Mobility and disease agent transmission.

mapped for the WHO *labor-sheds* within which people moved, even crossing international borders, to work in mines or plantations—areas within which migrants would be continually circulating and introducing or being exposed to the malaria plasmodia. Efforts to control and eradicate malaria internationally, he stressed, had to be coordinated to concentrate on the labor-shed, or one country's eradication efforts might well be thwarted by another's. Gould and Prothero (1975) developed a typology of mobility in which they classified on one axis increasing time (from hours to days, weeks, months, years, and so on into permanence) and on the other scale of movement (from local grazing to international travel). Into this time-space classification, they put purposes (herding, visiting, marketing, schooling, working, etc.). By their nature, circulation movements have reverse flows about equal to their forward flows and so have greater potential for disease circulation than does unidirectional migration. The relevance of this old typology of mobility for under-

standing the movement of malaria has recently been rediscovered (Martens & Hall, 2000).

The effects of these economically driven changes in mobility on patterns of disease diffusion are important subjects for future study, as we will see in the discussion of emerging diseases. Understanding the types of mobility and anticipating their transformations provide a basis for conceptualizing disease diffusion systems at any scale.

Measuring Mobility and Exposure

One can visualize different kinds of hazards varying in intensity from absent to intolerable as they are spread out on the surface of the earth. Each can be mapped. Each mapped “surface” can be overlaid on the others. At various scales, from backyard continuously up to national and continental, such overlaid distributions of hazards could be used to define hazard regions. Mapping how people move around and through such a multiple-hazard surface and how long they spend at any point or in any hazardous area/region (i.e., the duration of that exposure) can be used, conceptually at least, to model risk and predict future disease outcomes. For example, people may breathe different air pollution at home, highway, and work; may have benzene exposure at the gas station; and/or may be exposed to chemicals in the office, chemicals on the golf course, or gasoline chemicals in drinking water at home. The individual movements will need to be aggregated in all their variation, and the hazards measured across multiple scales of occurrence, to attempt to approach any population predictions. Population and economic geographers have developed several ways to summarize and analyze such population movement. These are usually classified as *time-space geography*. Future medical geographers need to conceptualize and develop this critical area better.

Time-space geographical analyses for epidemiology have been most fully developed and explained by Schaerstrom (1996). They build on the ideas of Torsten Hagerstrand, who first conceptualized such movements and developed “lifelines” in space to analyze them (Hagerstrand, 1957, 1975, 1982). Working with Swedish data, he located the moves made by each person in a town over his or her lifetime among the possible residence sites of that town. A vertical line represented the time spent in that site; a horizontal line indicated distance moved to another site. In these lifelines, people left home to work, got married, moved to new houses, grew old, left houses; these residences were then occupied by the lifelines of others replacing them in those locations. What a wonderful way to conceptualize exposure, to chemicals perhaps, over a lifetime. It has seemed impossible to do this for more than a few people, but now new modules of analysis on geographic information systems (GIS) have opened up the possibilities. (See the discussion of Bian’s work in Chapter 10.)

Population movement in space, mobility, is thus a continuum from the arrangement of furniture in a room to massive flows of people between continents. Perhaps the best way to understand the nature and measure of exposure is to begin at the micro scale. Within a district, valley, village, yard, house, or room, areal differences

in health hazards can be identified. The child standing in the kitchen next to a pot of boiling water is exposed to a different hazard than the one across the room tending the cat litter is; the man driving his car on a busy street in smog is exposed to different hazards than the farmer driving his tractor under direct midday sunshine through the spore-laden dust is. Among the key evidence of gasoline as a source of lead hazard, leading to its removal, were the concentrations of lead in dust on apartment windowsills adjacent to roads and intersections. They were many times higher than those of dust on windowsills farthest away, clearly showing distance decay from a source. People in buildings adjacent to landfills, incinerators, or simply industrial smokestacks breathe different air than those living in the same city under suburban trees do—but these trees are the ones that have birds infected with encephalitis! In a lifeline approach, the vertical line representing time in place is one of continuous exposure (e.g., to allergens, air pollution, or ultraviolet radiation); the horizontal line across space represents contact with, and perhaps transfer of, infectious agents. Having a goal of understanding who is exposed when and how long to whatever risk is being considered, one still needs to be able to measure the exposure. Vignette 5.2 discusses one method of collecting and analyzing such mobility data to examine individual exposures.

Larger areas of some repeated or maintained exposures may be identified by constructing elliptical cells. The axes of these cells are simply measures of the average population movements and individual deviations, such that a cell contains two standard deviations of all trips. They can be calculated to summarize areas within which people circulate for work, school, shopping, and other social activities. These are the areas in which people are actually exposed, rather than just the point locations of their residences. Each area can then be considered for the hazards it contains, or, more commonly, for the medical services and facilities accessible to people moving within the area. The construction and other uses of such elliptical cells are further described in Chapters 11 and 12. Since population mobility is very much involved in diffusion of disease, several ways of modeling and finding patterns in such movement are developed in Chapter 10 on disease diffusion.

At the opposite end of the time–space scale continuum, massive movements of population between countries, states, or provinces can occasionally be analyzed over lifetimes to separate etiological factors of environment, behavior, and genetics. The argument is that a sudden change in disease incidence upon migration would be due to some environmental factor; failure to change at all would suggest genetic disposition; and change over decades or generations would suggest behavioral roles and exposures. Migration to Hawaii from Japan set the classic epidemiological stage: Stomach cancer was high in Japan but low in the United States, and breast cancer was the reverse. Stomach cancer fell slightly in the migrants, fell more in their children, and by the next generation had approached American rates. The low Japanese breast cancer rates at first did not seem to change at all, suggesting less genetic susceptibility. Third-generation Americans of Japanese descent in Hawaii, however, did not have the low rates of their genetic homeland, but intermediary rates between Japanese women and European American women. In an updated study of women in Hawaii and California (596 cases and 966 controls), there was

a sixfold gradient in breast cancer risk by migration patterns (Ziegler et al., 1993). Asian American women born in the United States had a 60% higher risk than those born in Asia; among those born in the United States, the risk was determined by the number of grandparents born in the United States. Of course, habitat in Asia was changing, too. Among Asian Americans born in Asia, the risk was 30% higher if the prior community was urban, and 80% higher if they had lived in the United States over 10 years. Such evidence strongly suggests the importance of American lifestyles.

Migration streams have also been used to analyze the remarkable latitudinal pattern of risk for multiple sclerosis and a few other conditions, and possible etiological factors for esophageal cancer, as we shall see in Chapter 8. The existence of large (because such diseases are relatively uncommon, large numbers of people are needed to study them), coherent migrant flows with known origins as well as destinations is a rare phenomenon. Mobility data available in the United States are from few sources, designed to protect privacy, and offered at improper scales for such analyses. The decennial census asks respondents where they lived (same house, county, state?) 5 years previously; changing residence to another state requires paying taxes to license a car in that state, and lists of totals can be accessed; the much-researched 1:100,000 sample of Social Security address changes offers nothing that can be connected.

DISEASE ECOLOGIES OF THE AGRICULTURAL FRONTIER

As population grows rapidly in the middle of the demographic transition, and as farmland becomes fragmented by inheritance and/or acquired by larger landowners, farmers in the developing world have often set out for new farmland. Not every country has rain forest or other land that can be converted to intensive agriculture, of course, but surprisingly many do. Farmers in India and China have cleared land higher up the mountains, into more arid land, into colder land, and into marshy land. Pandas, tigers, and numerous species barely known are thus endangered, of course, but people too have come to live densely in land at the margins—land now flooded in extreme monsoons and cyclones, land newly infected with resurgent malaria, land under sliding mountains, land blowing away in the wind. Hundreds of thousands have recently been thus forced from their homes and their new land in India. China is struggling to rebuild and again resettle people displaced by earthquake, landslides, and flood from such lands. In Bangladesh, devastation seems to have become an annual event (and the seas have yet to rise). The Vietnamese have converted most of the now largely deforested central highlands to coffee or vegetable farming; tribal hill peoples have been displaced everywhere, along with their own particular nutritional ecologies. The Philippines encouraged the resettlement of millions of people from the degraded and impoverished Catholic Visayas (central islands) to Mindanao, to replace its forests with oil palm and marginalize its small Moro population. Indonesia's national endeavor of transmigration moved millions of farmers from Java and Madura to the outer islands of Sumatra, Borneo, Sulawesi,

Halmahera, and finally their province of Irian Jaya on New Guinea. Millions of hectares of forest were converted to oil palm in particular, and to rubber, coffee, sugar, and even rice. When Soeharto fell and the Indonesian central government lost control, local people in many places tried to throw out the intruders who had received the benefits of development and taken their land. But many thousands of those settlers, days away from any medical care, had already died from malaria, diarrhea, and acute infections compounded by malnutrition. In those six countries alone live more than 3 billion people (45% of earth's total). Their population's growth and push to the agricultural frontier have changed the global environment profoundly.

In Latin America farmers are pushing into the Peten Reserve in Guatemala and the forest sanctuaries of Belize; from all sides they convert the forests of the Amazon to agriculture. Soybean expansion in Brazil has received much of the blame for deforestation, but farmers moving in along new logging roads and highways, slashing and burning forest for the cultivation of a few years, have been at it for decades. Today large numbers of farmers with little experience of wet tropical soils have moved down from the altiplano, where ancestrally they farmed potatoes and other crops at 12,000 feet. They have created one of the world's new hyperendemic malaria hotspots. In sub-Saharan Africa, which is at an earlier stage in the demographic transition and just beginning to lower its fertility, population is growing regionally at 2.5% and will more than double before approaching stabilization. In East Africa the value of tourism has made possible the preservation of wildlife habitat and their ancient ecologies, but farmers still struggle to clear new land and grow crops in the paths of migrating elephants. Roads for logging and mining have opened some of the Congo forest up, followed closely by a great expansion of hunting for "bush meat." Butchering of primates was almost certainly the source of HIV transfer to the human population. The conversion and farming expansion are more famous for the episodic eruption of Ebola and Marburg viruses from their disturbances. In Gabon and the adjacent Congo basin remains a vast wild land yet unpenetrated by farmers. From disturbance of the African land and its ancient primate ecologies, surprises are yet likely to emerge.

This basic geography is needed to connect what are usually reported and studied as single episodes, unique developments, or unexpected consequences. The 1998–1999 outbreak of a new virus named Nipah in Malaysia and Singapore, for example, killed 40% of the 257 people hospitalized for encephalitis during those years (Chua et al., 1999). The virus made their pigs sick and passed to pig farmers and butchers, but fortunately never became contagious among people. Later studies showed that the reservoir was a *Pteropus* bat species that had newly taken to eating fruit around the peninsular Malaysian pig farms, and had apparently infected the pigs through their droppings and contamination of fallen fruit. The bats responsible for the outbreak are thought to have been displaced by the destruction for agricultural conversion of their foraging forest across the Strait of Malacca in Sumatra.

Readers are referred for review to the section in Chapter 4 about ecological complications of agricultural conversion and other changes resulting from ecological simplification (see especially Figure 4.9). The following connections and comparison of land development in the agricultural frontiers of Malaysia and Ecuador

draw upon the Chapter 4 discussion. They illustrate both the commonality of the political ecology of ecological disturbances and exposed migrants at the frontier, and the importance of detail (species-specific, cultural, and micro-scale).

Population Resettlement: Malaysia, 1970s

Some agricultural conversion by migrants is spontaneous, carried out by individuals or families without much organization or legal authority; some is highly planned, organized, and capitalized. One of the best examples of the latter is the program developed by the Federal Land Development Authority (FELDA) in Malaysia. In the 1970s and 1980s it resettled hundreds of thousands of farmers who needed land on “land schemes” converting rain forest land to grow rubber trees and oil palms. Although not without political and economic problems, these land development schemes became models for other countries. They played an important role in making Malaysia internationally dominant in palm oil production, and for a generation they created decent livings for poor farmers who otherwise might have gone to the urban squatter slums. FELDA schemes have been much studied by political scientists and economists for their costs and benefits and for their organizational successes and failures in increasing human capital. Interested in the ecological alterations being wrought worldwide by land development programs, Meade (1976, 1977, 1978) studied the FELDA schemes’ impact on health over a 2-year period through engaging in participant observation; conducting long interviews in Malay and English; conducting a census of the settlement; maintaining a register of 60 households interviewed weekly in both the scheme and a comparative existing rubber farming village 10 miles away; and sampling children’s stools, mosquito larvae, and field rats and their mites.

People who had applied for land were admitted from all over peninsular Malaysia. They spoke different dialects, and in those schemes to which Chinese and Indian Malaysians were admitted along with Malays, they had different religions, diets, and other cultural practices. Most were far from their home villages for the first time. For the first few years, while the tree crops matured, they faced raw land that had to be cultivated, fertilized, drained, and planted. Some settlers could not speak to their neighbors and in any case were not used to mixing with strangers. It took 2–3 years for communities to form. Occasionally some lonely settlers, unable to get to town or visit home villages on the deliberately unpaved roads, left the scheme. Women, accustomed to giving birth to their first child in their mothers’ houses, were especially stressed. Malay women did not traditionally hoe plantation fields or work on roads, but they did these and other jobs on the land schemes. Sometimes older children had to stay home from school to take care of small children while the mothers worked, leaving the children to eat plain cold rice for lunch. The men were required to report to work early in the morning even in poor weather—a kind of regimentation previously unknown to many of them. There were no local ways to earn money when there was no labor required on their holdings.

Settlers’ houses (see Plate 9) at first were too small for large families, but later houses were enlarged and smaller families were selected. The corrugated iron roofs

were long-lasting and vermin-free, but because the whole area had been clear-felled and was exposed to tropical sun, the houses frequently heated to over 110°F (43°C) during the afternoon. Each house had land around it for gardening. It takes years for fruit trees to mature, however, and it was difficult for government agents to get the Malay settlers to grow vegetables (which are usually grown by Chinese farmers in Malaysia in the highlands). There were no fresh fruit or vegetables, or meat since men weren't at leisure to hunt, unless vendors made their way in or male settlers out to Saturday market (both on motorcycles). FELDA had a store with dry and canned food, but even that was open only during hours when male settlers were supposed to be in the field. FELDA also provided latrines (Plate 9), which many settlers had not previously had. Almost all the adults did use the latrines, but because toddlers are not diapered, some house compounds were fecally contaminated. A few children's play groups maintained helminthic infections until school age, and the school became a focus for infecting the children who had grown up in more sanitary homes. Every few houses shared a standpipe that brought chlorinated water. Although the system often broke down because of leakage and siltation problems at the source, the enteric infections and fevers in one such scheme, when studied in detail, were clearly reduced compared to those in the nearby traditional village.

The vectored diseases were the worst affected, and the worst among these was malaria. Many of the schemes constructed in rain forest areas repeated the experience of rubber plantations early in the 20th century. The ecosystem was simplified, and epidemic conditions were created (see Plate 9 and Figure 4.9). Vignette 4.1 (see especially its Figure 2) analyzes how the hazard of scrub typhus was created by clearing trees unnecessarily, bringing field rats into the paths of village life, and inverting the age and role exposure to mites compared to the traditional village.

Trees were cut down; commercial hardwood logs were removed, and the remainder were burned. This exposed the streams to sunshine and created breeding habitat for *Anopheles maculatus*. In Chapter 4 on the landscape ecology of malaria, Figure 4.5 illustrates the range of *Anopheles* mosquitoes and their habitats in Malaysia. Some could be controlled by spraying inside houses, but other species never entered houses. *A. umbrosus*, for example, shared the forest with other *Anopheles* species. It liked the deep shade of the rain forest. Although abundant, it preferred to bite other animals than people. It nevertheless occasionally transmitted malaria between people. Only *A. maculatus* liked the direct sunlight on water for breeding that was created by clear-felling. It quickly bred in huge numbers, the only species, with only people to eat. When rubber trees were first planted in clear-felled forest along the coast of the Strait of Malacca, Tamil laborers contracted from India were carried out in the thousands.

On the FELDA schemes, houses were sprayed with insecticide twice a year as part of Malaysia's rather successful campaign against malaria, but the initial success of chloroquine and insecticides led to neglect of traditional concern for drainage, shading, oiling, and other techniques that had originally been developed in Malaysia for those rubber plantations. The mosquitoes across Southeast Asia were becoming resistant to the insecticides. The clinics serving the remote areas of low-density population where new land schemes were established were not enlarged before

the settlers came; their capacity had been adequate for the presettler, mostly Semai aboriginal population. Settlers sick with malaria were therefore driven in Land Rovers for hours over rutted and flooded land to clinics with no capacity to treat them, were given chloroquine pills, and were sent home. The resulting underdosage, combined with new strains introduced from older settlements, led to increased prevalence of chloroquine-resistant malaria and its unusual expressions. In at least one scheme, unfamiliar cerebral malaria led to a perception of spirit possession and consequent abandonment of the land. As the schemes matured and crop trees grew to shade the water, the epidemic receded.

In the less capitalized settlements of Indonesian transmigration, malaria, diarrhea, respiratory illness, and malnutrition plagued early decades of developments. The responsible ministry learned to do better over time. Construction of houses, roads, and economic supports; health service provision; food marketing; and labor opportunities in the interim before tree crop production matured were improved. Health education, family planning, and bed netting were all improved in established areas, although the insertion of land development and transmigration of ethnic groups into new lands of other islands with different ethnicities and religions caused increased instability. Despite all that was learned about the effects of agricultural conversion on disease ecology in the forests of Southeast Asia, little seems to have translated across the seas by government health or forest ministries, or by academic researchers. Policy issues and disease ecologies from conversion of forests in Latin America or Africa are repeating themselves.

Population Resettlement: Amazon, 2000s

Most Latin American countries are quite late in the demographic transition now, although a few haven't progressed as far as the others (data from Population Reference Bureau). Total fertility rates for the region as a whole are down to 2.5 children per family. Fully 76% of the regional population is urban. Yet, because of the relatively recent decline of fertility, more than 30% of the population is still under 15. This is shown in the 2% net out-migration to Europe and the United States. It is also shown in the changing push to the agricultural frontier. As Brazil, the world's fifth most populous country, has become an urban industrial giant, gold mining, commercial logging, and capitalized agribusiness have become greater threats than are poor farmers from the northeastern region slashing poor plots. Farmers still move into the Peten in Guatemala, but more are turning to cities and international work. In the western Amazon, however, agricultural clearance is increasing as Andean farmers seek new land. The Peruvian Amazon in the late 1990s became one of the world's hot spots for hyperendemic malaria.

South America offers unique challenges to frontier migrants, as well as reprising others from tropical Asia. Before the Columbian exchange brought malaria, yellow fever, river blindness, schistosomiasis, and so many contagious diseases, leishmaniasis and especially Chagas disease "protected" some tropical lowlands from settlement. The vectors of these diseases are listed in the tables of Chapter 4 on landscape epidemiology. Chagas disease in particular continues to kill sev-

eral times more people than malaria in the Western Hemisphere. The wild animal reservoir of this natural nidal disease includes more than 80 species, from sloth to dog. Its vector, a reduviid or “kissing bug,” has species variants that allow it to be transmitted from the lower Andes to the grasslands of the Grand Chaco and the central Mexican plateau. Its primary nidus is in the wet tropical lowlands. Chagas disease is also called American trypanosomiasis because its disease agent is a trypanosome, a kind of protozoan like the agent of African trypanosomiasis (sleeping sickness). After causing initial symptoms for a few days, the American trypanosome reproduces in muscle and circulates in the blood for years, even more than a decade, before people die from enormously enlarged hearts and dysfunctional colons. Metacolon has been found in many pre-Columbian mummies, indicating that the person had been unable to defecate for more than a month.

The reduviid bug, which transmits Chagas disease, resembles in its habitat, behavior, and size a cone-nosed cockroach. As noted in Chapter 4, however, this bug sucks blood and prefers the lips of sleeping people (hence its moniker, “kissing bug”). After its meal the bug defecates, and people then rub the trypanosome-laden bug feces into the wound or mucous membranes. The reduviid bug is one of the disease vectors most clearly affected by substitution of iron roofs for thatch ones. A memorable image from a reading decades ago (a lost reference) is of a “rain of feces upon the mucous membranes of people sleeping under the thatch roofs.” Like cockroaches, reduviid bugs live in piles of firewood, old leaves, cracks, and other such dark and dirty places (e.g., behind refrigerators). They have accompanied migrants to the urban slums and spread Chagas disease there. The trypanosome has even accompanied migrants to the United States and contaminated the U.S. blood supply. When blood is donated to the Red Cross now, a series of screening questions is asked in an attempt to eliminate possible exposure to Chagas. Remember that the latent period of this disease, during which hosts are unaware of growing infection and infectivity of their blood, is more than 10 years.

The best protection against Chagas disease is good housing in which the bugs find no habitat. Usually when settlers leave their home villages and family houses and strike off to the frontier today, they move to poor housing (see Plate 10), but at least seldom to thatch roofs. Houses in these settlements do offer entrance to mosquitoes, though—*Anopheles* at night and *Ae. aegypti*, the vector of yellow and dengue fevers, by day. The mosquitoes rest under furniture and clothing hanging from nails on the wall, and breed in abundant “container” water around homes. Water-borne disease is common. Latrines are rather haphazard, and wells are often nearby. The main improvement between the Ecuadorian frontier homes of Plate 10 and the Malaysian ones of Plate 9 from 30 years earlier is the presence of electricity; the worst comparative note is the absence of safe drinking water, which the Malaysian government did provide.

The satellite image in Plate 10 shows two rivers, two small urban areas marked with the orange circles, and two of the settlements studied by Polo (2009) and pictured. The roads and agricultural clearings (including a large planting of oil palm in the blue circle) can be seen to be cut from the depths of the Amazon rain forest in fields and pieces, not clear-felled and logged over 20 miles at a time. As in

Malaysia, rodents were disturbed and dispersed. Local field species must have been changed, along with their viruses and the likelihood of human contact with their urine aerosols, which can cause Bolivian (Machupo), Venezuelan (Guanarito), and Argentinian (Junin) hemorrhagic fevers. Monkeys and other mammal reservoirs (like bats); many species and whole ecologies of birds; and uncountable species of mosquitoes, bugs, acarids, and amphibians have been exterminated, replaced, displaced, promoted, and transported in this ecologically simplified habitat. With all this disruption, the potential for human disease and its epidemic eruption has been fundamentally altered.

The cultural ecology of malaria offers dramatic contrasts and similarities to those of agricultural conversion and population resettlement in Southeast Asia. Let's start with the fact that the large clear-felled conversions for agribusiness in Brazil have often changed dense rain forest into monocropped soybean fields or cattle grazing, but have resulted in relatively little malaria. For more than a century, there has been massive deforestation for cattle ranching, mining, and agricultural conversion. Malaria has been associated with this change in land use and land cover, but the powerful repeated links of epidemics with forest clearing have not been clear, as they were in the Malaysian experience.

Efforts to use satellite images of water flooding and malaria potential in Belize found that *Anopheles pseudopunctipennis* existed at densities too low for sampling change. The main vector of epidemic malaria in South America is *Anopheles darlingi*. Singer and Caldas de Castro (2006) explain how the linkage of deforestation, conversion, and malaria must be analyzed at several temporal and spatial scales, starting with the specific ecology of the vector and how it is affected by transformation of the ecosystem. These researchers characterize the larval habitat of *A. darlingi* "as zones of partial shade, proximal to a forest fringe or river edge, in relatively deep and clear water of high pH (relative to the more acidic Amazon streams), and with temperature greater than 25°C" (p. 1). Contrast this with the Malaysian vector, *A. maculatus*, which breeds in shallow water in direct sunshine and has intersecting vegetation to anchor eggs. The large area, fast clearance, and conversion of the corporate sector development in the Amazon has created habitats largely inimical to the larvae of the most efficient vector. In contrast, it is the government-sponsored and spontaneous colonization of the forest that has resulted in epidemic malaria. Compare the settlement habitats in Plates 9 and 10: notice the trees in the settlement and along the riparian background.

Singer and Caldas de Castro (2006) pick up and extend the relevance of Sawyer's (1988) stages of *frontier malaria*. At the micro/individual level of *colonistas* moving into the primary forest, the shaded habitat and riverine areas are very suitable for *A. darlingi* larvae, and indeed their population is dense. In the small clearings where the settlers work, the exposure is intensive. *Plasmodium falciparum* is the primary *Plasmodium* species in these conditions, spreading more rapidly than *P. vivax*. Settlers have limited understanding of transmission. Their houses readily admit mosquitoes, and the insides provide poor surfaces for residual spraying of insecticides (mosquitoes leaving). Little medical care or malaria prophylaxis is available. At the community level there are poor organization, few management efforts to

minimize malaria risk, little health service availability, and high rates of both in- and out-migration (which circulate infected and susceptible individuals). Some settlers try to escape poor land and malaria by development of new unplanned areas, but this process only promotes more transmission. After 6–8 years the population settles, and roads and impervious surfaces create habitat inhospitable to *A. darlingi* larvae. People move away from riparian areas. Clinics and health care increase. Population immunity grows, nutrition improves, and the epidemic subsides into endemic malaria as background.

This scenario of frontier malaria in transition also fits the experience of Southeast Asia in both time and space scales as well as human processes. There is one exception, posing a paradox that is important for conservation. In Malaysia conservation of forest habitat, which promotes biodiversity, thereby promotes multiple species as vectors of malaria—each with fewer members, less contact, lower intensity, and lower disease transmission. When it is pointed out that in the Amazon “altering the landscape likely plays an even larger role than people migrating into the jungle,” and reported that *A. darlingi* bites people 300 times more after areas are cleared than before, what is being described is the frontier malaria of colonization (Vittor et al., 2006). The statement “Conservation policy is one and the same with public health policy. It’s probable that protected conservation areas may ultimately be an important tool in our disease prevention strategies,” describes tropical Asia more than the Amazon and raises the paradox. There, it seems, the corporate large-scale total clearing and replacing are what most protect the workers and most threaten biodiversity; the individual settlers on their still-forested clearings are most at risk.

This chapter could go on for a hundred pages recounting the health impact of developmental changes, but self-control is in order for your authors. We focus in the remainder of this chapter on three aspects: other rural impacts, mostly as expressed in African ecologies and especially from impoundment of water; urbanization and the development of new urban forms and ecologies; and globalization of movements, including people, food, and lifestyles.

OTHER DEVELOPMENT IMPACTS ON RURAL ECOLOGIES

Many of the health problems of both new settlers and old villagers are simply those of protracted poverty. Farmers have their land become fragmented, leached of nutrients, eroded by wind and rain; their livestock are depleted. They become landless laborers, who must move to the frontier or the city or internationally to sell their labor. Little health care is available. The water is not safe to drink. The billion people living on less than U.S. \$1 a day cannot afford to eat balanced diets. Those who become involved in capital-intensive agribusiness may have money to buy food grown elsewhere; to get vaccinations and prophylactic drugs; and to have screened housing, deeper wells, less dangerous work, and access to primary health care when they do get sick. Their actions can induce different adverse disease ecologies, however. Increased cultivation of cotton and other crops for export usually leads to a greatly increased exposure to insecticides, the use of which produces

disease vectors highly resistant to it. The contaminated environment can initiate the late-transition-stage diseases of cancer and genetic alteration in fetuses. The stresses resulting from enormous change itself, and from ever-increasing economic and social inequity, have (as in India) caused suicide to become a major cause of death.

As has been mentioned earlier, sub-Saharan Africa is the world region at the earliest stage in the demographic transition (see Table 5.1). Shocking data easily accessible from the WHO statistical Information System (WHOSIS) database on the WHO website (www.who.int/whosis/en/index.html) confirm this. Infant mortality is almost 1 in 10; maternal mortality is high (45% of the world's); and life expectancy for the region as a whole is 49 years. Multiple infections with different types of helminths and protozoa remain common. Plates 6, 7, and 8 show the global prevalences of HIV, malaria, and tuberculosis, and it is clear that sub-Saharan Africa has more than its share of all these diseases. Southern Africa is hardest hit by all of them. With only 12% of the world's population, more than 40% of the world's diarrheal mortality occurs in this region. Much of the rural African population continues to have close contact with vectored disease. Yellow fever still kills "only" 20,000–30,000 annually. The three diseases that can render good agricultural land totally uninhabitable—malaria, river blindness, and sleeping sickness—evolved in the African realm. Indeed, 90% of the world's malaria mortality, and now 75% of its HIV mortality, occur in this region; only 6% of its cancer does, however.

In recent years, major progress has been made through the United Nations' millennial goals; WHO initiatives and global funds targeting malaria, HIV, and tuberculosis; the development of GIS for integration and surveillance; and major funding from the Gates Foundation for research on new drugs and treatments for the neglected tropical diseases. Onchocerciasis in West Africa has been the target of one of the most successful eradication efforts in world history; surveillance is now maintained by vigorous community participation and powerful GIS integration, so that in 2001 more than 40 million people, living in 117,000 hard-to-reach villages, regularly received ivermectin (first developed in the United States to combat the filaria responsible for canine heartworm) for prophylaxis. The WHOSIS database (see above) also shows 2000–2002 measles deaths were down 91%; leprosy had also been eliminated in 44 of 46 countries (but 41,239 people were newly infected in 2002).

The medical geography of developmental changes in Africa has been both the first, thanks to the work of John Hunter (Chapter 1), and the least investigated of any region. The initial comprehensive review by Hughes and Hunter (1970) of the literature on development and its impact on health in Africa sensitized a generation of scholars and experts. Hughes and Hunter described, for example, how the evacuation of tribes from their agricultural and grazing land was carried out by colonial governments because of outbreaks of rinderpest or smallpox or in order to stop tribal warfare. Such evacuations allowed brush and tsetse to claim the abandoned territory, so that returning herdsmen or farmers years later found the land denied to them by trypanosomiasis. They also described how roads had become linear transmission sites for disease. Tsetse flies (which are attracted to moving vehicles), black-

flies, mosquitoes, and the disease agents they vector became concentrated where roads crossed streams and people stopped to eat, wash, and urinate. New water reservoirs and roads meant new exposures and transmission to new areas. Hughes and Hunter also first described some of the nutritional and social-psychological consequences of development, as well as the difficulty of intervening in the complex disease systems and multiple levels of infection. They emphasized most of all, however, the need for a comprehensive, ecologically informed approach to developmental projects designed to break the synergism between poverty and disease. Since then, examples of the inadvertent consequences of developmental changes in Africa and everywhere else have multiplied. Yet their political ecology and the health promotion efforts of current massive, vigorous educational campaigns have been little studied. Guinea worm, river blindness, and sleeping sickness have been almost eradicated, for example, even as blood pressure, diabetes, and violent death soar. It is not only the rapidly urbanizing population that is so neglected by research, but the rural conditions of modernizing agriculture also.

Agricultural development in Africa, for example, affects the cultural–ecological process of animal herding. There are increased penning, greater control of breeding, and higher-quality feed. The nature of barns and enclosures often changes. Animals do less scavenging. They are harvested or culled more regularly, so that the age structure of the herd becomes younger, and the proportion of immune animals to susceptible ones changes. The socioeconomic role of contact with animals may change. It may come to involve more men than boys and may become an occupational specialty rather than a source of general population exposure. Replacement of draft animals by machines may deprive mosquitoes of their source of blood meals and displace them to less favored meals from people, causing new or intensified disease transmission. As developmental changes affect the quality and resistance of animal and human hosts, increase the intensity of animal herding, displace animals, change occupational connections and gender roles, modify water and food pathways, and alter habitat for arthropod vectors, the significance and distribution of various diseases transmitted from animals to humans will also be affected. These interactions, interventions, and promotions remain in great need of research.

Irrigation and Water Impoundment

Nothing except the use of fire can compare with the ecological consequences of the human ability to control water. There are today tens of thousands of large dams and water impoundments; millions of tube wells pumping water up from the depths; and hundreds of millions of hectares of land under irrigation. Mountainsides have been terraced for wet rice cultivation; swamps have been drained over millions of square miles. The particular health hazards resulting from impoundment of water and irrigation of crops have long been recognized, although not often evaluated by engineers and economists. Much of the present appreciation of their importance stems from the construction of the Aswan High Dam in Egypt and the spread of schistosomiasis.

Basin irrigation was an ancient practice in upper Egypt. The basins (*tanks*, in British English) constructed to store the floodwaters of the Nile were eventually drained to supply a continuing flow of water to irrigated fields. Subsequently they were themselves planted to crops, and eventually dried up completely before again being flooded by the rising river. At the height of Ethiopia's rainy season, the flooding Nile brought vital silt and water to the lower river valley in Egypt. It also washed out and damaged the snails and their eggs that had survived the dry season, thus serving to control the snail population. With the coming of the Aswan High Dam and perennial irrigation, the snail population expanded greatly. Because the nutrient-rich irrigation water flowed only slowly through the agricultural fields, was cleared of silt by reservoir sedimentation, and often supported aquatic vegetation, it was an excellent snail habitat. Schistosomiasis, which had been endemic at low levels in the region from the time of the Pharaohs, exploded to become a universal scourge. The same situation has been repeated monotonously, from the Gezira irrigation scheme in the Sudan to Lake Volta in Ghana. Africa is a dry continent. Irrigation is absolutely crucial to countries' hopes of feeding their growing populations, but the price is high. In Kenya, in a district on the shores of Lake Victoria, the rate of schistosomiasis infection for children entering elementary school went from virtually nothing to 100% in less than 10 years when an irrigation system was developed.

In Brazil, too, expanding irrigation means expanding the territory of schistosomiasis. In Southeast Asia there are only a few foci, including a small but critical one in the Mekong River. As some of the plans for damming and controlling that great river materialize, now that there is peace and many dams on tributaries have been completed, one of the feared consequences is the spread of schistosomiasis throughout the Mekong River basin. Much research concern for the spread and control of schistosomiasis (*S. japonica* in the Yangzi lake basins) has also followed the completion of the Three Gorges Dam in China.

Water impoundment and irrigation have affected vectored diseases ranging from California encephalitis to lymphatic filariasis in Ghana, from clonorchiasis (liver fluke) in Thailand to the release of heavy metals and poisons from sediments as old dams are removed in the United States today. The way water was impounded by the Tennessee Valley Authority in the U.S. Southeast without increasing endemic malaria has been discussed in Chapter 4. Throughout the ecologically more diverse tropical world, however, impoundment continues to increase malaria hazards.

Water impoundment schemes/projects/structures have undoubtedly contributed to economic development in many Third World nations. In Bangladesh, for example, government and NGO-funded water containment structures have protected rural populations from annual floods caused by monsoon rains and typhoons. Earthen embankments circling households and fields decrease deaths due to drowning and allow farmers to double- or triple-crop in the growing season, as well as to plant genetically modified high-yield varieties of rice. At the same time, however, such structures may act to increase the risks of infection from bacteria and viruses, causing diarrheal diseases like cholera or dysentery. By disrupting

the annual cycling of water through the system, embankments prevent the flushing and replenishing of the ponds and streams that supply households' drinking and cooking water. In some cases, this increases the densities of microorganisms to which residents are exposed.

As we discuss in Chapter 6 on the hazards of climate, on the other hand, water stored in reservoirs is absolutely essential for providing "safe drinking water" and water for hygiene and sanitation to growing urban populations. The water-borne and water-unwashed diseases (see Chapter 2) can kill far more people more quickly through contagion than all the vectored diseases combined.

WORLD URBANIZATION AND CHANGING DISEASE ECOLOGIES

One of the most profound changes in human history, a truly new thing, is occurring: As noted at the start of this chapter, over half of the human beings on earth are now urban.* A few decades ago, as thousands of years ago, more than 80% of people were rural and agricultural. A few decades in the future, all projections indicate that more than 70% of people will be urban. This momentous population migration is transforming livelihoods, economic conditions, and basic needs for shelter, schooling, and health care. It is also altering habitats and disease ecologies fundamentally, and connecting them through mobility of hypercirculation. The enormous quantitative changes in conditions may dialectically create qualitative changes in the state of human health, on a scale not seen since the first cities led to the creation of most contagious diseases. Boyden (1970), McKeown (1988), and Dubos (1965) have described historically how urbanization can change the state of everything. We humans are now writing a new chapter of civilization.

Dimensions of Urban Change

Urbanization is the process of transition from rural settlements and society, to urban settlements and society primarily through rural–urban migration. This destination in the mobility transition became dominant in the now more developed countries during the first half of the 20th century. In the United States, for example, those farm boys who had seen Paris couldn't return to their rural isolation. Southern sharecroppers headed for new northern assembly lines. African Americans migrated by the millions from southern rural poverty to northern cities. The erstwhile agricultural

* *Urban* settlements or localities are defined for the United Nations by national statistical agencies according to remarkably different criteria. In India, for example, a nuclear settlement can have schools, a hospital, and a population of over 40,000, but may not be defined as urban because its population is more than 80% agriculturally engaged. The U.S. Census Bureau defines urban areas as those consisting of a large central place (density at least 1,000 people per square mile in core block groups and 500 in surrounding census blocks) and a total population of at least 2,500 for urban clusters, or at least 50,000 for urbanized areas. Urban classification cuts across other hierarchies, and urban areas can be in metropolitan or nonmetropolitan areas.

frontier, in the Great Plains generally, “hollowed out” America as population moved to live in cities within 200 miles of the coast or Great Lakes (for reasons of economic geography). As the developing world goes through the demographic and mobility transitions so much faster with such a larger population base, the “second wave” of urbanization involves human relocation and adjustment at a size and speed that are almost incomprehensible.

Table 5.4 shows the dimensions of the urban change occurring around the world. Altogether, the population of countries in the developing regions of the world (as of the year 2000) is almost 5 billion. It is only 40% urban. Three-fifths of that population lives on less than U.S. \$2 a day. Sanitation is adequate for 88% of that population, buoyed by Iran, Turkey, and great old cities. (*Improved sanitation* is defined by the United Nations Human Settlements Programme [UNHSP, 2003, p. 243] as having either private or shared access to connection to a well-functioning sewer, a septic tank with sufficient capacity, or a well-ventilated improved latrine that prevents contact between human excreta and human bodies.)

Two-fifths of the urban population lives in slums—a word that has come to incorporate former usages such as *squatter settlements* and *shantytowns*, not just the former old-city overcrowded dilapidated areas. Most of Latin America, latest and almost done in the demographic transition, is already highly urban; its great rural–urban migration was a phenomenon of the 1970s and 1980s. Brazil, the population giant, is already more than 80% urban, but it is projected to grow by 2020 to 89%. This means it will add to its cities 48,994,000 people. The Asian subcontinent (India, Pakistan, and Bangladesh) is still growing from 1.7% to 2.2% a year. That natural increase, whether it is born in the cities or the rural areas, will become almost entirely urban. Tens of millions now underemployed and malnourished in rural areas will migrate to the cities. The subcontinent will add to its cities a phenomenal quarter of a billion people within little more than a decade, and even more people over the following decades as the three countries become majority urban. China alone will add to its cities more than 300 million by 2020, 200 million of whom are estimated to be the “floating population” living on trains, on boats, and in unofficial urban settlements today. Sub-Saharan Africa still will more than double in population by natural increase, even as it urbanizes. Already 72% of its urban population lives in squalid slums, and only 80% has sanitation. As the rural population migrates to the cities, this poorest of regions will add through natural increase and migration a quarter of a billion people to those cities. How can cities in developing countries possibly provide adequate shelter, sanitation, water, schooling, and health care to that urban population?

Cities grow by three methods. The first of these is through natural increase. Second, they add or annex surrounding territory to themselves, usually because it too has become urbanized and needs sewerage, transportation for commuting, electricity, and medical and police services. And most of all, cities grow through in-migration of people. As the demographic transition progresses, access in cities to vaccination and health care, basic education, better nutrition, more information, jobs, and opportunities for women to contribute to their families by means other than producing another child all lead to lower fertility rates even as child mortality

TABLE 5.4. Dimensions of Urban Change

Location	2000 population (thousands) ^a	% below U.S. \$/day ^b	% urban ^a		Urban population added (thousands), 2000–2020 ^a	Urban: % slum ^a	Sanitation: % improved ^c
			2000	2020			
Developed regions	1,191,275	18.5	75.3	79.9	75,724	6.0	99.9
Developing regions	4,865,393	60.4	40.4	51.3	1,299,446	43.0	88.1
Northern Africa Egypt	143,054	21.3	51.1 42.7	59.7 48.2	39,345 14,282	28.2	87.0
Sub-Saharan Africa Nigeria Kenya	650,572	73.5	34.0 44.1 33.4	45.7 58.3 50.8	254,834 57,253 11,476	71.9	80.4
Latin Amer. and Caribbean Brazil Haiti Mexico	518,935	33.2	75.4 81.2 35.7 74.4	81.8 88.9 48.8 79.2	151,838 48,994 2,419 25,486	31.9	86.2
Eastern Asia China	1,353,978	34.9	38.2 35.8	55.1 53.4	328,958 315,521	36.4	69.5
South Central Asia India Bangladesh Pakistan	1,480,868	57.8	29.8 27.7 25.0 33.1	37.8 34.7 37.7 42.4	307,962 168,490 40,078 49,777	58.0	65.9
Southeastern Asia Indonesia Vietnam	522,121	52.2	37.5 41.0 24.1	51.1 58.4 34.7	141,578 66,063 15,954	28.0	75.8
Western Asia Iran Turkey	188,275	19.4	64.7 27.7 65.8	69.8 34.7 73.7	73,138 66,063 17,216	33.1	97.9
Oceania Fiji	7,590	—	26.4 49.4	33.8 62.8	1,793 190	24.1	87.2

Note. Country data from UNHSP (2003, Table B.2).

^aData from United Nations Human Settlements Programme (UNHSP) (2003, Table A.1).

^bData from UNHSP (2003, Table A.3).

^cData from UNHSP (2003, Table A.2).

falls. Born-urban dwellers lower their fertility toward replacement level. In country after country, rural–urban immigrants have adjusted to urban life and values, lowering their fertility rates to between the city rates and those of the rural areas they came from. Since 70% or more of urban populations can be migrants from rural areas, early on these populations grow substantially through natural increase. This was the European American experience also. The migration experience in the developing world is different, however.

In the first wave of urbanization, most migration was what is known in population geography as *step migration*. Few people went straight from the sharecropped, eroded Southern fields directly to Chicago, for example; instead, like jazz, they moved up the river cities. People went first to nearby cities, then moved to larger regional cities, and generally moved up the urban hierarchy (see Chapter 10) as they improved their urban living skills and extended their information fields. A city like Durham, North Carolina could become a financial, insurance, job, and cultural “metropolis” for African Americans driven by poverty from the cotton fields. In the colonized developing world, by contrast, one or two cities (usually ports) became—by virtue of their infrastructure (railroads, ports, electricity, sewerage, plumbing) and institutions (government, military, educational, hospital, commercial, and [after independence] manufacturing and investment)—*primate* cities that dominated the urban hierarchy. A normal ratio in an urban hierarchy of largest city to second is 2:1; for instance, Charlotte is twice the size of Raleigh, North Carolina, or New York City is twice the size of Los Angeles. New York City, while reaching 17.9 million people in 2015, will have only 6.9% of the urban American population. In comparison, think of the ratio in the urban hierarchy of Mexico City, Manila, or Kinshasa to the second city in Mexico, the Philippines, or the Democratic Republic of the Congo. This situation is known as *primacy*, and in such urban systems the rural–urban migration does not go to places throughout the urban hierarchy. Rural people move to the one primate destination through *chain migration*. Thus Bangkok, which has been more than 40 times the size of Thailand’s second city, could grow from 5.9 million in 1990 to 7.4 in 2000, 8.9 in 2010, and to 9.8 million in 2015, while decreasing only from having 59.2% to having 55.9% of all the urban people in the country. In contrast, India, which had a regular subcontinent-wide urban hierarchy before dominance by Europe, has seen Mumbai (formerly Bombay) grow from 12.3 million in 1990 to 16 million in 2000 to a projected behemoth 22.6 million in 2015, but it will have “only” 5.7% of India’s urban population. (All data on cities in this paragraph are from UNHSP, 2003, Table C-1.)

Most of the global accretion of population to urban areas will occur in medium-sized cities with hundreds of thousands or a few million today. The chain migration into existing primate cities, however, will lead to the growth of *megacities* of more than 15 and even more than 20 million (bigger than almost three-quarters of the countries in the world today). Table 5.5 shows the projected population of cities over 10 million in 2015—a projection only a few years away at this writing. Among cities in currently developed countries, only Tokyo and New York City remain in the top 10, with Los Angeles and Osaka among the top 23. The great cities of Europe are not among them. Sinking, polluted, gridlocked Lagos will have more than 20 million

TABLE 5.5. Megacities Soon: Cities with More Than 10 Million People

2015 rank	City	2015 projected population (thousands)	% urban share
1.	Tokyo, Japan	26,444	25.7
2.	Mumbai, India	26,138	6.0
3.	Lagos, Nigeria	23,173	27.3
4.	Dhaka, Bangladesh	21,119	38.6
5.	Sao Paulo, Brazil	20,397	11.8
6.	Karachi, Pakistan	19,211	18.5
7.	Mexico City, Mexico	19,180	20.7
8.	New York City, United States	17,432	7.0
9.	Jakarta, Indonesia	17,256	12.6
10.	Calcutta, India	17,252	4.0
11.	Delhi, India	16,806	3.9
12.	Manila, Philippines	14,825	22.6
13.	Shanghai, China	14,575	2.5
14.	Los Angeles, United States	14,080	5.7
15.	Buenos Aires, Argentina	14,076	34.9
16.	Cairo, Egypt	13,751	31.5
17.	Istanbul, Turkey	12,492	18.4
18.	Beijing, China	12,299	2.1
19.	Rio de Janeiro, Brazil	11,905	6.9
20.	Osaka, Japan	11,013	10.7
21.	Tianjin, China	10,713	1.9
22.	Hyderabad, India	10,457	2.4
23.	Bangkok, Thailand	10,143	50.2

Note. Data from UNHSP (2003, Table C.1).

souls, as will Dhaka. Notice their primacy in the percentage of urban population in their countries. Six of these extremely large populations will be in the Asian sub-continent, three in China, three in Southeast Asia, and four in Latin America. The social, political, and economic consequences can be considered elsewhere; think of the disease ecologies!

New Urban Forms and Cultural Ecologies

Both the numbers and the speed of this second wave of urbanization make the growth of world cities an amazing phenomenon with an unknown future. Except for Chinese cities, these growing habitats are mostly tropical. Two new urban forms have been identified, each with its own implications for disease ecology. The first has been introduced above: *megacities*. A second form has been labeled a *desakota*, from the Indonesian words for “large village” and for “city” (McGee, Ginsburg, & Koppel, 1991). In Asia, the labor needs and food produced by wet-rice cultivation have often resulted, as on Java, in high agricultural population densities of many hundreds of people per square mile. This rural population comes with market cen-

ters and nuclear settlements in large villages with religious centers, schools, often medical facilities, and (recently) outsourced industrial piecework. As the cities grow into megalopolises like Jakarta, their sprawl encompasses these existing dense populations within it. Thus rice field mosquitoes, viruses of pig waste and waterfowl, road tars and oils, industrial chemicals such as dyes and lysing agents, batteries, and pesticide containers all come to coincide in space with very high densities of highly mobile populations susceptible to contagious diseases, from cholera to AIDS to tuberculosis. As discussed below, vectored infectious diseases are a new threat in large cities.

On the left side of Plate 11, a sequence of three photos, part of a longitudinal comparison photographed every 5–8 years over a 40-year period, shows what the growth of Bangkok looked like. In 1966 (upper left) Bangkok had about 4 million people. It was a sprawling city of Nanyang (Southern Seas) Chinese two-or three-story shop-houses in the upper delta of the Chao Praya River. Most of its canals had recently been paved over for roads, which interfered with the flushing of the remaining canals. Its skyline was dominated by hundreds of stupas marking active monasteries and the daily presence of Theravada Buddhism in people's lives. Besides the government buildings and shops, many of its buildings were of wood (often on stilts in this land of monsoon rain), with corrugated iron roofs. Most of the city dwellers and businesses burned charcoal for fuel. Buses burned a smoky, dirty diesel loaded with lead (which was present in gasoline, too). The most serious disease was tuberculosis, and of course some diarrhea. The city's water system was many decades old. Water left the treatment plants safe, but after flowing under sewerage canals, old cracked pipes sometimes sucked in contamination. Rural migrants from the impoverished northeast—coming especially when the crop failed for lack of rain—crowded squatter slums along the sides of the railroad tracks for miles out of Bangkok, and filled every lowland and empty building lot held for speculation. Rats and roaches were uncountable. *Aedes aegypti* were legion, breeding in boats, rain barrels, garbage, and stagnant canals. Epidemics of dengue hemorrhagic fever were beginning to put tens of thousands of children into hospitals.

By the second picture (middle left), taken from the same temple top, the canal had been cleaned up. The squatter dwellings have been cleared out, and brick and cement housing built. High-rise buildings were sprouting up. The Vietnam War, in which Thailand was an active ally, was over. American soldiers' R&R had ended, so the wartime commercial sex trade was scrambling for new customers. The lead was still in the gasoline, but bottled natural gas had replaced charcoal for most commercial establishments. A large proportion of the squatters had been rehoused in high-rise apartments.

The country then entered upon a decade of double-digit economic growth. The third picture (lower left) shows the transformations brought by that growth. Now a city of 9 million, Bangkok still housed more than half the urban population of Thailand. Think what a different disease ecology it offered from the first picture! Most of the mosquitoes and the open sewerage were gone, but the city was sinking into the river mud as the aquifer was pumped for potable water. The leaded gasoline was gone, but cheap, dirty (particle) diesel and gridlock new-car traffic created

smothering smog. Tuberculosis was still a problem, but now AIDS had appeared and was spreading through the sex industry in the worst epidemic outside Africa. (Thailand a few years later became the first country to stop and reverse the epidemic.) Arriving new rural migrants by the tens of thousands had been pushed out toward the port, forming a very large slum. Family planning services had lowered fertility to two children, and almost total vaccination coverage had been achieved. Bangkok had now become a cosmopolitan world hub, a center of global tourism and other mobility, soon to exceed 10 million (Table 5.5).

The top right picture in Plate 11 was taken on the streets of Bombay (as it was then still called), but could have been taken in Calcutta. That baby was probably born on the sidewalk. People cooked and ate there, defecated there, slept there, and died there. The sweetmeats being cooked could be carried and sold all over the city. The water came from the roof gutter. The fuelwood to cook cost more than half of their income, and these people were definitely living on less than U.S. \$1 a day. The next step up was to get a plastic sheet over one's head, or perhaps a roof made of crushed tin cans, along the tidal estuaries full of raw sewerage. Dengue fever, tuberculosis, and sexually transmitted diseases were here. At night the rats gathered for the filth. From such conditions, food stored in local cities when the monsoon and crops had failed, a bubonic plague epidemic began in Surin, India in the 1990s and triggered a global alert. Every day the buses and trains rolled in and thousands more rural–urban migrants got off to stay because chances for food and health care and jobs were so much worse in the rural areas at the maximum natural increase of the demographic transition.

These pictures are of Asian cities, but could be matched today in Lagos or Nairobi, or in Lima or Quito. The picture (middle right) is of a settlement near, or perhaps now within, Dhaka: a *desakota*. High-rise buildings form a frame along a highway past a large village whose population farms intensively, and markets its food and goods. Chickens, goats, children, wild birds, rats, fleas, and mosquitoes share their viruses promiscuously. Increasingly there is risk of vectored disease—something that used to be almost unknown in urban areas, except for flea- and louse-borne disease.

Urbanization of Vectored Diseases

The health hazards discussed previously as part of the urbanization process today are not very different in kind from the European American experience of the past century. One important difference evolving, however, is the urbanization of vectored diseases. The European model of urbanization assumes that the city population is removed from vectored diseases, except initially for those vectored by fleas and human lice. This is not necessarily true today (Azad, Radulovic, Higgins, Noden, & Troyer, 1997; Schriefer, 1994). Surtees (1971) was prescient in his concern for urbanized vectors. The vectored diseases of malaria, dengue hemorrhagic fever, Chagas disease, and especially lymphatic filariasis are spreading in today's developing cities.

Open drainage and septic ditches, construction sites, garbage mountains, and water storage have created simple ecosystems with abundance of vectors. In India, *Anopheles stephensi* likes to bite people; has become highly insecticide-resistant; and breeds prolifically in the gravel mines, construction digs, and garbage pooling of the periurban area. Assisted by population mobility and some other *Anopheles*, it has created epidemic urban malaria. Similarly, Keiser and colleagues (2004) have estimated that 200 million people in Africa—a quarter of the total population currently living in an urban setting (on 1.1–1.6% of land)—are at risk of getting malaria.

The mosquito *Culex fatigans* especially likes polluted drainage and septic systems for breeding. (The picture at lower right in Plate 11 shows such an open-ditch habitat in an urban area in Malawi.) As Surtees had foreseen, it made Brugian filariasis such an endemic urban vectored disease today that Indian scientists (Dhanda, Das, Lal, Srinivasan, & Ramaiah, 1996) wrote that filariasis, which they considered an urban disease, was rapidly emerging as a major problem in rural areas! Hunter (1993) described filariasis (elephantiasis) spreading among growing towns in Ghana. Erlanger and colleagues (2005) estimate that there are 2 billion people at risk of lymphatic filariasis in the world, 394 million of them urban dwellers without improved sanitation. Environmental changes due to water resource development and management, they find, consistently lead to a shift in vector species composition and generally to a strong proliferation of vector populations.

As discussed in Chapter 4, dengue fever transmitted by the domiciliated *Aedes aegypti* seems to have been transformed into a hemorrhagic form through the intense transmission found in urban areas. It is now, with the proliferation of reintroduced *Ae. aegypti*, spreading in the Caribbean and Central America. Yellow fever remains an urban threat in Africa. As mentioned earlier, Chagas disease has gone with people to the cities, where the reduviid bugs thrive in poor urban housing and domiciliated animals. Controlling its spread through the blood supply has become a problem reaching, through population mobility, even into the United States.

New Urbanization in the United States

Urban areas are not only growing; they continue to change form and structure in both more developed and less developed countries. During the second half of the 20th century, for example, the U.S. population became more suburban than “city.” Those greener urban areas intermingled domesticated, domiciliated, wild, and human animals in new ways. The suburbs that grew nationwide were still nuclear settlements, although at lower densities (which has proved the nemesis of mass transit today). Increasingly population sprawled beyond suburbs into dispersed “exurban” territory, and intruded into the natural nidi of zoonotic diseases (see Chapter 4).

One result has been the reemergence of rabies as a danger to human health in the Eastern United States. In the U.S. Southeast raccoons, supplemented by the occasional bat or fox, bring the disease into backyards and give it to domestic animals and people. In the North foxes are more involved, and in the Midwest skunks, but

the result is a new contact between a deadly disease of wild animals and people. A second result has been the emergence of tick-vectored diseases as major problems. Lyme disease and Rocky Mountain spotted fever (see Chapter 4) continue to spread and intensify as population decentralization pushes housing into wooded areas and agricultural fields. These habitats support resurgent deer, rabbit, and rodent populations in suburban “salad bars” called ornamental planting. It seems that urban park rats may be better for feeding and infecting ticks because they are larger and have higher viremia and a longer lifespan than the rural white-footed mice (Steere, 1994). In some Western states, murine typhus has been brought into human contact via cat fleas. Perhaps the biggest and least appreciated development is the spread of *Aedes albopictus*, the Asian tiger mosquito, which was brought by an Asian shipload of tires for recycling to Texas. In less than 10 years it spread to the Carolinas and Pennsylvania. Although, as discussed in Chapter 4, in Asia the domiciliated and people-loving *Ae. aegypti* displaced *Ae. albopictus* to become the most dangerous vector of dengue and other viruses, the U.S. suburban forms seem to be reversing the relationship. *Ae. albopictus* is thriving in the vegetation and wildlife of suburban sprawl. This species, capable of surviving winter conditions in a range of territory from India north to Japan and Korea, may extend and intensify the possible range of dengue in the United States if introduced again. Perhaps more importantly, however, the facility with which *Ae. Albopictus* feeds on opossums, raccoons, birds, dogs, and rodents as well as people creates a high risk of cross-species transmission of new arboviruses, even while diluting the intensity of any local focus of dengue.

In this new millennium, urban form and settlement are changing again. A movement known as the *new urbanism* aims to bring back clusters of social interaction, community stores, entertainment centers, schools, and sports facilities, as well as population densities sufficient for mass transportation. Such community activities would again encourage walking, it is argued. Street patterns need to connect like the old grids; suburban cul-de-sacs discourage walking, as does lack of sidewalks. People also need someplace to walk to. These developing nuclear settlements are surrounded with *green belts*, natural areas protected from development (or at least strip malls) but accessed by networks of *greenways* (Fabos, 2004). Central features of greenway plans are connectivity and accessibility (for micro-level circulation) (Jongman & Punnett, 2004). If the new urban forms and their greenways are to encourage convenient and safe walking, cycling, and exercise activities, then issues of social equity, neighborhood environmental justice, use, and exclusion become questions of public health (Powell, Martin, & Chowdhury, 2003; Saelens, Sallis, Black, & Chen, 2003; Saelens, Sallis, & Frank, 2003).

GLOBALIZATION OF MOVEMENTS

One of the fundamental activities of human behavior in disease ecology is to move people and elements of the habitat from place to place (see Chapter 2). In the midst of “globalization” of economic activities and communication of every kind, this

movement counts as a new thing if only because of the speed with which it is happening. This aspect of human agency is more important than ever. We have seen it in this chapter in the spread of Western diet and inactive lifestyle; in the introduction to the United States of the tiger mosquito in a shipload of tires; and in the relocation of disease agents of schistosomiasis, encephalitis, and malaria through impoundment of water and irrigation. To this should be added the transformation of the world's food supply.

Consider first, as one example of new development changes, what has to be called "the industrialization of food animal production." Like the urbanization of the human population, the herding of animals into enormous, densely packed populations reprises the ecological consequences of the first domestications. Cattle in feedlots at the giant slaughterhouses and meat-processing plants can be penned by the hundreds of thousands at a single facility. Chicken raising has been specialized into brooding; egg production; chicken (broiler) raising; and assembly-line slaughter, cleaning, parts production, packaging, and marketing. Fowl being raised for food are penned so densely that their beaks commonly must be cut. Following the business success with chickens and turkeys, in recent years swine production (according to the so-called "North Carolina model") has followed suit. Piglets are now nursed through bars on trapped sows, raised densely packed by tens of thousands into a facility, and fed by conveyor belts; their wastes are flushed into septic "lagoons" and eventually sprayed on the land. Food contamination from such industrialized production and marketing poses new challenges for inspection and for risks of pandemic. Food thus produced is shipped not only across the United States but internationally, as from Thailand to Japan.

Even if we put aside for present purposes ethical issues about the humane treatment of animals, serious public health consequences need consideration. The waste discharge and spills thus produced have attracted a lot of attention through putrefaction of streams, river, and offshore fish kills, and recently attacks by *Pfiesteria* (a genus of dinoflagellate protozoa) on fish and production of neurological toxins. Animals cannot be herded so densely without massive use of antibiotics, which are passing newly resistant strains of microbes on to people. The salmonella and other intestinal fauna are increasingly contaminating carcasses and parts in the mass, speeded-up, assembly-line processing. The combination of birds and pigs close together at such densities raises several red flags. The influenza virus is thought to "shift" when avian flu and other strains get combined in pigs, producing pandemics. The emergence of bird flu in chickens in Hong Kong in 1998 caused a panic and the slaughter of all the chickens; similar situations have continued to occur elsewhere. Avian encephalitis viruses, such as Japanese encephalitis (and probably Eastern equine encephalitis), can be amplified by the high viremia in pigs into major epidemics. Animals are not raised on food grown on the same farm any more, but by agribusiness production that spans continents. (Since this paragraph was written and submitted for review, a new influenza—"swine flu," or H1N1—has emerged, apparently from just such a hog facility as described above in Mexico. See Chapter 10.) When cattle in Britain began to get bovine spongiform encephalopa-

thy (BSE, also known as “mad cow disease”), it was traced through epidemiological methods to feed that included sheep brains and body parts (although cows are famous as vegetarians). This British feed had already affected cows in Italy and other European countries in trade. Although the string of protein molecules that seems to be involved (prions) defies any definition of life, it seems to reproduce the same brain destruction in chains of animals that consume it. Epidemiologists have made an analogy with *kuru*, previously found among cannibals in Papua New Guinea. The Western public’s greatest alarm, however, is about a rare human brain disease, Creutzfeldt–Jakob, which may be part of the same food chain as BSE. More common, if little understood, health challenges from animal feed include synthetic estrogens, growth hormones, and similar elements of high-tech mass production. In such ways, new international technologies of agricultural production, processing, and marketing are creating a new disease ecology.

The globalization of food production and marketing thus brings not only the benefits of abundant food available in any season, but risks of worldwide exposure to risky local conditions, whether these stem from industrial-model chicken and pig raising or the intensive use of pesticides on fruits and vegetables. It has been said that you may eat in the most sanitary gourmet restaurant of the 21st century, but your food was grown in the 16th. The dubious sanitary conditions and unsafe water sources of the developing world are daily encountered on the vegetables of the French table (from Mali) or on the grapes of the American table (from Chile). Food is distributed from centralized sources as never before. Twenty years ago there was concern that a sick milk cow could spread contamination through a dairy and its equipment to a supermarket outlet system, or that contaminated machinery in a meat-, candy-, vegetable-, or dairy-processing system could poison vending machines, school cafeterias, and kitchen counters nationally through the system of marketing. Now the threat is international. National food inspection systems themselves are largely antiquated bureaucracies. Contamination of beef with the intestinal bacteria of cattle, of water with runoff from feedlots, of apple juice with untested insecticides, and of Chinese children’s milk with an industrial chemical for plastics that tests “protein” are not random events. There is a way through the holistic approach of cultural ecology to find patterns, explanations, and understanding, but the work has barely been begun. It remains for our readers and other researchers of the future to develop the network of connections, theorize its relations, and speculate whether or how they will change with more organic ways and more local consumption.

We need briefly to return to the third destination of the mobility transition—international wage labor—and to ever-growing circulation. Globalization has many important effects, whether on wage structures and employment, information flow, belief systems, or scale of the pollution syndrome. One of the most profound for the ecology of disease has been the escalation of international migration and circulation. The numbers are “staggering,” as the cliché goes. There are over 250 million people legally working outside their home countries. There are over 250 million people a year who move between the tropics and developed latitudes for business, tourism, study abroad, and so on. Nigerians are trying to reach work in Germany, Tongans

in New Zealand, Samoans in California, and Pakistanis in the Persian Gulf. There are countries in which more than a fifth of the labor force consists of foreign workers (Switzerland, Côte d'Ivoire); in Saudi Arabia and other Gulf states, more than half of it does. There are other countries, like the Philippines, in which more than half the population over large areas can be overseas working—and actually making remittances the largest source of foreign exchange. Look at how the age structures of countries (Figure 5.2) fit together to connect countries with young adults needing work to countries without young adults to do needed work. People go from villages to the agricultural frontier and back; they go to cities and back; they circulate internationally for tourism, religious obligations, or education. Most of this they do at the speed of jet aircraft, arriving before symptoms can develop. Quarantine has become almost useless (as illustrated again by the emergence of the H1N1 influenza from Mexico in spring 2009). In addition to this, the United Nations High Commission on Refugees (2008) is responding to an unprecedented 16 million refugees (up from 5 million in 2000), plus 51 million people displaced internally by natural disaster or armed conflict. These 67 million people (astounding, unprecedented numbers) often live under desperate conditions. They are a virtual nation of the most vulnerable people struggling to live in health conditions that render our interconnected and mobile world susceptible to their pathologies of body and mind.

CONCLUSION: EMERGING DISEASES IN YOUR FUTURE

Human population growth and change, supported by our cultural technology, are the driving forces in changing the earth's climate, biological diversity, water availability and quality, agricultural production systems, food marketing, settlement patterns, urban forms, movements of disease exposure and introduction, creation of disease systems, and health status. This chapter has tried to explain the multiple transitions the world's population is undergoing and some of the consequences for human health that have resulted.

The end of the demographic transition may well be ushering in a stage of renewed parasitic and infectious disease. Infectious agents are newly discovered to be involved in everything from ulcers to hardening of the arteries. Instead of disappearing at the end of the demographic transition, diseases that were once almost eradicated are reemerging. Some of these are *iatrogenic* (from the Greek word for "cure"), or caused by medicine and doctors. The most notorious of these is the multiple-drug-resistant new tuberculosis, but resistant strains of staphylococcus (MRSA) threaten to make hospitals places to get sick. The widespread use of, and demand for, antibiotic treatments even when useless (as against viruses), and the general overprescription of drugs, have helped create drug resistance. Inadequate medical follow-up and patient compliance have also contributed to this resistance. New technologies of transferring human organs and tissues have created new means for agents to pass; future use of pig and other animal parts for humans (*xenografts*) may create *xeno-zoonoses*. Use of immunosuppressant drugs to prevent tissue rejection or fight the rapidly growing cells of cancer has allowed a new set of

infectious agents to become unsuppressed. Although these health hazards develop from the behavioral dimension of the triangle of human ecology, they have not been examined in this chapter.

The simplifications of ecosystems produced by agricultural conversion and, even more, by urban conversion when connected with the speed and multiple dimensions of population movement, have established new global epidemic propensities. Some disease agents are as yet purely zoonotic. However, as farmers in the frontiers contact the viruses of the rodents, bats, and birds they have disturbed, they bring them to home villages connected to cities where the farmers' siblings live amidst densities of human pathogens, market animals, and water/sanitation/housing conditions that couldn't be better designed to promote the jump by disease agents to the human species—and then into global connections. Children in the most developed places are again dying from intestinal fauna acquired in the meat and water of their countries.

Certainly the emergence of diseases can be approached as a political ecology. Governance, capitalism, and the struggles for power by groups at every scale are inherent. What those forces do, however, is shape the cultural ecology. Whatever the power and whoever holds it, the needs of the local disease vectors, the adaptability of enteric microorganisms, the need for more food production and for employment in cities, and the changes in age structure and distribution of susceptible individuals will create new disease systems. This chapter has presented a cultural ecology of how population change is resulting in changes in the creation and maintenance of disease systems and their emergent possibilities. Medical geographers need to become involved in developing prospective medical geographies of health and disease.

REFERENCES

- Audy, J. R. (1965). Types of human influence on natural foci of disease. In B. Rosicky & K. Heyberger (Eds.), *Theoretical questions of natural foci of disease: Proceedings of a symposium* (pp. 245–253). Prague: Czechoslovak Academy of Science.
- Azad, A. F., Radulovic, S., Higgins, J. A., Noden, B. H., & Troyer, J. M. (1997). Fleaborne rickettsioses: Ecologic considerations. *Emerging Infectious Diseases*, 3, 319–327.
- Bell, A. C., Ge, K., & Popkin, B. M. (2002). The road to obesity or the path to prevention?: Motorized transportation and obesity in China. *Obesity Research*, 10, 277–283.
- Boyden, S. V. (Ed.). (1970). *The impact of civilisation on the biology of man*. Toronto: University of Toronto Press.
- Brown, L. (2004). *Outgrowing the earth: The food security challenge in an age of falling water tables and rising temperatures*. Washington, DC: Earth Policy Institute.
- Centers for Disease Control and Prevention (CDC). (2008). State-specific prevalence of obesity among adults—United States, 2007. *Morbidity and Mortality Weekly Report*, 57(28), 765–768.
- Chua, K. B., Goh, K. J., Wong, K. T., Kamarulzaman, A., Tan, P. S., Ksiazek, T. G., et al. (1999). Fatal encephalitis due to Nipah virus among pig-farmers in Malaysia. *Lancet*, 354, 1257–1259.
- Dhanda, V., Das, P. K., Lal, R., Srinivasan, R., & Ramaiah, K. D. (1996). Spread of lymphatic filariasis, re-emergence of leishmaniasis and threat of babesiosis in India. *Indian Journal of Medical Research*, 103, 46–54.

- Dubos, R. (1965). *Man adapting*. New Haven, CT: Yale University Press.
- Erlanger, T. E., Keiser, J., Caldas de Castro, M., Bos, R., Singer, B. H., Tanner, M., et al. (2005). Effect of water resource development and management on lymphatic filariasis, and estimates of populations at risk. *American Journal of Tropical Medicine and Hygiene*, 73(3), 523–533.
- Fabos, J. G. (2004). Greenway planning in the United States: Its origins and recent case studies. *Landscape and Urban Planning*, 8(2–3), 321–342.
- Gould, W. T. S., & Prothero, R. M. (1975). Space and time in African population mobility. In L. A. Kosinski & R. M. Prothero (Eds.), *People on the move*. London: Methuen.
- Hagerstrand, T. (1957). *Migration and area in Sweden* (Lund Studies in Geography). Lund, Sweden: Department of Geography, Royal University of Lund.
- Hagerstrand, T. (1975). Space, time and human conditions. In L. Karlquist, L. Lundquist, & F. Snickars (Eds.), *Dynamic allocation of urban space* (pp. 3–12). Farnborough, UK: Saxon House.
- Hagerstrand, T. (1982). Diorama, path and project. *Tijdschrift voor Economische en Sociale Geografie*, 73(6), 323–339.
- Haggett, P. (1994). Geographical aspects of the emergence of infectious diseases. *Geografiska Annaler*, 76B, 91–104.
- Hughes, C. C., & Hunter, J. M. (1970). Disease and development in Africa. *Social Science and Medicine*, 3, 443–493.
- Hunter, J. M. (1993). Elephantiasis: A disease of development in northeast Ghana. *Social Science and Medicine*, 35, 627–649.
- Jacobsen, K. (1997). Refugees' environmental impact: The effect of patterns of settlement. *Journal of Refugee Studies*, 10, 19–36.
- Jongman, R., & Punnett, G. (Eds.). (2004). *Ecological networks and greenways: Concept, design, implementation*. Cambridge, UK: Cambridge University Press.
- Keiser, J., Caldas de Castro, M., Maltese, M. F., Bos, R., Tanner, M., Singer, B. H., et al. (2005). Effect of irrigation and large dams on the burden of malaria on a global and regional scale. *American Journal of Tropical Medicine and Hygiene*, 72(4), 392–406.
- Keiser, J., Utzinger, J., Caldas de Castro, M., Smither, T. A., Tanner, M., & Singer, B. H. (2004). Urbanization in sub-Saharan Africa and implication for malaria control. *American Journal of Tropical Medicine and Hygiene*, 71(Suppl. 2), 118–127.
- Kloos, H. (1985). Water resources development and schistosomiasis ecology in the Awash Valley, Ethiopia. *Social Science and Medicine*, 20, 609–625.
- Lewis, N. D., & Kieffer, E. (1994). The health of women: Beyond maternal and child health. In D. R. Phillips & Y. Verhasselt (Eds.), *Health and development* (pp. 122–137). New York: Routledge.
- Looi, L. M., & Chua, K. B. (2007). Lessons from the Nipah virus outbreak in Malaysia. *Malaysian Journal of Pathology*, 29(2), 63–67.
- Martens, P., & Hall, L. (2000). Malaria on the move: Human population movement and malaria transmission. *Emerging Infectious Diseases*, 6(2), 103–109.
- Matuschka, F. R. (1996). Risk of urban Lyme disease enhanced by the presence of rats. *Journal of Infectious Diseases*, 175, 1108–1111.
- May, R. M. (1994). Changing diseases in changing environments. In B. Cartledge (Ed.), *Health and the environment* (pp. 150–171). Oxford, UK: Oxford University Press.
- McGee, T., Ginsburg, N., & Koppel, B. (Eds.). (1991). *The extended metropolis: Settlement transition in Asia*. Honolulu: University of Hawaii Press.
- McKeown, T. (1988). *The origins of human disease*. Oxford, UK: Blackwell.
- Meade, M. S. (1976). Land development and human health in west Malaysia. *Annals of the Association of American Geographers*, 66, 428–439.
- Meade, M. S. (1977). Medical geography as human ecology: The dimension of population movement. *Geographic Review*, 67, 379–393.
- Meade, M. S. (1978). Community health and changing hazards in a voluntary agricultural resettlement. *Social Science and Medicine*, 12, 95–102.
- Meade, M. S. (1980). Potential years of life lost in counties of Southeast Asia. *Social Science and Medicine*, 14D, 277–281.

- Meslin, F. X. (1997). Global aspects of emerging and potential zoonoses: A WHO perspective. *Emerging Infectious Diseases*, 3, 223–228.
- Mott, K. E., Desjeux, P., Moncayo, A., Ranque, P., & de Raadt, P. (1990). Parasitic diseases and urban development. *Bulletin of the World Health Organization*, 68, 691–698.
- National Center for Health Statistics. (2008, July 14). Births, marriages, divorces, and deaths: Provisional data for 2007. *National Vital Statistics Reports*, 56(21).
- Odum, E. P. (1978). *Ecology: The link between the natural and the social sciences* (2nd ed.). New York: Holt, Rinehart & Winston.
- Omran, A. R. (1977). Epidemiologic transition in the U.S.: The health factor in population change. *Population Bulletin*, 32, 2–41.
- Polo, P. E. (2009). *Nature–culture interactions: Understanding the prevalence of malaria in the northern Ecuadorian Amazon*. Unpublished master's thesis, University of North Carolina at Chapel Hill.
- Popkin, B. M. (2002a). An overview on the nutrition transition and its health impacts. *Public Health Nutrition*, 5(1A), 93–103.
- Popkin, B. M. (2002b). The shift in stages of the nutrition transition in the developing world differs from past experiences. *Public Health Nutrition*, 5, 205–214.
- Popkin, B. M. (2003). The nutrition transition in the developing world. *Development Policy Review*, 21(5–6), 581–597.
- Popkin, B. M., & Gordon-Larsen, P. (2004). The Nutrition Transition: Worldwide obesity dynamics and their determinants. *International Journal of Obesity*, 28(Suppl. 3), S2–S9.
- Population Reference Bureau. (2008). *Demographic data and estimates for the countries and regions of the world*. Washington, DC: Author.
- Powell, K. E., Martin, L. M., & Chowdhury, P. P. (2003). Places to walk: Convenience and regular physical activity. *American Journal of Public Health*, 93(9), 1519–1521.
- Prothero, R. M. (1961). Population movements and problems of malaria eradication in Africa. *Bulletin of the World Health Organization*, 24, 405–425.
- Prothero, R. M. (1965). *Migrants and malaria*. London: Longmans.
- Prothero, R. M. (1994). Forced movements of population and health hazards in tropical Africa. *International Journal of Epidemiology*, 23, 657–664.
- Saelens, B. E., Sallis, J. F., Black, J. B., & Chen, D. (2003). Neighborhood-based differences in physical activity: An environment scale evaluation. *American Journal of Public Health*, 93(9), 1552–1558.
- Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine*, 25(2), 80–91.
- Sawyer, D. R. (1988). *Frontier malaria in the Amazon region of Brazil: Types of malaria situations and some implications for control*. Brasília: PAHO/WHO/TDR.
- Schaerstrom, A. (1996). *Pathogenic paths?: A time geographical approach in medical geography*. Lund, Sweden: Lund University Press.
- Schriefer, M. E. (1994). Murine typhus: Updated roles of multiple urban components and a second typhus-like rickettsia. *Journal of Medical Entomology*, 31, 681–685.
- Service, M. W. (Ed.). (1989). *Demography and vector-borne diseases*. Boca Raton, FL: CRC Press.
- Sharma, S. N., Subbarao, S. K., Choudhury, D. S., & Pandey, K. C. (1993). Role of *Anopheles culicifacies* and *Anopheles stephensi* in malaria transmission in urban Delhi. *Indian Journal of Malariology*, 30, 155–168.
- Singer, B., & Caldas de Castro, M. (2006). Enhancement and suppression of malaria in the Amazon. *American Journal of Tropical Medicine and Hygiene*, 74(1), 1–2.
- Soares, R. (2007). On the determinants of mortality reductions in the developing world. *Population and Development Report*, 33(2), 247–287.
- Steere, A. C. (1994). Lyme disease: A growing threat to urban populations. *Proceedings of the National Academy of Sciences USA*, 91, 2378–2383.

- Surtees, G. (1970). Effects of irrigation on mosquito populations and mosquito-borne disease in man, with particular reference to rice field extension. *International Journal of Environmental Studies*, 1, 35–42.
- Surtees, G. (1971). Urbanization and the epidemiology of mosquito-borne disease. *Abstracts on Hygiene*, 46, 121–134.
- United Nations. (1950). *Demographic yearbook 1949–50*. New York: Author.
- United Nations High Commission on Refugees. (2008). *2007 global trends: Statistics*. New York: United Nations.
- United Nations Human Settlements Programme (UNHSP). (2003). *Global report on human settlements 2003: The challenge of slums*. Sterling, VA: UN-HABITAT, Earthscan Publications.
- United Nations Population Division (UNPD). (2001). *World population prospects 2000*. New York: United Nations.
- United Nations Population Division (UNPD). (2002). *World urbanization prospects: The 2001 revision*. New York: United Nations.
- U.S. Census Bureau. (2008). *International database*. Retrieved from www.census.gov/ipc/www/idb
- Vandegrift, D., & Yoked, T. (2004). Obesity rates, income, and suburban sprawl: An analysis of U.S. states. *Health and Place*, 10, 221–229.
- Vittor, A. Y., Gilman, R. H., Tielsch, J., Glass, G., Shields, T., Sánchez Lozano, W., et al. (2006). The effect of deforestation on the human-biting rate of *Anopheles darlingi*, the primary vector of falciparum malaria in the Peruvian Amazon. *American Journal of Tropical Medicine and Hygiene*, 74, 3–11.
- Warnes, A. M. (1999). UK and western European late-age mortality: Trends in cause specific death rates, 1960–1990. *Health and Place*, 5, 111–118.
- Wolfe, B. L., & Behrman, J. R. (1983). Is income overrated in determining adequate nutrition? *Economic Development and Cultural Change*, 31, 525–549.
- World Health Organization (WHO). (2005). *World Health Report 2005: Making every Mother and Child Count*. Geneva: Author.
- Zelinsky, W. (1971). The hypothesis of the mobility transition. *Geographic Review*, 61, 219–249.
- Ziegler, R. G., Hoover, R. N., Pike, M. C., Hildesheim, A., Nomura, A. M. Y., West, D. W., et al. (1993). Migration patterns and breast cancer risk in Asian-American women. *Journal of the National Cancer Institute*, 85, 1819–1827.

FURTHER READING

- Akhtar, R. (Ed.). (1987). *Health and disease in tropical Africa*. London: Harwood Academic.
- Desowitz, R. S. (1981). *New Guinea tapeworms and Jewish grandmothers*. New York: Norton.
- Drexler, M. (2002). *Secret agents: The menace of emerging infections*. Washington, DC: National Academy Press.
- Farvar, M. T., & Milton, J. (Eds.). (1972). *The careless technology: Ecology and international development*. New York: Stacey.
- Gade, D. W. (1979). Inca and colonial settlement, coca cultivation and endemic disease in the tropical forest. *Journal of Historical Geography*, 5, 263–279.
- Garrett, L. (1994). *The coming plague: Newly emerging diseases in a world out of balance*. New York: Farrar, Straus & Giroux.
- Gopalan, C. (1992). *Nutrition in developmental transition in South-East Asia*. New Delhi: World Health Organization.
- Harpham, T. (1994). Urbanization and mental health in developing countries: A research role for social scientists, public health professionals, and social psychiatrists. *Social Science and Medicine*, 39, 233–245.
- Howe, G. M. (1972). *Man, environment and disease in Britain*. New York: Barnes & Noble.
- Howe, G. M. (Ed.). (1977). *A world geography of human diseases*. New York: Academic Press.

- Hubbert, W. T., McCulloch, W. F., & Schnurrenberger, P. R. (Eds.). (1975). *Diseases transmitted from animals to man* (6th ed.). Springfield, IL: Thomas.
- Karlen, A. (1995). *Man and microbes: Diseases and plagues in history and modern times*. New York: Putnam.
- McMichael, T. (2001). *Human frontiers, environments and disease*. Cambridge, UK: Cambridge University Press.
- Meade, M. S. (1992). Implications of changing demographic structures for rural health services. In W. M. Gesler & T. C. Ricketts (Eds.), *Health in rural North America* (pp. 69–85). New Brunswick, NJ: Rutgers University Press.
- Mena, C. F., Barbieri, A., Walsh, S. J., Erlien, C. M., Bilsborrow, R. E., & Lu, F. (2006). Pressure on the Cuyabeno Wildlife Reserve: Development and land use/cover change in the northern Ecuadorian Amazon. *World Development*, 34(10), 1831–1849.
- Mena, C. F., Bilsborrow, R. E., & McClain, M. E. (2006). Socioeconomic drivers of deforestation in the Northern Ecuadorian Amazon. *Environmental Management*, 37(6), 802–815.
- Moeller, D. W. (1997). *Environmental health* (rev. ed.). Cambridge, MA: Harvard University Press.
- Murray, C. J. L., & Lopez, A. D. (Eds.). (1996). *1990 global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020* (Global Burden of Disease and Injury Series, Vol. 1). Cambridge, MA: Harvard University Press.
- Mutatkarak, R. K. (1995). Public health problems of urbanization. *Social Science and Medicine*, 41, 977–981.
- Pan American Health Organization. (1996). New emerging and re-emerging infectious diseases. *Bulletin of the Pan American Health Organization*, 30, 176–181.
- Phillips, D. R., & Verhasselt, Y. (1994). *Health and development*. London: Routledge.
- Reardon, T., & Berdegue, J. A. (2002). The rapid rise of supermarkets in Latin America: Challenges and opportunities for development. *Development Policy Review*, 20(4), 371–388.
- Rodriguez, A. D., Rodriguez, M. H., Hernandez, J. E., Dister, S. W., Beck, L. A., Rejmankora, E., et al. (1996). Landscape surrounding human settlements and *Anopheles albimanus* (Diptera: Culicidae) abundance in southern Chiapas, Mexico. *Journal of Medical Entomology*, 33, 39–48.
- Sawyer, D. R. (1992). *Malaria and the environment*. Brasília: Instituto SPN.
- Sawyer, D. R., & Sawyer, D. O. (1992). The malaria transition and the role of social science research. In L. C. Chen (Ed.), *Advancing health in developing countries: The role of social research* (pp. 105–122). Westport, CT: Auburn House.
- Sharma, V. P. (1996). Re-emergence of malaria in India. *Indian Journal of Medical Research*, 103, 26–45.
- Singhanetra, R. A. (1993). Malaria and mobility in Thailand. *Social Science and Medicine*, 37, 1147–1154.
- Stanley, N. F., & Alpers, M. P. (Eds.). (1975). *Man-made lakes and human health*. New York: Academic Press.
- Takemoto, T., Suzuki, T., Kashiwazaki, H., Mori, S., Hirata, F., Taja, O., et al. (1981). The human impact of colonization and parasite infestation in subtropical lowlands of Bolivia. *Social Science and Medicine*, 15D, 133–139.
- Thompson, K. (1969a). Insalubrious California: Perception and reality. *Annals of the Association of American Geographers*, 59, 50–64.
- Thompson, K. (1969b). Irrigation as a menace to health in California: A nineteenth century view. *Geographical Review*, 59, 195–214.
- Vachon, M. (1993). Onchocerciasis in Chiapas, Mexico. *Geographical Review*, 83, 141–149.
- Weil, C. (1981). Health problems associated with agricultural colonization in Latin America. *Social Science and Medicine*, 15D, 449–461.
- Williams, B. (1990). Assessing the health impact of urbanization. *World Health Statistics Quarterly*, 43, 145–152.

- Wills, C. (1996). *Yellow fever, black goddess: The coevolution of people and plagues*. Reading, MA: Addison-Wesley.
- World Health Organization (WHO). (1976). *Water resources development and health: A selected bibliography* (Document No. MPD/76.6). Geneva: Author.
- World Health Organization (WHO). (1993). *The urban health crisis: Strategies for health for all in the face of rapid urbanization: Report of the technical discussions*. Geneva: Author.

VIGNETTE 5.1. Age Standardization

Few things can be so misleading as comparisons between populations with different age structures. Because the very old and the very young have particularly high mortality rates, differences in the proportion of the population in these categories alter the number of deaths occurring in the population, and so alter the crude death rates. Differences in the possibility of a disease's occurring are not accurately reflected, as discussed previously.

Consider the crude death rates in three countries: Country *A* has 4.3 deaths per 1,000; *B* has 6.0 per 1,000; and *C* has 4.3 per 1,000. It would appear that the population of country *B* is far worse off. Vignette 5.1, Table 1, also shows the age structure and the age-specific death rates, however, and it can be seen that country *B* has the same mortality experience as country *A*. The difference is the proportion of its population that is at risk of the higher rates because of age structure. In contrast, country *C* has the same crude death rate as country *A*, and an identical population structure, but its health conditions are characterized by higher mortality rates for the elderly and lower ones for infants.

One can compare mortality among these three countries by using the age-specific death rates. If individual years of age were used instead of 10-year groupings, however, or if more countries were involved, the tables would soon become too complex for comprehension. A single, summary figure is needed—one that is similar to the crude death rate, but takes into account differences in age structure.

This summary, comparative figure is the *standardized death rate*. To calculate it, a standard population is needed. The age-specific death rates for all the populations are then applied to the age structure of the standard population. This answers the question "How many deaths would there be if the study populations, given their separate mortality experiences, had the same age structure?" The total number of expected deaths is then divided by the total standard population to create a summary death rate that is valid for comparison.

Sometimes it is difficult to identify an appropriate standard population. Usually if one is comparing subdivisions of a whole, such as states or provinces of a nation or counties of a state, the total population by race or sex (as appropriate) is used. A world standard and separate regional standards are used by the WHO. Sometimes geographers engaged in fieldwork find no appropriate standard population available, and they must resort to adding the village or other studied populations together and using the

VIGNETTE 5.1, TABLE 1. Age-Specific Death Rates for Three Hypothetical Countries

Age	Population	Deaths (per 1,000)	Age-specific death rate
<u>Country A</u>			
0-1	300	15	50.0
1-9	2,250	11	4.8
10-19	1,700	5	2.9
20-29	1,400	2	1.4
30-39	1,350	1	0.7
40-49	1,200	1	0.8
50-59	1,000	2	2.0
60+	800	6	7.5
Total	10,000	43	CDR = 4.3/1,000
<u>Country B</u>			
0-1	600	30	50.0
1-9	3,500	17	4.8
10-19	1,950	6	3.0
20-29	1,400	2	1.4
30-39	1,050	1	0.9
40-49	700	1	1.4
50-59	500	1	2.0
60+	300	2	6.7
Total	1,000	60	CDR = 6.0/1,000
<u>Country C</u>			
0-1	300	4	13.3
1-9	2,250	2	0.9
10-19	1,700	1	0.6
20-29	1,400	2	1.4
30-39	1,350	3	2.2
40-49	1,200	5	4.2
50-59	1,000	10	10.0
60+	800	16	20.0
Total	10,000	43	CDR = 4.3/1,000

Note. CDR, crude death rate.

total as a standard. The choice of a standard population needs to be carefully considered with respect to the research question.

Vignette 5.1, Table 2, illustrates the calculation of standard death rates. In this example the three populations were added together to make the standard. If *B* had been standardized on *C*'s population, or some other combination had been used, the result would have been different. The general formula for direct age adjustment is

$$\frac{\sum M_x x P_x^s}{\sum P_x^s}$$

VIGNETTE 5.1, TABLE 2. Direct Age Standardization for the Death Rates of Table 1

Age	Standard population (A + B + C)	Population A		Population C	
		Age-specific death rate	Expected deaths	Age-specific death rate	Expected deaths
0–1	1,200	.0500	60.0	.0133	16.0
1–9	8,000	.0048	38.4	.0009	7.1
10–19	5,350	.0029	15.5	.0006	3.2
20–29	4,200	.0014	5.9	.0014	5.9
30–39	3,750	.0007	2.6	.0022	8.2
40–49	3,100	.0008	2.5	.0042	13.0
50–59	2,500	.0020	5.0	.0100	25.0
60+	1,900	.0075	14.3	.0200	38.0
Total	30,000		144.0		116.0

Note. Total population crude death rate per 1,000 = $146/30,000 = 4.87$.

Standard death rate per 1,000: population A = $144/300,000 = 4.8$.

population B = same age-specific death rate = 4.8.

population C = $116/30,000 = 3.8$.

Standard rate ratio:

population A = $4.8/4.87 = 1$.

population B = $4.8/4.87 = 1$.

population C = $3.87/4.87 = 0.8$.

where M is the mortality rate, P is the population at risk, x is the age category, and s is the standard population. For example, Mx is the mortality rate of population age x of the study population, and Px^s is the population of age x in the standard population.

To calculate the directly adjusted death rate, one needs to know the age-specific death rates of the study population. It sometimes happens that one knows the total number of deaths, but not the ages of the deceased, and thus the data cannot be directly standardized. In that case, if age-specific deaths are available for a standard population, the death rates may be indirectly standardized. The direct process is then reversed. The age-specific death rates of the standard population are applied to the study population's age structure to answer the question "How many deaths would there be if the study population, given its age distribution, had the mortality experience of the standard?" The observed deaths are then compared with the expected deaths to create the *standardized mortality ratio* (SMR). If this ratio of mortality experience is multiplied by the crude death rate of the standard population, a standardized death rate results. The indirectly standardized death rate is commonly used by geographers because age information for death is often not available at the micro scale. The SMR, which is calculated by the indirect method, is also convenient for comparing the spatial patterns of common and rare diseases in the same place as it lends itself to mapping. The direct method of age standardization is statistically more valid, however, and should be used whenever practicable.

In fact, despite the widespread usage of the SMR both in geography for mapping and in epidemiology, biostatisticians find indirect standardization so clearly invalid that the equation is not presented here. The problem is that the SMR is actually a weighted average in which the weights are derived from the study population, and so the standard is always the experience structure of the study population. The idea that

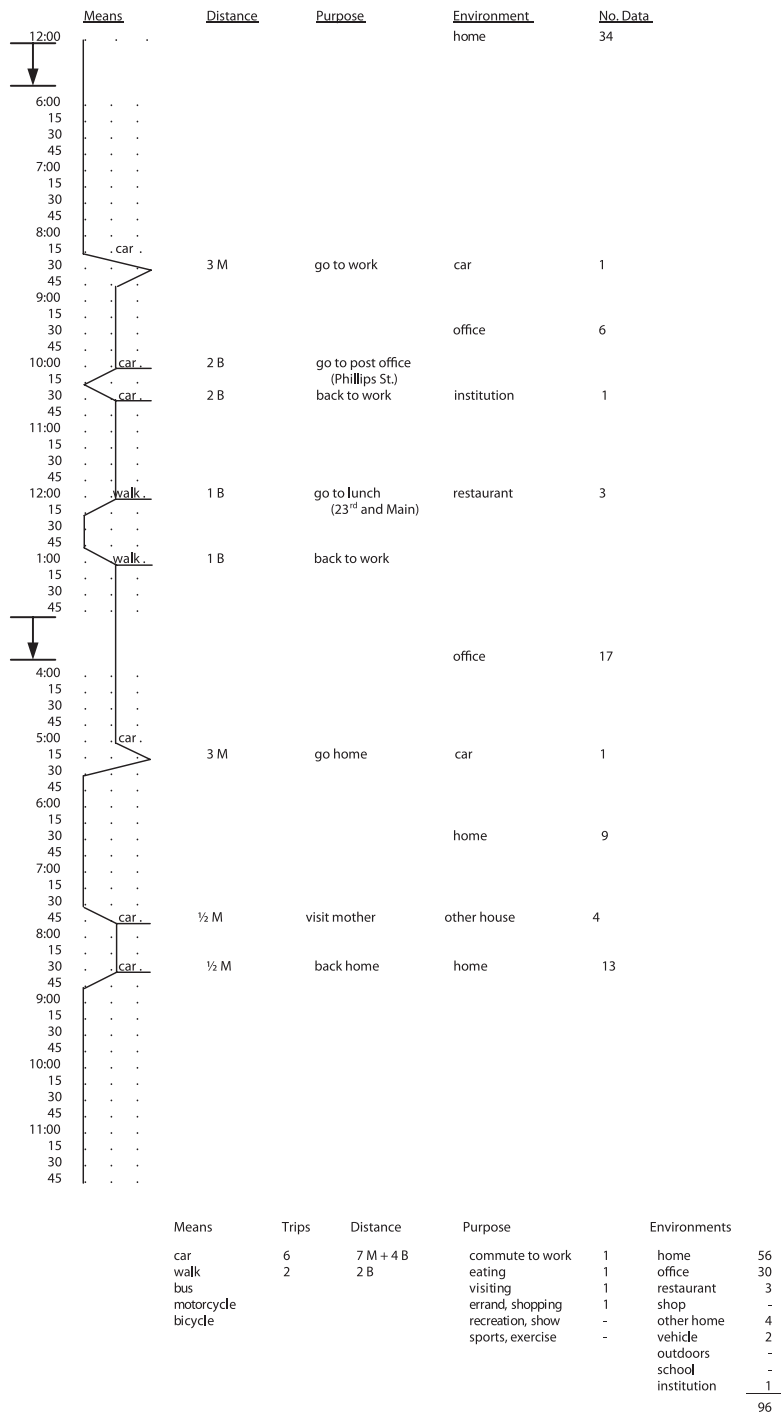
indirect standardization is based on the nonstudy population as the common standard is a misconception that results in a common methodological error: mapping SMRs for comparison, although they do not share a standard population. Whenever age-specific rates are available, which is almost always the case today, a mappable ratio can be easily developed: the standard rate ratio. As the table shows, the standard rate ratio results from dividing the standard death rate (directly calculated) of the study population by the crude death rate of the standard population, in the same way that the SMR was calculated from indirect standardization.

VIGNETTE 5.2. Microspatial Exposure Analysis

Detailed mobility information is essential for study of exposure to specific environments. For example, knowledge of water contact behavior is critical to understanding the incidence of schistosomiasis. Knowing about exposure to orchard mosquitoes in the evening or stream mosquitoes in the morning helps to identify hazards from specific vectored diseases. Often the environment must be carefully described. The home at night, for example, affects health differently depending on whether the house is screened, the beds are netted, the room is smoke-filled, and so on. At a public health level, the appropriate population for education, intervention, or monitoring can often be targeted through knowledge of critical environments and population mobility. Knowledge of microspatial mobility may help identify risk factors for diseases of unknown etiology.

Vignette 5.2, Figure 1, illustrates one means of collecting such microspatial mobility data. Respondents are asked to recall their mobility over a 24-hour period, yesterday. "Yesterday" should not be a holiday or a special day of the week. The day is marked off in equal time segments by three columns of dots (15-minute segments are used in the figure). The interviewer talks the respondent through the day. A line is drawn down the first column of dots until the first trip, to the yard or to another place, depending on the scale being analyzed. The shift of the vertical line from column one to column two and back to column one represents a change in environment. Notice that the line passes on the diagonal to the dot in the next column. A horizontal line would result in counting the same time segment twice. A line to the third column marks the means of travel. When the trip takes 15 minutes (one time segment) or more, the third-column dots represent environmental exposure. The distance of the trip can be noted in appropriate measures (in the example, miles and blocks), and the location can be specified so that map coordinates can be assigned later. Coding requires that one specify environments appropriately for the research purpose. The number of dots that the vertical line covers for each environment provides an easy summary. The number of trips, the distances traveled, and the means of travel by purpose are also easily summarized from the form. The scale and environments involved in another interview might specify particular fields, water holes, marketplaces, rooms within a factory, or districts within a city.

Aggregating the responses to mobility interviews permits assessment of differences in exposure to environments and of the extent and means of mobility. Vignette



VIGNETTE 5.2, FIGURE 1. Mobility recall form. Diagonal lines between the first two columns of dots show change of habitat, while the third column is used to indicate the means of transportation, which may itself become a habitat exposure.

VIGNETTE 5.2, TABLE 1. Risk of Exposure

Environment	Dots of exposure of population ^a		Control population relative risk
	Sick	Healthy	
Total home	1,613	1,498	1.08
Home 7 A.M. to 7 P.M.	260	196	1.3
Office	86	490	.18
Factory	518	58	8.9
Restaurant	28	86	.33
Shop	29	55	.53
Other home	86	60	1.43
Vehicles	115	58	1.98
Outdoors	29	144	.20
School	346	403	.86
Institution	5	18	.28
Church	25	10	2.5
Total dots of exposure	2,880	2,880	

^aPopulation, 30 sick and 30 healthy individuals.

5.2, Table 1, represents the results obtained from a hypothetical study population with an illness and a healthy control population, matched by age, sex, ethnicity, income level, and other appropriate criteria. Because there are 96 of the 15-minute segments (dots) in a day, each population of 30 has a total of 2,880 dots of exposure to some environment. The sick population has a high relative risk for exposure to a factory environment and to vehicles during the commute to work. It is much less exposed to the outdoors (perhaps for exercise or recreation). There are also social differences between the two populations, as reflected in more time spent by the sick population in visiting friends, going to church, and staying home, and less time spent in restaurants or institutions such as banks or post offices. The significance of differences between the population can be assessed statistically by using a *t*-test.

CHAPTER 6

Climate and Weather

Influences on Health

Climate, weather, and biometeorological conditions affect the health status of people both directly and indirectly. In terms of the triangle of population–habitat–behavior interactions, this chapter is especially concerned with influences of the natural environment. All living things on earth have evolved to cope with either the presence or absence of light. They have to cope with cold and/or heat. They have to get water. None of these essentials are uniform over the earth. Because of the tilt of the earth's axis, and the time period of the rotation on that axis and of that axis around the sun, predictable patterns of climate have been created; these are associated with predictable patterns of vegetation, or *biomes*. Realms of evolution and biomes of vegetation are each partially characterized by the species and life cycles of arthropods, amphibians, reptiles, birds, primates, and other mammals living there, and of their fungi, bacteria, viruses, and ectoparasites. Culture, of course, has a lot to do with how this natural environment affects people, how they are protected, and how people alter that environment. Climatology is one of the oldest studies in geography, but too often the physical geographers who write about correlations of heat waves and deaths, or flood and mosquito hazards, completely ignore cultural buffers, adjustments, and active responses.

Climate is the large term referring to weather conditions averaged and prevailing over long periods, many decades or centuries, of time. Such broad descriptive average parameters as how cold the winters or how hot the summers get, average last or first days of frost, onset of the monsoon rains, general humidity levels, and direction and strength of winds—these parameters describe climate. Climate determines what crops can (not will) be grown where, and how many blood meals mosquitoes usually take before they die and disease transmission is broken (or not). The daily and weekly conditions of temperature, humidity, precipitation, wind, and

cloud cover that are experienced constitute the *weather*. The fluctuations that always occur in heat, rain, cold, drought, snow cover, wind strength, and storms from week to week and year to year are weather. Fluctuations in the weather—extended periods of severe droughts, several years of mild or severe winters—have always occurred. Cycles of hurricane activity—decades of more or fewer storms occurring and hitting the mainland United States—have been known for more than a century. One of the great questions of our times has been when extremes of weather and concurrence of cycles become enough to change the average (i.e., to change climate). Such a climate change, to a warming earth, is upon us. It will affect temperatures, frost dates, patterns of precipitation (as well as its amounts and forms), wind strength, humidity, surface ultraviolet light, floods, and droughts. It will thereby affect crops; insect production; the synchronization in timing of birds, bees, and mosquitoes; water availability for irrigation, sanitation, and hygiene; and the potability of drinking water. The importance of these for human health is discussed in the final major section of this chapter.

So *climate* is composed of long-term averages over large areas; *weather* consists of daily and weekly local experience; and *biometeorology* consists of the variation and change in the physical and chemical characteristics of the atmosphere that affect variation and change in the physicochemical systems of living organisms. The human mammal is sensitive to a far wider range of atmospheric characteristics (Audy's physical *insults*, or *stimuli*) than is commonly realized. These include not only temperature, humidity, air movement, atmospheric pressure, solar radiation, sound, and gaseous pollution, but also infrasound, magnetism, and electrical charge. In addition, a host of indirect influences derive from the biometeorology of each location on earth.

Elements of the built environment, such as microwaves or buildings' central heating, can also be important influences on the physiological status of the population—that is, on the chemical–electrical systems of people's bodies. Physiological status is influenced by the time of year and the time of day; it responds to the light, temperature, and altitude of a place. A person's susceptibility to disease, toxins, and pharmaceutical drugs can be altered by biometeorological changes. Although there is a tendency in medicine to treat all people as a homogeneous population, in fact human physiological characteristics differ from place to place as well as from individual to individual. The temperature at which water boils varies with altitude around the earth, so it shouldn't be surprising that the potency of a drug or infectious dose does too. However, some of the most talked-about influences of the weather (such as the short tempers and increased violence associated with a long, hot summer) may be more constructs of social rhythms in employment and school vacation, sports, and what might be called the political economy of air conditioning than they are biometeorological effects of the weather itself.

This chapter briefly examines how humans receive and biologically respond to such stimuli, and surveys the most direct effects on human health. Biological rhythms and acclimatization to temperature and altitude are examined. The physiological basis of climatic influence is described in some detail because sound geographic hypotheses and choice of relevant variables depend on an understanding

of the disease process. The influence of heat waves, air masses, and their passage is described. Seasonality in the basic vital events of birth and death is analyzed. Finally, attention is turned to the impact of changing climate upon all these, and upon the water-related and vectored infectious diseases. With global warming, disaster preparation and response will undoubtedly become important applications of medical geography.

DIRECT BIOMETEOROLOGICAL INFLUENCES

The Radiation Spectrum

Figure 6.1 illustrates the spectrum of electromagnetic radiation. Electromagnetic radiation is propagated through space in the form of packets of energy called *photons*, which travel at the speed of light. Each photon has a frequency, wavelength, and proportional energy. Short wavelengths of radiation are known as *ionizing* radiation because they can detach electrons and damage atomic structure. These photons of the higher-energy ranges—cosmic, gamma, and X-rays—can penetrate into genes and disrupt the DNA sequences of our design. X-rays are principally a human-made health hazard; the atmosphere mostly shields us from their natural sources.

Ultraviolet light is best known for causing sunburn. Because the damage it does depends on the intensity and duration of exposure, it creates some important occupational hazards. Welding, for example, produces such intense ultraviolet radiation that an instant's unshielded exposure can cause severe inflammation of the eye's membrane. Only a small band within the range of ultraviolet radiation can penetrate the atmosphere to any extent because it is blocked by a stratospheric layer of ozone. This layer in the past was called the "germicidal band," because air-borne bacteria and fungus spores are killed in this zone, and the atmosphere is cleansed. This ozone layer was partially destroyed—unintentionally, of course—by industrial production of halogenated chemicals such as the chlorofluorocarbons (CFCs, used in refrigeration and spray-can propellants). CFCs, inert and seemingly harmless at normal surface temperatures, react with the ozone when catalyzed by the extreme cold. International treaties now banning CFCs and related chemicals, although not

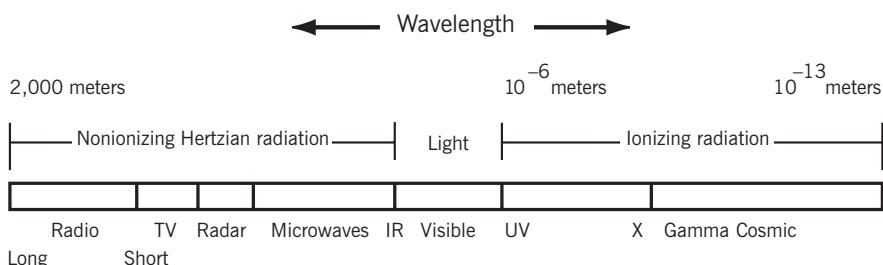


FIGURE 6.1. The spectrum of electromagnetic transmission. Radiation, from cosmic to radio, is a continuum of wavelength. IR, infrared; UV, ultraviolet.

universally subscribed or enforced, seem to have stopped enlarging the hole in the ozone layer over the Antarctic. We now know, however, that other chemicals are involved in this process, and that when the greenhouse gases trap heat in the lower atmosphere, they will accelerate the destruction of ozone in the then colder stratosphere. The weakened shield of the ozone layer against the harmful effects of ultraviolet radiation for human and other life on land is expected during the first years of this 21st century to let the highest amount of ground-level radiation through in the midlatitudes.

Humans evolved in the presence of ultraviolet radiation and are dependent upon it. It activates a human enzyme to produce vitamin D. The dependence of the human body on ultraviolet radiation to catalyze the body's synthesis of vitamin D, and the health dilemma this has posed for human evolution as people spread over the globe, are described in Chapter 2.

Visible light is biologically very powerful. It stimulates photoreceptive tissue, such as the retinas in human eyes. It also changes the electrical charge of protoplasm (the physical base of all living activities) and the viscosity, permeability, and colloidal behavior of proteins. It thereby causes living things to turn toward light, and it energizes photosynthesis. Light seems to be the most important factor influencing biological rhythms. Only these photons in the intermediate range of electromagnetic radiation can be detected by human senses.

Infrared and longer-wave radiations are important for heating the earth's surface and its cold-blooded plants and animals, either directly or through re-radiation by the earth. The health hazards of these long-wave radiations result from their function as penetrating heat sources. Infrared radiation used to be a cause of blindness in such occupations as the manufacture of glass, but careful shielding and new industrial processes have removed most such occupational hazards. Today microwaves are of some concern as the machinery that produces them proliferates. The waves can penetrate the body at considerable distance and cause internal heating that the body's sensors do not detect. The eyes are at greatest risk because the lens of the eye has no blood circulation to remove the heat. At the end of the spectrum are the longer radio waves. They are all around us, but are not known to have ever done anyone harm, not even people intensely exposed by occupation.

All of these types of electromagnetic radiation vary over the surface of the earth. The earth's magnetism affects the path and concentration of much radiation. The tilt of the earth results in seasonal and latitudinal differences described in this radiation. High elevations that project through thousands of feet of atmosphere—and the people who live at those elevations—experience different amounts and compositions of radiation. Human-made sources of several kinds of radiation are becoming important, and they have a very irregular distribution. Consequences of radiation have been much studied in terms of cell chemistry, microbiology, and physiology, but the spatial distribution of hazards and health effects has received little attention, except for risk of skin cancer.

One of the purposes of this text, as mentioned in the Preface, is to serve as a basic reference. The following section falls into that category. It offers the basic biological foundation for, and medical associations with, the effects of weather and

climate. Some readers will find it a good grounding in background with which they are unfamiliar. Others will want to skip it entirely and proceed to the influences of the weather.

Biological Rhythms

The presence or absence of light is one of the oldest and most universal selective pressures to which all living things have had to adapt. An important part of that adaptation has been the development of daily and seasonal rhythms of many body processes. It is clearly beneficial to an organism to be able to anticipate when cold, night, rain, high tide, or spring flowers and grass are going to come, so that the slow physiological processes of growing hair, losing leaves, storing food, or coming into rut can be initiated in good time. The precise synchronicity of breeding swarms of many sea creatures and insects, the beaching and egg laying of turtles and crabs that travel great distances to appear simultaneously at a specific place, and the migration of birds and bats are phenomena that have long fascinated scientists and testified to the existence of *biological clocks*. The equivalent body rhythms of human beings have been recognized much more recently. Some social scientists have resisted recognizing any environmental influences on people, and some religious belief systems resist any suggestions that humans are animals living in the biosphere.

The existence of endogenous clocks is now well established for a wide variety of animals, including humans. Even in constant temperature and total darkness, some biological processes continue to oscillate rhythmically. The rhythms with a span of about 24 hours are known as *circadian*. In humans they include sleep and wakefulness; body temperature; cognitive performance; serum hormone levels; urinary cycles and excretion of ketosteroids, chloride, sodium, and urea; and ionization of blood calcium and phosphate, which affects hormone regulation. It is clear that the physiological state of individuals is constantly changing. A tenfold difference in the susceptibility of mice to bacterial toxins at different times of night has been demonstrated, and it is likely that human circadian rhythms also influence susceptibility to infection and toxins.

The study of circadian and other rhythms can help us understand the ways in which environment directly affects and alters the status of a population. Reception and interpretation of environmental stimuli influence virtually all the body's major organs and endocrine systems (Figure 6.2). There have been major advances in understanding the nature of the body's clock (Barinaga, 1999; Somers, Devlin, & Kay, 1998; Thresher et al., 1998). A recently discovered cluster of nerve cells in the brain, the *suprachiasmatic nucleus* (SCN), is now believed to be the organ responsible for synchronizing body rhythms with environmental light. Pigments called *cryptochromes* have been found in the eye, skin, and part of the brain. They are believed to drive the body's clock. The two forms of cryptochromes located in a part of the retina absorb blue light and transfer the light signal to the SCN—a process for which the individual does not have to be awake or conscious. In individual neurons there, two proteins alternately build up and then turn on and off each other's gene for production, thus forming an accurate internal clock in each of tens of thousands of cells

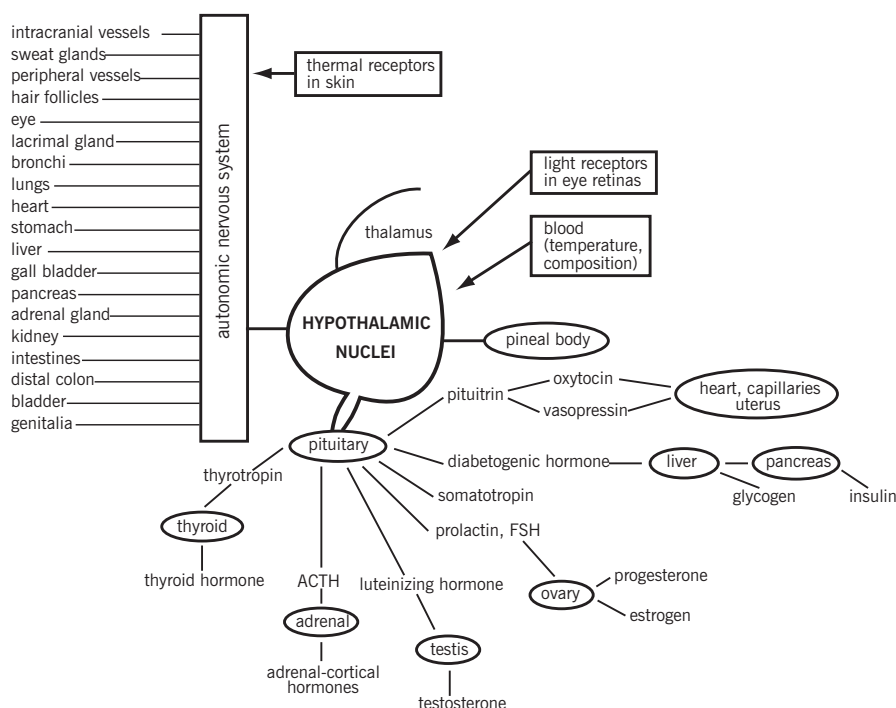


FIGURE 6.2. Hypothalamic environmental mediation. Sensing the environment, the hypothalamus acclimatizes the body through its direct neural connections to many organs and through the hormones it induces in the pituitary gland. ACTH, adrenocorticotrophic hormone; FSH, follicle-stimulating hormone.

in the SCN. This oscillation also turns on and off a gene for production of vasopressin and maybe other endocrines. The photoreception of the cryptochromes sets the body's elegant master clock of oscillating proteins. Individual light sensitivity can result in seasonal affective disorder (SAD; winter depression) and sleeping problems among the elderly, as well as difficulties adjusting to changes in longitude.

The main body mediator, the hypothalamus, controls the pineal gland in mammals, which acts as an endocrine transducer (i.e., it converts one form of energy to another). It synthesizes and secretes the hormones melatonin and serotonin. These are two of the most powerful chemical transmitters in the brain and can penetrate everywhere in the body. During the day, the pineal gland synthesizes serotonin. As darkness falls, the pineal gland produces a surge of melatonin. The cycle in the pineal gland is implicated as the clock that controls the development of puberty (Fellman, 1985). A sharp decline in melatonin production at puberty seems a trigger. Middle-aged adults secrete only half as much melatonin as children do. Besides causing sleepiness, melatonin dampens release of estrogen (in fact, it is being worked on as a contraceptive) and so may be related to timing of conception.

The hypothalamus also controls the pituitary, the body's master gland, and so it can adjust most of the body's metabolism, endocrine system, blood vessels, and other organs as appropriate for environmental conditions. As we shall see, this is important for acclimatization. In the 1980s SAD was recognized—the “winter blues,” with depression, excessive sleep, weight gain, and loss of interest in daily activities. The major therapy that has been developed is phototherapy, which consists simply of bright lighting for a short period in the morning, to communicate with the pineal gland.

Most people become aware of circadian rhythms when they travel across longitudes. The physical and mental effects of desynchronizing the body's rhythms from those of the environment are familiar as jet lag. It is not a matter of fatigue; the effects are the same even if one slept well on the airplane. Rather, body temperature and kidney/hormone cycles are no longer appropriate for the new schedule of social activities and daylight. Each biological clock is reset, advancing or slowing an hour or two each cycle, until all cycles are again synchronized with the environment. During the period of jet lag, however, the body may be more susceptible to infectious agents and (through impaired judgment and coordination) to accidents. There is growing use of melatonin to help reset the clock.

The large volume of international travel has made circadian rhythms and their adjustment a matter of concern to business, sports, and tourist industries, as well as to diplomats and the military. One topic of interest is the importance of circadian rhythms for work shifts. When people continuously work at night, first the sleep cycle, then the body temperature cycle, and eventually (it is thought) all the other cycles will be appropriately set to the social activity pattern. In some professions, however, shifts are regularly changed around or even rotated. This means that nurses, police officers, and other 24-hour-service personnel may frequently be involved in critical situations while effectively in a state of jet lag.

Most research on circadian rhythms has been medical. Are people who live in the long winter nights of high latitudes, for example, more susceptible to infectious agents than those in equatorial lands? Light is used today to induce chickens to lay more eggs and to bring cattle and sheep back into estrus after a miscarriage or failed fertilization. Does nighttime electric lighting affect human fecundity at all? Reading under bright light until late at night, as academics do, delays sleep onset by changing the circadian rhythms. Geographically designed studies of circadian rhythms would include contrasts between places and cultures (and occupations/activities by gender and class) and could be very illuminating.

Acclimatization

Acclimatization is the genetically given capacity of humans to adapt their physiological systems over long periods to heat, cold, and altitude. People are adapted to living in an extraordinary range of temperatures and other environmental conditions. They live and reproduce in hot deserts and Arctic tundra, as well as in the high Andes and Himalayas.

Behavioral and genetic adaptations are of great importance to acclimatization. Cultural geographers have studied adaptive differences in house types, designs, and construction; the materials and design of clothing; and the cultural rhythms of daily activity patterns. Genetic adaptations often involve the morphology of the body. In many animal populations, for example, long limbs and slimness serve to dissipate heat efficiently, while squat bodies with a low surface-to-volume ratio serve to conserve heat. Physical anthropologists have identified some such tendencies among human populations, but extensive migration has created as much exception to the rules (tall Norwegians in the Arctic and compact Chinese in tropical Asia) as conformity (compact Inuit in the Arctic and tall, slender Tutsi in tropical Africa). As important as genetic and behavioral adaptations are, this chapter looks at the physical processes of acclimatization and the implications for health that result from variations in physiological status of the population.

The physiological processes of thermoregulation are known in some detail. The human body must maintain a temperature of about 98.6°F (37°C), regardless of the environmental temperature. Over short periods, people react by shivering to generate heat from muscle movement and sweating to cool the body by evaporation. People everywhere sweat and shiver at the same skin temperature. Acclimatized people, however, use other physiological means to control body temperature. Their skin temperatures may not reach the threshold values for sweating or shivering.

The hypothalamus learns about environmental temperature from nerve channels to thermoreceptors in the skin, and about body temperature from nerves and the hypothalamic blood vessels. The hypothalamus causes the peripheral blood vessels to dilate (*vasodilation*) under heat stress, so that more blood can be brought to the surface for cooling, and to constrict (*vasoconstriction*) in protection against heat loss, so that more blood is held in the core of the body. Over time, the density of the capillary network near the skin will change. Most animals sweat from a few sites (armpits in humans, for example) controlled by the sympathetic nervous system. Humans have evolved an unusual capacity for eccrine (excretory) sweating by glands located all over the body. This involves excreting dilute solutions of water, salts, urea, sugar, and lactic acid for purposes of cooling the surface of the body through evaporation. Under conditions of prolonged heat, the efficiency of this sweating may be greatly increased. An acclimatized person sweats from the body's entire surface; the flux of water can reach several liters in an hour. The consequent drain of blood electrolytes (such as magnesium and sodium) is enormous, even though urinary output is considerably diminished.

By far the most complex adaptations to temperature involve the endocrine system (see Figure 6.2). Endocrines alter the basal metabolic rate to generate more or less heat from food. Because different kinds of food, such as fats or carbohydrates, have different capacities for heat generation, dietary needs change. Levels of thyroxin (which regulate carbohydrate metabolism) change, as do levels of iodine required to make thyroxin. Thus someone living at a cold, high elevation would need more iodine to maintain a high basal metabolic rate and would be more at risk of goiter than someone in a warm place, even if iodine were equally available in the environment. In fact, iodine is less available in glaciated mountains, which have

soils deficient in iodine and are usually remote from the transportation systems that would facilitate trade and the import of seafood, so goiter is often a serious problem in such areas.

Another endocrine system, the adrenocortical system, stimulates the pancreas to produce more or less insulin and controls the metabolism of the liver. It also affects capillary resistance and the removal and excretion of waste products from the blood. Even the relative amounts of types of blood cells change from winter to summer. Blood volume also increases under heat stress and decreases under cold stress. An important adaptation to cold is deposition of a layer of fat under the skin. The work of the heart, fat deposition in the blood vessels, blood-clotting time, blood sugar levels, and innumerable other physiological characteristics also differ in people acclimatized to different degrees of heat or cold, and even within an individual in different seasons.

Acclimatization to high altitude affects far fewer people, but in some ways is more dramatic than adaptations to heat or cold. The greatest adaptive stress is from low oxygen tension in the rarified atmosphere: More red blood cells must be produced, and hemoglobin (and iron nutrition) must be increased. The body must change its endocrine balance to prevent acidosis resulting from changed levels of carbon dioxide in the blood. Permeability of cell and capillary walls, and the facility with which red blood cells give up oxygen to body tissue, must be increased. Lung size, surfaces, and permeability are altered. When fully adjusted to elevations of over 10,000 feet (3,000 meters), the physiology of people alters sufficiently to create much more susceptible, and often hazardous, reactions to therapeutic doses of drugs. Fecundity also seems to be lower at high altitudes. When people acclimatized to high altitudes descend to sea level to seek work, they may be susceptible to respiratory infections.

THE INFLUENCES OF THE WEATHER

Weather is experienced by people on an immediate basis. It changes hour by hour and mile by mile. The frequency with which, and the amount that, it changes—the variability of weather—are its most important characteristics for health. The extremes of conditions reached are also significant.

Air masses—huge chunks of relatively homogeneous air flowing over the earth's surface, with a depth of up to a few kilometers—are as familiar to television weather report watchers as the fronts that separate air masses. Although tremendous energy exchanges occur at the fronts, the air masses do not generally mix. They retain their own characteristic temperature, humidity, atmospheric (barometric) pressure, and other properties. The fronts that separate them are accompanied by winds and, especially if associated with squalls and thunderstorms, by the generation of ionization, extremely low frequency (ELF) waves, and infrasound (too low in frequency for humans to hear). Let us review these individual elements and their health implications before considering what happens when fronts pass or extreme conditions develop.

Temperature, Humidity, and Air Movement

Temperature, humidity, and movement of air together determine how readily the human body loses its heat. Such heat is generated by metabolic activity, even when the body is at rest. We have considered the major physiological processes by which the body controls the generation, conservation, and dissipation of heat during acclimatization. The ability of the body to radiate heat depends on the surrounding temperature. The effectiveness of evaporative cooling depends on the surrounding humidity and air movement. Temperature–humidity, “comfort,” and wind chill measures portray environmental stress much more accurately than a simple measurement of temperature, which is commonly used by medical scientists.

The most obvious health impact of the weather involves extremes of heat and cold. Death from exposure to cold involves no mysterious mechanisms: When the body’s core temperature falls below a certain point, its chemical reactions slow down and finally cease. Heat causes increased peripheral circulation, which can be stressful in its own right, and copious sweating can drain the body of electrolytes. In heatstroke, the sweating mechanism shuts down, allowing body temperatures to rise rapidly. Certain proteins in the brain can permanently change, even if the body is cooled before death results.

Heat waves have long been associated with increased mortality. This association seems to be discovered *de novo* after every event. Persistently high temperatures seem to be more closely related to mortality than are the peak temperatures reached in an event. The most consistent predictors seem to be the average temperature and the number of successive hot days. Increased mortality usually occurs in the first 2–3 days following the heat wave’s peak. Greenberg, Bromberg, Reed, Gustafson, and Beauchamp (1983) studied the 1980 Texas heat wave and compared it with the heat wave of 1950 and the period 1970–1979. Temperatures of over 100°F (37.7°C) existed for 61 of 71 days in 1980, and 107 deaths were attributed to heat. Not surprisingly, the risk of death was highest among those engaged in heavy labor and among the elderly. Death rates were higher during heat waves of former years, Greenberg et al. found; this was presumed to be because of the increased prevalence of air conditioning today. In terms of our triangle, the lack of access to the (behavioral) technological intervention of cooling in the built habitat is a killer. Yet climatologists, who are geographers, can focus so intently on the natural habitat that they fail to consider cultural buffers or even the built environment. Kalkstein and Davis (1989), for example, conclude that “regional acclimatization” appears to be especially important in summer, since the strongest associations of heat with mortality occur in parts of the United States where hot weather is uncommon (the Pacific Northwest and the continental North), and the weakest associations occur where heat is most common (the Southeast and continental South). This they took to be evidence that people respond to weather in a relative rather than an absolute fashion because the population responds more dramatically where hot weather occurs rarely. They also found in their models of nonwhites, however, strong mortality responses to heat in southern cities (Atlanta, Birmingham, Memphis) despite lack of such response in white models. This they took to indicate that nonwhites

may be more sensitive to heat! (Of course, the justifications for slavery were just the opposite, as African ancestral lands are not known for coolness.) In social sciences, addressing such associations, one speaks instead of the *vulnerability* of a population, usually due to socioeconomic circumstances. Today the social risks involved in vulnerability include the increased numbers of elderly persons (whose physiology is less efficient at cooling) living alone; the risk of living alone, compounded when people decline to open their windows for improved circulation because of fear of crime; and, of course, lack of money to buy cooling technologies. Chicago, St. Louis, and many other U.S. cities have responded to these risks by organizing the provision of fans and warning family members and neighbors to check on the elderly.

In August 2003 in Europe, almost 15,000 more people died than expected for the average August; in Italy alone there were 4,000 excess deaths; altogether at least 30,000 lives were lost in the largest European natural disaster in 50 years. The heat wave was caused by a high-atmospheric-pressure cell that sat unmoving and blocked any instability, winds, or rain. Switzerland and other places had the hottest summer in 250 years, which was of course taken to prove global warming and confirm predictions of future trends. In addition to the heat, however, it was important for France that during August most of the population goes on vacation, including doctors and hospital staffers. Elderly persons at home without air conditioning in this unexpected weather, with few cultural buffers in place and few social networks, were found dead (sometimes days later) or taken to hospitals which were not prepared to receive them. Shocked and now forewarned, French social welfare and medical care systems responded and planned, so that in the heat of the summer of 2008 there was no notable excess of deaths.

There are several population predisposing factors for suffering heatstroke. These include the presence of such degenerative diseases as cardiovascular disease, renal (kidney) disease, and hypertension; preexisting acute diarrheal or febrile disease; salt deficiency and dehydration; and the use of drugs that affect the thermoregulatory system (alcohol, amphetamines, and such therapeutic drugs as diuretics and anticholinergics). People who undertake sustained activity in the heat are at risk, as are people whose lack of fitness or acclimatization makes their cardiovascular systems less able to cope with the necessary heat dissipation.

Heat waves have also been associated with increased crime, especially homicide, but the pattern is not systematic or consistent. A plausible biological pathway is that heat interferes with sleep and therefore with dream cycles. This topic cries out for some behavioral-vertex research.

As noted above, an increase in the number of deaths following heat waves has been repeatedly observed. It is not clear, however, how much of the increase in mortality occurs in those about to die from degenerative disease anyway. Several studies have reported less-than-normal mortality shortly after the peak associated with a heat wave, but the decrease does not usually offset the previous increase. There is some concern that people surrounded by air conditioning may be losing their acclimatization to heat, so that they are less able to cope with extreme heat stress. Mortality, however, is certainly higher among those socioeconomic groups least likely to enjoy air-conditioned homes.

Air-Borne Life

The air is charged with living things. The transport and survival of bacteria, viruses, fungi, and allergens such as pollen depend on certain conditions of atmospheric temperature, humidity, condensation, and movement. Agents of human disease are injected into the air by coughing and sneezing; by the shedding of hair and dead skin; and by the spray of cooling towers, air conditioners, and irrigation systems. Soil bacteria, fungi, and pollen are picked up by the wind. Dispersal depends on atmospheric turbulence. To all such life forms, ultraviolet radiation ultimately is lethal.

Temperature and humidity are the limiting factors for survival in air; they act together with different effects on different organisms. Bacteria often can withstand extremes of temperature, but may be seriously affected by humidity. The ubiquitous intestinal bacterium *Escherichia coli* survives for less than 3 hours at 68°F (20°C) and 50% relative humidity, but *Mycoplasma* of human pulmonary origin can survive up to 5 hours at 82°F (28°C) and 50% humidity. Viruses can multiply only in living organisms, but they can survive in the air anywhere from a few seconds to a few hours. Influenza viruses survive longest in cool but nonfreezing air, and stay suspended longer when there is low humidity (Lowen, Mubareka, Steel, & Palese, 2007). Polio viruses have been shown to be progressively desiccated in the air, but to have sufficient longevity for dissemination over several miles. Tolerance and preference for environmental conditions are highly specific to the type of organism. In one study, for example, a polio virus survived best at 80% relative humidity, but a vaccinia virus survived best at 20% (Hyslop, 1978).

Air-borne bacteria and viruses have caused epidemics miles from their sources. An outbreak of Q fever in San Francisco, for example, was traced to the fumes from the fat-rendering plant of a slaughterhouse. The combined effects of temperature and humidity are thought to result in different survival times for air-borne microorganisms in summer and winter, and thus to be critical factors in the seasonal incidence of disease (described later in this chapter).

Both fungi and pollen are suspended in the air. A few fungal infections of the lungs are serious diseases, especially histoplasmosis and coccidioidomycosis. Their distribution is clearly related to temperature, humidity, rainfall, and local winds, which pick up the fungi from disturbed soil. The distribution of human disease may be limited by prevailing winds as well as soil temperature and moisture. Ragweed, the most common pollen source for the hay fever that afflicts more than 6% of the U.S. population, grows best on cultivated land. Again, rain, wind, and humidity affect dispersal.

Atmospheric Pressure

The weight of the atmosphere presses on the surface of the earth. Subsiding air exerts more pressure than rising air. One atmosphere of pressure, usually defined as 1,013 millibars (mb), is felt as 14.7 pounds per square inch (1,034 grams per square centimeter) at sea level. Atmospheric (barometric) pressure changes are slight in

equatorial zones. Over North America normal pressure changes between air masses are on the order of 25 mb. At high latitudes pressure changes may reach 120 mb. The most extreme low and high pressures ever recorded on earth amount to a difference of about 3 pounds per square inch (211 grams per square centimeter). (Altitudinal changes, such as those encountered in balloons or depressurized aircraft, can cause much greater pressure differences.)

The major mechanism postulated for why normal changes in atmospheric pressure affect the body's biochemistry is that body volume expands slightly, leading to retention of water and therefore to alteration of electrolyte balance. Eventually levels of disequilibrium are reached that trigger intervention by higher homeostatic control systems to restore proper fluid levels. These changes would result in water storage in certain parts of the body—in joints, eyes, and so forth—and could cause joint pain, glaucoma pain, increased blood pressure, blood clotting, and general irritability.

Winds

Winds blow between the different atmospheric pressures that characterize air masses, and between zones of subsiding and rising air, as part of the heat balancing of atmospheric circulation. The strength, frequency, and duration of winds vary greatly from place to place. Winds transport dust, fungi, bacteria, and insects long distances. Early approaches to eliminate river blindness in local areas of West Africa, for example, were stymied by seasonal winds carrying in new infected adult blackflies.

Winds have another effect. The movement of air, especially air with low humidity, promotes the ionization of atmospheric gases. Electrons are stripped from their atoms, producing positive and negative charges. In the extreme form associated with convective buildup of cumulus clouds, segregation of electric charges may result in lightning. The degree of atmospheric electrical charge is one characteristic of an air mass. Dry winds, such as the chinook, Santa Ana, foehn, sirocco, and harmattan, are associated across cultures with irritation, bad temper, accidents, and violence. Despite an extraordinary richness of folklore, however, there has been little in the way of controlled studies. The heat itself used to be implicated, but suspicion is now directed at ionization, the electric charge in the air.

Atmospheric Ionization

Ionization in the atmosphere has several sources. It results from radiation impact as well as the friction of movement. The charged molecules are electrical and chemical stimuli with undetermined consequences for human health. Persinger (1980) has reviewed and explained many of the phenomena involved with the "electromagnetic stimuli of the weather matrix." Much of the following discussion draws on his book.

Atmospheric ions can react with chemical processes in the body by giving or taking electrons if they have access to biochemical pathways; therefore, their potential

to be culprits in the etiology of suicide, homicide, aggression, infection, migraines, conjunctivitis, and respiratory congestion needs to be considered seriously. Their physical influence as environmental electrical stimuli is generally dismissed on the grounds that any current the ions might induce in the human body by their presence in the atmosphere would be too small to have an effect, even around the brain neurons' electric field. Chemically, however, there is a pathway for possible interaction: By absorption into the bloodstream through the lungs, the ions may alter the pH by acting as oxidizing or reducing agents. This can influence a host of chemical pathways and homeostatic sensors/controls.

The serotonin irritation syndrome of migraines, nausea, irritability, edema, conjunctivitis, and respiratory congestion has been correlated with the accumulation of positive ions that occurs as much as a day before the weather changes. The serotonin hypothesis, in brief, is that in the brain a major chemical transmitter whose activities are associated with aggression, sexual activity, pain sensitivity, and other functions is affected by bursts of atmospheric ions. The mechanism is not explained yet, but if it exists, it may be relevant to depression, irritability, suicides, report of pain, and other cross-cultural correlates of the weather.

Unpleasant effects have usually been correlated with positive ions, whereas negative ions are generally associated with feeling good. Positive ion charges occur naturally before thunderstorms, and they accompany hot, dry winds. Humans generate positive ions by such means as heating systems, in which cold, dry air is heated (making it drier) and blown through metal pipes. Dry throats, mouths, and membranes, which become more permeable to viruses, may result from sleeping and living in heated environments. Therapeutic claims made for machines that generate negative ions for bedrooms have not been proven.

Extremely Low Frequency Waves and Infrasound

A few scientifically sound studies have been done on the effects of geomagnetic storms upon moods or brain function. Most of these have focused on reaction times, which are quantifiable measures. It is possible that the intensity and quantity of ELF waves (3 hertz [Hz] to 3 kilohertz [kHz] or so) generated by such storms could influence the brain because they include the frequencies of the brain's own transmissions. Most ELF energy (generated, for example, by lightning and human-made sources) remains between the earth and the ionosphere, and travels long distances without appreciable attenuation. The major variations in the number and intensity of ELF waves and their sources result from diurnal and seasonal changes.

Infrasound shares some of these ELF characteristics, in that it can travel long distances with little loss of energy. At frequencies too low for humans to hear, the waves have the potential to generate whole-body vibration by resonance. They are generated by severe weather disturbances and, traveling at the speed of sound, may be detectable weather precursors. In laboratory studies of infrasound, 10 Hz at 115 decibels can cause lethargy, euphoria, and loss of time judgment; these are frequencies generated inside closed automobiles traveling at 100 kilometers an hour. At other frequencies, infrasound is associated with nausea and dizziness. The research

literature, however, is sparse. There are few weather stations that measure infrasound, and fewer correlation studies under controlled conditions.

Passage of Fronts

The passage of a front marks a change of air mass. Especially when the passage is accompanied by squalls, it entails winds, electric charge, and the generation of ELF waves and infrasound. As the air masses change, so do many of the environmental conditions that have been affecting people: temperature, humidity, brightness or cloud cover, suspended living organisms, and atmospheric pressure. The body must cope with all these stimuli and their changes. Accumulated stresses can lower the level of health so that individuals fail to rally (Chapter 2). It is thus not surprising that the passage of fronts has often been related to increased mortality.

There have been numerous studies on the influence of frontal passage and weather changes on morbidity and mortality rates. Weather types have been classified in many ways, and various weather elements have been examined singly or together. The best correlations have been attained when weather is treated as a complex of elements. Those variables most often found to be important include mean temperature, diurnal change, mean dew point, mean visibility, mean barometric pressure, day-to-day change in barometric pressure, mean wind speed, and various comfort indexes used to adjust temperature for humidity and wind speed. The most consistent relationships of the rates have been to intensity of change in the weather. The most consistent mortality and morbidity associations with weather conditions have been overall mortality, cardiovascular disease, and asthma; respiratory illness and psychiatric conditions also have some associations with weather.

SEASONALITY OF DEATH AND BIRTH

Death Seasonality

It has been known for at least 2,500 years that some causes of death occur more often in one season than in another. A century ago it was common knowledge in the United States that in summer people died of malaria, yellow fever, cholera, typhoid, gastroenteritis, and tuberculosis, whereas in winter people died of influenza, stroke, and cold-related causes. The seasons are not much changed, but the former mortality patterns have been obliterated. How have people caused this?

The late-20th-century literature on seasonality of mortality was dominated by one person, Masako Sakamoto-Momiyama of the Meteorological Research Institute in Tokyo. Her work over 25 years discovered, explained, and popularized the importance of seasonal patterns and how they are changing. After analyzing the seasonal pattern of occurrence of scores of diseases at various latitudes; in various climates; and among different age, socioeconomic, race, and ethnic groups, Sakamoto-Momiyama developed general models of variation (Figure 6.3). Mortality patterns in most midlatitude countries were bimodal; that is, they had both winter and summer peaks, as noted above. With economic and social development, the

summer peak disappeared, and mortality now had its highest incidence in winter. Sakamoto-Momiyama differentiated several types of shifts in seasonality. In the *transitory type* (cancer in Figure 6.3), the summer peak of incidence gradually shifted into autumn as it decreased, perhaps moving toward a winter peak. The *reversing type* reversed from a summer to a winter peak. It had two subcategories: diseases for which the former summer peak became a trough and a new winter peak was created (reversing type A), and diseases that formerly had both a winter and a summer peak but lost the summer one (reversing type B). Sakamoto-Momiyama classified gastroenteritis, tuberculosis, and beriberi as type A, and many of the degenerative diseases associated with aging as type B.

As Sakamoto-Momiyama (1977, p. xiv) related, while doing research in 1965–1966 in New York, she was surprised to find the winter peak disappearing for many causes of death in the United States. Her detailed studies of infant mortality in particular led her to revise her previous idea that technological progress would lead to concentration of mortality in winter, as she had concluded from her experiences in Japan. Instead, she hypothesized that a “deseasonalization” of mortality would occur. Her later work noted that Japan and Western Europe were following the United States and Scandinavia into this deseasonalized pattern of mortality (Sakamoto-Momiyama, 1977).

Why does this matter? It is amazing that people, society, could do this. The relevant question is how human behavior could alter such a strong and stable pattern of mortality association. The seasons are constant, but obviously human–

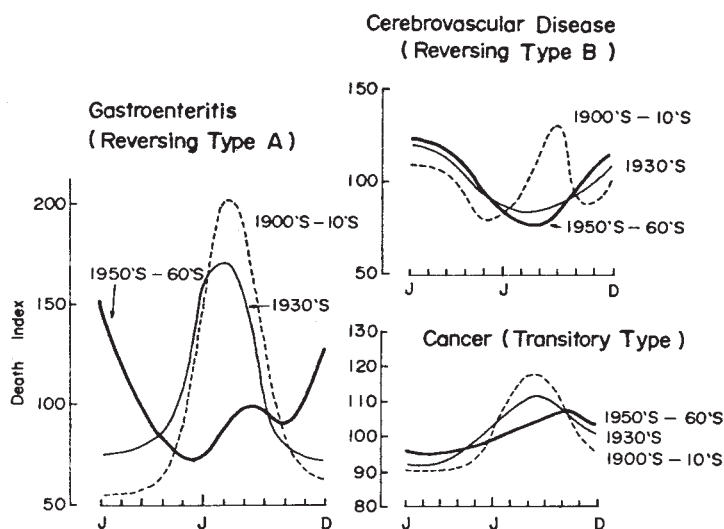


FIGURE 6.3. Models of seasonal shifts in mortality. When the death index by month is plotted for several time periods, the peak of mortality for some diseases reverses from summer to winter; for others the summer peak gradually disappears, leaving a previous winter peak as the only seasonality; for still others the summer peak decreases and shifts toward fall. From Sakamoto-Momiyama (1977, p. 33). Copyright 1977 by University of Tokyo Press. Reprinted by permission.

environmental interaction has been altered fundamentally. There can be many reasons in cultural ecology why mortality incidence would lose seasonality under conditions of economic and technological development. Heavy labor and exposure to the elements decrease. Diet changes, and a variety of nutritious foods become available all year. Medical technology advances. Health care becomes more accessible in winter. For some causes of death, one can point to specific behavioral developments: Refrigeration and the chlorination of water supplies control summer gastroenteritis, for example; vaccination for polio means that swimming pools no longer start summer epidemics; changes in house type, screening, occupation, and other factors have led to the eradication of such diseases as malaria and yellow fever in some countries. Sakamoto-Momiyama (1977) advanced the hypothesis that much of the deseasonalization of mortality was due to central heating and air conditioning. She pointed to the relationships between temperature and cerebrovascular disease (Figure 6.4). Sweden and New York City, which had widespread central heating by 1960, showed consistent linear relationships without changes of slope at high temperatures, whereas Western Europe and Honshu in Japan, where many houses were still not heated in 1960, had seasonal changes of slope. In Hokkaido, which is cold in winter, people used heating systems that were turned off when the temperature reached 54°F (12–13°C); the winter slope was small until that point and afterwards resembled Honshu's (represented by Tokyo in Figure 6.4).

Much research has been stimulated by these studies about what influences seasonal mortality today. Are heart attacks due to shoveling snow? To heavy winter diets and sedentary lifestyles? Vignette 6.1 discusses some of the statistical procedures essential to seasonality research. Although some diseases have been analyzed in statistical detail by epidemiologists, few studies have been done on geographic

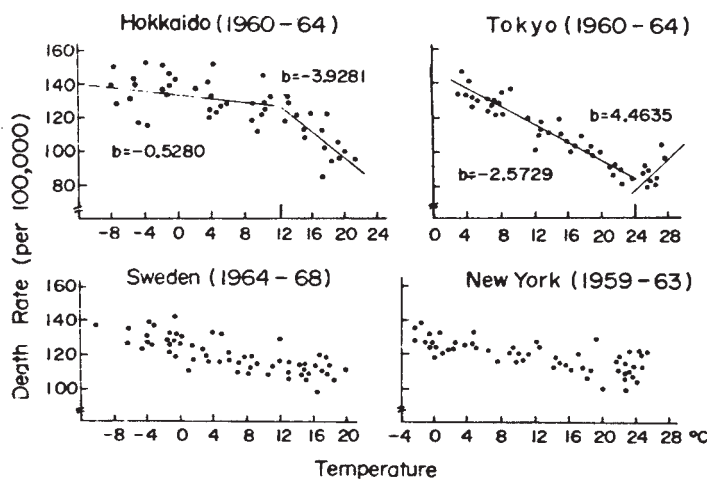


FIGURE 6.4. Association between cardiovascular mortality and temperature. In Hokkaido (northern Japan), Sweden, and New York, deseasonalization of mortality seemed to be associated with use of central heating. From Sakamoto-Momiyama (1977, p. 165). Copyright 1977 by University of Tokyo Press. Reprinted by permission.

patterns, and few national studies have been completed by geographers. Kevan and Chapman (1980), for example, confirmed the deseasonalization of many diseases in Canada, but found great seasonality continuing for bronchitis, pneumonia, influenza, and circulatory diseases (winter) and for violence, accidents, and poisoning (summer).

Influenza continues to be a winter disease. The association of cooling winter temperatures with increased influenza mortality, after a certain lag, is indubitable. The association persists over widespread locations and different climates. Speculation has blamed several factors:

1. Biometeorological conditions cause the nasopharynx and trachea to be dry, and membranes to become more susceptible to virus penetration. (This is also a problem often associated with air travel.)
2. The virus can survive in the air between hosts more easily when the air is relatively dry and cold than when it is hot and humid. (See new studies on influenza in Chapter 10.)
3. The lower solar radiation, and hence ultraviolet radiation, promote virus survival.
4. The body's seasonal metabolic changes make it more susceptible.
5. People gather indoors in closer proximity for longer periods in winter; schools are in session; and even recreation tends to be indoors—points emphasized by the importance of room density as a correlate in many influenza epidemic studies. However, prisons are crowded all the time, and influenza is a winter disease there too. Any explanation of influenza's seasonality must account for its winter incidence in Florida and Hawaii, on the Great Plains and in New England, and in perpetually crowded institutions such as prisons.

Not only infectious diseases are involved in seasonality. Mortality from heart disease is negatively correlated with temperature in several midlatitude countries. Between 30° and 70°F (−1° and 21°C), correlation coefficients of −.95 have been found. A 2.5% increase in mortality has been estimated for each degree Celsius decrease in mean monthly temperature, although the exact relationship differs by climate area. Sometimes mortality from ischemic heart disease but not cerebrovascular disease, and sometimes the reverse, have been correlated with temperature change.

In general, respiratory disease shows the strongest association with winter, and cancer shows the least seasonality—which is hardly surprising, given its long incubation period. The meteorological conditions and physiological stresses of acclimatization are undoubtedly involved in seasonality, but so are culture rhythms and technology. The etiology of continuing seasonal patterns of mortality remains puzzling.

Birth Seasonality

Seasonality of birth, found everywhere, has been much less studied than seasonality of death. The seasonal incidence of birth in Japan is quite different from that in

European countries, and that of the United States is again distinct. The index of birth (see Vignette 6.2) in Figure 6.5 illustrates the bimodal pattern that has persisted in the United States for decades but has shown signs of changing lately. There have been a few good descriptions of the seasonal patterns by region, race, and other population characteristics in the United States and Europe, including at least one study in which no seasonality was found in a population (Arcury, Williams, & Kryscio, 1990). In some studies, the amplitude (height above and below the index line) has been found to be greater for illegitimate births and births to the poor, to the less educated, and to blacks, although there are a few indications that these relationships may be changing. Despite minor differences in the amplitude of seasonality among population subgroups in the United States, however, the pattern of seasonality has been remarkably similar. The relative peaks vary in magnitude and timing across the country; they have changed over time, but the manner of this change has not yet been studied.

One possible reason for birth seasonality is that high temperatures may have an effect on spermatogenesis (the creation of sperm). High temperatures may also injure existing sperm. As with other biometeorological effects, threshold and range are presently unknown. How high a temperature over how long is necessary to affect fertility? Is a single episode, such as playing tennis in tight athletic clothing

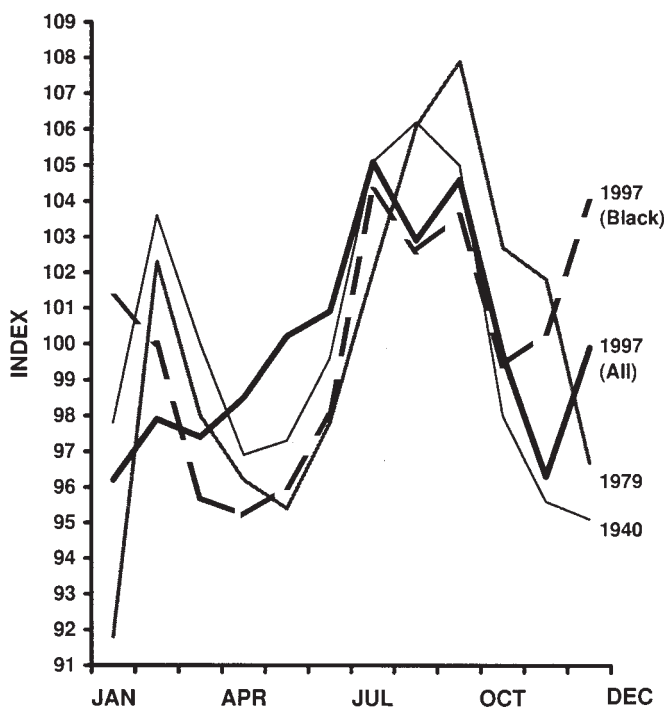


FIGURE 6.5. Seasonality of birth in the United States. The bimodal pattern has remained remarkably constant for decades. In recent years, the April trough (summer conceptions) and February minipeak have been lost overall, but African Americans still retain the southern pattern.

on a hot day, enough exposure? Or are weeks of hot, humid weather needed? Is birth, like mortality, also becoming deseasonalized?

The seasonal patterns of birth may have nothing to do with temperature. They may be a result of agricultural patterns; school calendars; the timing of holidays and vacations, Christmas–New Year celebrations, and leisure time in the United States; and the timing of equivalent events in other cultures. When people migrate, does their seasonality of birth pattern change relatively suddenly under the new environmental conditions or gradually, over a generation or more, through cultural adaptation? Vignette 6.2 discusses the various hypotheses about the cause of seasonality of birth and its relevance to the triangle of human ecology.

Birth seasonality has innumerable consequences for health and disease. The most obvious is the variable need for health services. Obstetrical wards built for peak periods will have many idle beds during the trough. Birthing classes and maternity leaves from work are affected. More significant effects, however, may involve congenital birth defects, premature births, and neonatal mortality (Keller & Nugent, 1983). There is considerable seasonal variation in neonatal mortality, but its correlates are not clear. The point often made by scientists who study congenital disease is that if there is seasonality to it, the condition is not the result of genetics alone. Something in the environment after conception—whether in the womb, at birth, or after birth—must actually cause the expression of the disease. As bones, nerves, palates, and endocrine systems form at different times in the development of the fetus, biometeorological influences that may be relevant to congenital disease may not occur at the time of conception at all. This is especially true for one mental disease, schizophrenia. A strong peak of birth for individuals with schizophrenia seems to occur during late winter and early spring in many countries. These individuals would have been fetuses in their third month, when the central nervous system is forming, during the late summer heat that is associated (in the United States) with the minimum of conception. This, of course, is also (especially at high latitudes) the time of maximum sunlight and effect on serotonin production and other hormone cycles.

The problem of separating environmental and cultural factors is complex. Mothers are likely to be eating different foods, exercising in different ways, and following different work and sleeping habits during summer than they are in winter. They may spend different amounts of time outside breathing polluted air, or the air may be seasonally polluted. It is suspected that for some congenital diseases, seasonality may result from the activity of a viral agent. This is an analogy to the congenital effects of rubella (German measles) infections. Most physicians studying a seasonal pattern in a congenital disease will look to changes in food, drink, sickness, and medicines before they will consider the weather.

Other Biometeorological Effects

There are many other, more indirect biometeorological effects on mortality and natality. Chief among these are the effects of climate on agricultural cycles and

cropping patterns, and hence on many nutritional problems. The “hunger season” recognized in wet–dry climatic zones of Africa, for example, creates nutritional susceptibility to infectious diseases and other stresses. In many places the weather cycle creates periods of high and low employment, of transportation difficulties and poor access to health services, and of stagnant air or atmospheric inversions and intense pollution episodes (see Chapter 7).

HOW CLIMATE CHANGE IS LIKELY TO AFFECT HEALTH AND DISEASE

Media of all kinds scream at us today that global warming is going to be catastrophic for human health. They don’t actually put it that way. They say, “If MERCURY SOARS, SO WILL HEALTH HAZARDS,” or proclaim that “TENS OF THOUSANDS HAVE ALREADY DIED FROM ASTHMA” because of global warming. Or they say that “MALARIA WILL SPREAD, EVEN TO THE UNITED STATES,” “THE HEAT WILL BREED MORE MOSQUITOES,” “FOOD WILL SPOIL,” “STARVATION, MASS DISPLACEMENT OF POPULATIONS, AND WARS WILL BE UPON US.” Maybe the media are right—but they are picking up and spreading impressions of information, facts still being gathered, and half-baked hypotheses in the scientific studies addressing the momentous process that is upon us. The climate *is* changing, and the biosphere of earth is warming. The media reports, however, leave out many aspects of scale and proportion; and, like those who discovered the effects of heat waves described earlier in this chapter, they almost always leave out the cultural buffers and social organization we first discussed in Chapter 2. The issue of health consequences from global warming needs to be addressed (metaphorically speaking) in lowercase letters, not capitals. There are also many misunderstandings about what climate change means for a particular place, and whether events there right now prove or disprove that warming is happening.

One common misperception is that global warming will happen wherever one happens to be, and at the amount forecast globally. The questions then become whether Minnesota, Texas, or France is warming 1° or 3° and by what decade. One may then be tempted to respond, “So what? That’s not so much hotter than it is now.” Energy comes mostly from the vertical rays of the sun moving over the equator from one tropic to the other. This tropical heat has to be circulated and dispersed around the globe. Vignette 6.3 describes the climatic zonation and associated biomes that currently result. The process of global warming will cause some climatic zones to expand latitudinally (and altitudinally) and may alter seasonal timing in a particular place. The exchange of energy will result in warmer seas, causing stronger hurricane winds, shifting ocean currents that transport heat, and (probably most important for human health) shifts in the pattern and amounts of precipitation. Precipitation that falls as rain instead of mountain snow will run off and not be available to sustain vegetation and agriculture when it melts in summer. Monsoon extremes and unpredictability will increase. There may or may not

be more storms, but the hurricanes and tornadoes that occur will frequently have stronger winds. The initial forecast of the 1970s climatic models was for “growing extremes”—the blizzard of the century again, the 100-year flood repeated, the “worst drought” recurring. Using the classifications that we have used so far in this text, Table 6.1 summarizes these and other concerns. It incorporates health-related ideas from the Centers for Disease Control and Prevention’s Policy on Climate Change (J. Gerberding, policy statement, 2007); Committee on Climate, Ecosystems, Infectious Diseases, and Human Health (2001) of the National Research Council; and the International Panel on Climate Change’s Fourth Assessment Report of impacts on human health (Confalonieri et al., 2007).

TABLE 6.1. Climate Change and Warming: Health Impacts

Potentially affected	Weather event	Health effects
Biometeorology	Heat waves	Heat stress; heatstroke
	Cold; seasonal change	Respiratory distress
	Fronts, air masses	Arthritic attacks, CVD effects
Air	More heat, humidity, atmospheric inversions	Allergies to increased pollen, fungi; aerosol contagion exposure to pollutants
Water-borne pathogens	Drought	Shortage of water for hygiene and sanitation increases disease transmission, virulence, compromises buffers
	Flood	Groundwash of enteric pathogens, chemicals into water supply
	Warming	Greater and longer survival of pathogen transmission
Water-based pathogens	Drought	Reduction and elimination of vector breeding except for domiciliated vectors/ reservoirs, which increase contact
	Flood	Increase in mosquito and other vector breeding
	Warming	Greater winter survival and altitudinal extent of vectors
Nutrition	Drought, flood, warming causing salinization, failed pollination, and other ecosystem disruption and economic consequences; repetitive crop failure	Regional: severe malnutrition; local: human and livestock starvation; forced migration to cities and consequences of crowding in inadequate sanitation
Hazards	Number and severity of storms, flooding increase; coastal land loss to sea rise; mass population displacement	Increased trauma (drowning, injury); exposure to elements; mental health difficulties (PTSD, depression, substance abuse, domestic violence)

Note. CVD, cardiovascular disease; PTSD, posttraumatic stress disorder.

Biometeorology

As this whole chapter has emphasized, human health is affected by biometeorology. So fortunately, the tilt of the earth's axis and the earth's rotation around the sun are not going to be changed by greenhouse gases. Solar intensity, sun angles, and seasonal shifts in the vertical rays remain constant. There may be some shift in visible radiation (see the earlier discussion of the radiation spectrum). Greater differences between air masses may affect arthritis or the timing of heart attacks as the fronts pass. Cold in severe winter storms, "dropping" from the Canadian Arctic deep into the U.S. South, may affect winter mortality. These are minor concerns, indeed; the main effect will be through heat waves. As you have read, however, we humans can erect strong cultural buffers and use our cultural adaptability to cope. France was much more prepared, with community and medical support, for the heat in the summer of 2008 than for its historic heat wave in 2003.

Air

Most of the attention about nongaseous changes to the air we breathe has been paid to the particles (spores of fungi, pollen, etc.) suspended in it and to carbon monoxide, ozone, and other pollutant gases (Chapter 7). Mold allergies and hay fever will probably be adversely affected. Concerns that the warmer, more humid air will raise the risk of aerosol contagion, however, overlook the fact that viruses such as influenza like it cool and dry so that they can stay suspended in the air and maximize contagion. In some places, warming may make for less influenza! Stronger winds over dryer land will also pick up more dust. This has been a historic health problem for northern China, as well as for West Africa, where the latitudinal band of dust along the grassland biome (Vignette 6.3, Figure 1) that has been associated with linear expansion of meningococcal meningitis epidemics may worsen.

Precipitation Effects

The central importance of water for transmission of human disease has been developed in Chapter 2. Water-borne pathogens do not have a simple relationship with warmer weather, but decreased precipitation is sure to make health conditions worse. In water that is warmer, or that is warm for longer, some organisms like *Vibrio cholerae* may survive longer. The predators of amebas, helminth eggs, bacteria, and mold may also thrive, however. During drought—as water sources shrink in area and as wells become almost airless masses of algae—not only can pathogens become concentrated, but humans and other animals lose the option of drinking elsewhere. In wet-dry climatic areas like northeastern Thailand, the cycle has always been that when the monsoon rains break, they wash the chicken, pig, bovine, and human manure on the ground into sheets of water that contaminate the wells; and during the dry season, when other surface water dries up and even fewer wells have water, people get sick from using the contaminated water. Something

like this threatens to occur on a much larger scale. The cultural buffers of chlorinating and otherwise treating water, pumping deep water, or finding renewable fuels to boil water cheaply will become more important. Such buffers will need improved social and economic organization as well as new physical infrastructure to support them. This impact of warming works against the United Nation's millennial goal of greatly decreasing the proportion of the world's population without safe drinking water.

Perhaps the most attention has been given to the vectored diseases, especially those classified as water-based. Remember that effects of climate warming are going to be species-specific and not simple generalities. Greater drought and less water on the land after rain should mean less mosquitoes in general and less malaria and snails for schistosomiasis. But on the other hand, the need for greater container storage can easily result in more *Aedes aegypti* mosquitoes and greater risk of dengue hemorrhagic fever and other arboviruses. More reservoirs of water will also need to be constructed, and these will include drainage and irrigation channels that will support more snails. Without human intervention to make it worse, however, drought should be bad for more than water-based vectors, as tsetse flies, sand flies, and even reduviid bugs need humidity.

Climate warming does not simply mean greater drying, however. It means greater storms, more flooding, erratic monsoons, and displacement of downpours from where they have historically occurred to other places entirely. In recent years the U.S. Southeast has repeatedly had prolonged drought as a ridge of high pressure remained stationary over it for months, blocking passage to other air masses and the frontal rain these would bring, as well as suppressing any instability from ground convection that could produce thunderstorms. The Midwest and Pennsylvania got the Southeast's rain as well as their own normal precipitation, resulting in devastating floods in those regions. Is this just an example of normal weather fluctuations, or is it the beginning of a new climate pattern? It can be difficult to understand what one is in the middle of! The consensus, nevertheless, is that government should build more reservoirs and that citizens should learn and implement better water conservation.

Satellite imagery of water on the land in East Africa and after hurricanes in the United States is now used to predict mosquito-borne outbreaks of Rift Valley fever, West Nile encephalitis, Eastern equine encephalitis (in the United States), and of course malaria. More subtle is the effect of warming on the survival of mosquitoes over winter. La Crosse encephalitis in the Midwest flyway is a virus that has been shown to pass transovarially under favorable conditions, such as a warm winter. *Ae. albopictus* has a range extending from India to Japan in Asia; in warm winters it can expand in the United States. The eggs and nymphs of ticks usually winter under the leaf debris. When the disease agent is passed transovarially, as is the rickettsia that causes Rocky Mountain spotted fever, more infection will survive the start of the spring cycle.

The scariest journalism reports that malaria will be spreading and malaria-causing mosquitoes will be breeding even in the United States. As described in Chapter 4, most of the United States was malarious in the past. It still has *Anoph-*

eles mosquitoes everywhere; the mosquitoes have not been eradicated. Malaria gets reintroduced to the United States frequently when students and other travelers come back infected and are treated. What prevents the possibility of new malaria here, however, is not that the temperature is 1° lower than it might become, but that strong cultural buffers have been established—ranging from laws on water impoundment to air-conditioned, glass- and brick-walled houses with indoor plumbing. Might malaria and other mosquito-borne diseases occur over a wider area, and especially at higher elevations? There are already reports of malaria's being transmitted in schools and villages higher in the East African highlands than occurred even a decade ago. Like secondary nidi, however, these are marginal lands for the mosquitoes. The best defenses against expansion of the disease range are public policy, economic development, and better education.

Nutrition and Hazards

For all the attention given to mosquitoes and hay fever, the risk of all the possibilities described above is trivial compared to the much more likely impacts on human health of global warming: destruction of good agricultural land and crop cycles; impoverishment of regional populations; ecological disruption to birds, bees, and many other seasonal cycles; coastal flooding; and mass displacement in forced migration of hundreds of millions of people from such areas as Bangladesh, the Pacific islands, and the Mediterranean coast. In hurricanes (typhoons, cyclones) and other extraordinary rain events, people drown as their houses and the cars they're driving are washed away. Crops are spoiled; rice fields and other agricultural land are spoiled with salt ocean water. Whole populations become depressed as they lose all they had, including family members, and there is devastation for hundreds of miles as far as they can experience. Even when temporary shelters are provided, mental health is badly affected. People who experience trauma and witness terrible things often suffer posttraumatic stress disorder. Among the displaced, unemployed, homeless, landless, and lost, abuse of alcohol and drugs increases; and with families packed into trailers or simply tents, domestic violence and suicide increase. Mass displacements of thousands or even millions of people into other countries where they are not wanted or add to existing resource and health issues are likely to lead to higher levels of violence, anger, and war.

The displacement of populations because of rising sea levels is the extreme, even if realistic, scenario. It is limited to specific, identifiable areas. The impact of global warming and climate change on food insecurity is much more pervasive and insidious. Not only will changed rain patterns cause droughts, floods, and storms, but frost and thaw will become erratic and difficult to predict. Warm winters will draw fruit trees out in February, when there are no insects to pollinate them. Insects and caterpillars will hatch out before bird migrations arrive for the nestlings to eat them. The biological rhythms of millennia with which agriculture has evolved will be disrupted. Not only will there be local crop failures of many kinds and widespread malnutrition, but systems of food trade, marketing, and transportation will need to raise their flexibility to barely imaginable levels.

Whereas increases in these nutritional, epidemic, and hazardous risks to human health are clearly based in the biology of climatic effects, their impact is almost entirely a matter of human agency and organization—the behavior vertex of the triangle. Changes will not occur uniformly over all the globe or to all people in any place. Low coastal areas are especially at risk. The poor are everywhere the most vulnerable. They have the least education and knowledge (human capital), the least social contacts and political power (social capital), and the poorest current health (reserve). Where will new physical, economic, and health care infrastructure go, and who will be served by it? Whether *Plasmodium falciparum* expands its territory, as has so often been predicted, depends more on success in distribution of insecticide-impregnated bed nets than on 1° of warming (Chaves, Kaneko, Taleo, Pascual, & Wilson, 2008). As Fothergill (1998) points out, men and women have different degrees of exposure in all phases of disaster, including risk, warning, preparation, response, and recovery. Some relief efforts have effectively targeted women with relief kits, building grants, and control over resources (e.g., after the 1999 cyclone in Orissa, India, and the 2004 tsunami in Thailand). The elderly, the very young, the poor, and those with disabilities are everywhere the most vulnerable to the health effects of climate change.

Remember the fundamental premise of Dubos and Audy: Health is adaptability. This is true not only of individuals, but of community, social, and economic systems; political organizations; and world organs of cooperation and support (the strengthening of which is the target of one of the United Nations' millennial goals). Minimizing the deleterious health impacts of climate warming needs to be a project in political ecology.

CONCLUSION

Climate change is going to affect human health in both grand, sweeping patterns and displacements, and in local and particular ways. For example, vectored diseases (especially those vectored by mosquitoes) on a warmer earth will extend their distribution a few degrees further in latitude and a couple thousand feet higher in altitude. Across large areas where they already exist, however, they do no harm because of local standards of living, education, health care systems, and vigorous cultural buffers. Expansion of vectors' extent even by miles in such areas will not matter. Since impacts are species-specific, broad patterns of impact may not reflect the local. Some mosquitoes (including some species of *Anopheles*), for example, prefer to breed in brackish water, although most prefer fresh water. Increased rainfall and runoff will freshen some brackish areas, even estuaries. Raised sea levels, including storm surge overwash, will salinate deltaic lakes, groundwater, and wells. Changes in local vector species will be locally affected to an extent little appreciated.

Climate change, of course, will have an impact on more than disease ecology. It will affect all the biometeorological factors described early in this chapter as well. There is a vast store of folklore on weather sensitivity. Old war wounds, creaking joints, and restlessness are notorious for enabling people to predict changes in the

weather. Mood swings, migraine headaches, asthma, rheumatism, and heart problems are also popularly associated with various weather conditions. As we have seen, it is certainly possible that popular wisdom is not just myth. Changes in barometric pressure, temperature, ionization, sunlight, and other weather elements may plausibly affect human biochemistry.

There is remarkably little proof, however, of the effects of weather on health, aside from temperature extremes. Conditions and data vary with age, season, socioeconomic status, and cultural buffering (e.g., through the use of air heating and cooling). The present status of research is such that little is known of the relevant periodicities. Variations in temperature, pressure, or other weather elements and their rates of change occur over seconds, minutes, hours, days, months, and years. Finding the pathways for influences is difficult because of the problem of specifying the relevant periodicity (should we correlate heart attacks with temperature change in a day or over a number of days? with passage of a front or with the number of fronts that passed within a limited time?) and our ignorance of threshold or relevant range (is a temperature change from 80° to 60°F the same in effect as one from 40° to 20°F? can winter in countries whose mean daily temperatures never get below freezing be compared with that in countries whose temperatures stay below freezing for weeks?). The general scientific law of initial value, applied to biometeorological studies, would state that the magnitude of any weather influence is going to depend on both the level from which it started and the level of the physiology it is affecting. For example, if a certain influence raises blood pressure, the health result depends to a large degree on the initial level of the blood pressure. At what point, then, can we detect effects? The tremendous range of individual variation in physiology further complicates study. In sum, biometeorological changes are believed to have effects on health, and plausible pathways for many influences exist. Most of the research done so far has been accomplished in Europe, with its generally moderate marine climate. The climates of North America and the seasonality of its latitudes vary more substantially and offer many opportunities for research. Although some interventions involve geographic planning, much of this research is very “medical” rather than “public health” and involves mainly geographers as climatologists.

The impacts of climate change on health are going to be overwhelmingly upon food security, nutrition, and the availability of usable water (for drinking, sanitation, and hygiene), for cities and the countryside both. We already know who the vulnerable are: the poorest, the children, the elderly, those living in low-lying areas. We can already develop geographic information system (GIS) applications that map their location block by block, mile by mile, pretty much anywhere in the world. Planning for disaster relief—actually coping with the aftermath—is proving the integrative powers of GIS. ESRI, the creator of much of the major GIS software, was involved with a response to the 2004 tsunami in Indonesia that connected scientists, international aid organizations, and other nongovernmental organizations, government workers, local citizens, and policy makers as nothing before had done. Several examples of the use of GIS in disaster and health planning are posted on the ESRI website (www.esri.com; search “disaster” for .pdf files). North Carolina has developed a model public health program of rapid needs assessment that involves

the use of linked hand-held Arc GIS pads; these pads have been used for hurricane response within 72 hours. Much additional research and development work is needed in this area of medical geography and disaster.

What new technologies, organizations, and policies can develop and deliver which new *cultural* buffers? How can these be optimally located and designed in space? How can local populations, not satellites, be involved in information gathering, processing, construction, and implementation of effective helpful policies? These are research questions and applications for geography as social science.

REFERENCES

- Arcury, T. A., Williams, B. J., & Kryscio, R. J. (1990). Birth seasonality in a rural U.S. county, 1911–1979. *American Journal of Human Biology*, 2, 675–689.
- Barinaga, M. (1999). The clock plot thickens. *Science*, 284, 421–422.
- Bovallius, A., Roffey, R., & Henningson, E. (1978). Long range air transmission of bacteria. *Applied and Environmental Microbiology*, 35, 1231–1232.
- Buor, D. (2004). Water needs and women's health in the Kumasi metropolitan area, Ghana. *Health and Place*, 10, 85–103.
- Campbell, D. E., & Beets, J. L. (1979). The relationship of climatological variables to selected vital statistics. *International Journal of Biometeorology*, 23, 107–114.
- Cech, I., Youngs, K., Smolensky, M. H., & Sargent, F. (1972). Day-to-day and seasonal fluctuations of urban mortality in Houston, Texas. *Biometeorology*, 23, 77–87.
- Chaves, L. F., Kaneko, A., Taleo, G., Pascual, M., & Wilson, M. L. (2008). Malaria transmission pattern resilience to climatic variability is mediated by insecticide-treated nets. *Malaria Journal*, 7, Article No. 100.
- Coleman, J. S. M. (2008). Atmospheric circulation patterns associated with mortality changes during transitional seasons. *Professional Geographer*, 60(2), 190–206.
- Committee on Climate, Ecosystems, Infectious Diseases, and Human Health. (2001). *Under the weather: Climate, ecosystems, and infectious disease*. Washington, DC: National Research Council.
- Confalonieri, U., Menne, B. R., Akhtar, R., Ebi, R. K. L., Hauengue, M., Kovats, R. S., et al. (2007). Human health. In M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden, & C. E. Hanson (Eds.), *Climate change 2007: Impacts, adaptation and vulnerability* (pp. 391–431). Cambridge, UK: Cambridge University Press.
- Cowgill, U. M. (1966). Season of birth in man: Contemporary situation with special reference to Europe and the Southern Hemisphere. *Ecology*, 47, 614–623.
- Dalen, P. (1975). *Season of birth: A study of schizophrenia and other mental disorders*. Amsterdam: North-Holland.
- Driscoll, D. M. (1971). The relationship between weather and mortality in ten major metropolitan areas in the United States, 1962–1965. *International Journal of Biometeorology*, 15, 23–39.
- Driscoll, D. M. (1983). Human biometeorology in the 1970's. *International Journal of Environmental Studies*, 20, 137–147.
- Emch, M., Feldacker, C., Islam, M. S., & Ali, M. (2008). Seasonality of cholera from 1974 to 2005: A review of global patterns. *International Journal of Health Geographics*, 7(31), 1–33.
- Fellman, B. (1985). A clockwork gland. *Science*, 85, 76–81.
- Fothergill, A. (1998). The neglect of gender in disaster work: An overview of the literature. In E. Enarson & B. Morrow (Eds.), *The gendered terrain of disaster: Through women's eyes* (pp. 9–25). Westport, CT: Praeger.
- Greenberg, J. H., Bromberg, J., Reed, C. M., Gustafson, T. L., & Beauchamp, R. A. (1983). The epidemiology of heat-related deaths in Texas—1950, 1970–79, and 1980. *American Journal of Public Health*, 73, 805–807.

- Huntington, E. (1938). *Season of birth*. New York: Wiley.
- Hyslop, N. S. G. (1978). Observations on the survival of pathogens in water and air at ambient temperatures and relative humidity. In M. W. Loutit & J. A. R. Miles (Eds.), *Microbial ecology* (pp. 197–205). Berlin: Springer-Verlag.
- Kalkstein, L. S., & Davis, R. E. (1989). Weather and human mortality: An evaluation of demographic and interregional responses in the United States. *Annals of the Association of American Geographers*, 79, 44–64.
- Keller, D. A., & Nugent, R. P. (1983). Seasonal patterns in perinatal mortality and preterm delivery. *American Journal of Epidemiology*, 118, 689–698.
- Kevan, S. M., & Chapman, R. H. (1980). Variations in monthly death rates in Canada. In F. A. Barrett (Ed.), *Canadian studies in medical geography* (pp. 67–77). Downsview, ON, Canada: York University, Department of Geography.
- Lowen, A. C., Mubareka, S., Steel, J., & Palese, P. (2007). Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathogens*, 3, 3151.
- McMichael, A. J. (2004). Climate change. In M. Ezzati, A. Lopez, A. Rodgers, & C. Murray (Eds.), *Comparative quantification of health risks: Global and regional burden of disease due to selected major risk factors* (Vol. 2, pp. 1543–1649). Geneva: World Health Organization.
- McMichael, A. J., Campbell-Lendrum, D., Corvalan, C., Ebi, K., Githeko, A., Scheraga, J., et al. (Eds.). (2003). *Climate change and human health: Risk and responses*. Geneva: World Health Organization.
- Moeller, D. W. (1997). *Environmental health* (rev. ed.). Cambridge, MA: Harvard University Press.
- National Center for Health Statistics. (1998). *Births: Final data for 1997*. Bethesda, MD: Author.
- Perkins, S. (2004). Dead heat: The health consequences of global warming could be many. *Science News*, 166, 10–12.
- Persinger, M. A. (1980). *The weather matrix and human behavior*. New York: Praeger.
- Rodgers, J. L., & Udry, J. R. (1988). The season-of-birth paradox. *Social Biology*, 35, 171–185.
- Rosenberg, H. M. (1966). Seasonal variation of births in the United States, 1933–63. In *Vital and Health Statistics* (Series 21, No. 9, pp. 1–42). Washington, DC: U.S. Department of Health, Education and Welfare.
- Sakamoto-Momiyama, M. (1977). *Seasonality in human mortality*. Tokyo: University of Tokyo Press.
- Sakamoto-Momiyama, M., & Katayama, K. (1967). A medical-climatological study in the seasonal variations of mortality in the United States of America. *Papers in Meteorology and Geophysics*, 18, 209–232.
- Sakamoto-Momiyama, M., & Katayama, K. (1971). Statistical analysis of seasonal variation in mortality. *Journal of the Meteorological Society of Japan*, 49, 494–509.
- Somers, D. E., Devlin, P. F., & Kay, S. A. (1998). Phytochromes and cryptochromes in the entrainment of the *Arabidopsis* circadian clock. *Science*, 282, 1488–1490.
- States, S. J. (1976). Weather and death in Birmingham, Alabama. *Environmental Research*, 12, 340–354.
- States, S. J. (1977). Weather and deaths in Pittsburgh, Pennsylvania: A comparison with Birmingham, Alabama. *International Journal of Biometeorology*, 21, 7–15.
- Takahashi, J. S., & Zatz, M. (1978). Regulation of circadian rhythmicity. *Science*, 217, 1104–1110.
- Thresher, R. J., Vitaterna, M. H., Miyamoto, Y., Kazantsev, A., Hsu, D. S., Petit, C., et al. (1998). Role of mouse cryptochrome blue-light photoreceptor in circadian photoresponses. *Science*, 282, 1490–1494.
- United Nations Development Program (UNDP). (2008). *Fighting climate change: Human solidarity in a divided world. Human development report 2007/2008*. New York: United Nations.
- West, R. R., & Lowe, C. R. (1976). Mortality for ischaemic heart disease: Inter-town variation and its association with climate in England and Wales. *International Journal of Epidemiology*, 5, 195–201.
- Yang, G. J., Vounatsou, P., Zhou, X. N., Tanner, M., & Utzinger, J. (2005). A potential impact of cli-

- mate change and water resource development on the transmission of *Schistosoma japonicum* in China. *Parasitologia*, 47(1), 127–134.
- Zhou, X. N., Yang, G. J., Yang, K., Wang, X. H., Hong, Q. B., Sun, L. P., et al. (2008). Potential impact of climate change on schistosomiasis transmission in China. *American Journal of Tropical Medicine and Hygiene*, 78(2), 188–194.

FURTHER READING

- Epstein, P. R. (1997). Climate, ecology, and human health. *Consequences*, 3, 3–19.
- Folk, G. E., Jr. (1974). *Textbook of environmental physiology*. Philadelphia: Lea & Febiger.
- Jusatz, H. J. (1966). The importance of biometeorological and geomedical aspects in human ecology. *International Journal of Biometeorology*, 10, 323–334.
- Kavaler, L. (1981). *A matter of degree: Heat, life, and death*. New York: Harper & Row.
- Terjung, W. H. (1966). Physiologic climates of the conterminous United States: A bioclimatic classification based on man. *Annals of the Association of American Geographers*, 56, 141–179.
- Tromp, S. W. (1980). *Biometeorology*. London: Heyden.
- Warren, J. A., Berner, J. E., & Curtis, T. (2005). Climate change and human health: Infrastructure impacts to small remote communities in the north. *International Journal of Circumpolar Health*, 64(5), 487–497.

VIGNETTE 6.1. Monthly Indexes

The study of seasonality of events is complex because the pattern of the whole year has to be compared (whether as months or as weeks). Various, often complex statistical analyses (such as correlation–regression, Box–Jenkins ARIMA modeling, and spectral analysis) have been used to study seasonality. Beware: It is easy to summarize and compare patterns in a careless and invalid manner. Whatever methods of analysis are used, three methods have proved essential to preparing the pattern for analysis or comparison. These are *smoothing*, *adjustment*, and *indexing*.

Smoothing serves to tone down random fluctuation and other “noise,” which can be especially important when total numbers are small. A pattern that was clouded and confused in the raw numbers often shines through smoothed data. There are several methods of smoothing. The simplest and most common is to create a running average. For example, 1980, 1981, and 1982 yield a mean value used for 1981; 1981, 1982, and 1983 yield a mean value for 1982; and so on. Smoothing in this manner is more appropriate for yearly running totals, however, than for months because of the variation in the number of days in a month. When smoothing monthly data in this manner, one ends up with periods of 92 days for one average and 89 days for another. The irregular pattern of month length increases such period differences rather than smoothes them. Another method of smoothing (illustrated in Vignette 6.2, Figure 2) involves trigonometric functions such as sine waves (in harmonic analysis) to identify the periodicity and strength of variation.

The most important step is to adjust for the number of days in a month. Comparing deaths in February and August will not do. There are different types of adjustment formulas, but they all do essentially the following:

$$A_i = \frac{M_i}{D_i} \times \frac{365}{12}$$

where i is the month; A is the monthly adjusted number; M is the number of cases, deaths, or frequency of whatever is being studied; and D is the number of days. The average number of days in a month, $365/12$, is of course a constant, 30.42, but remember that for leap year you need $366/12$. The adjusted monthly deaths (or whatever) may be used to create monthly adjusted death rates by dividing by the total population. The procedure also works for age categories, and can be used to create age-adjusted monthly rates.

Indexing portrays each month's adjusted numbers as above or below the average of what would be, under conditions of even distribution throughout the year, the monthly average after adjustment for the number of days (see Figure 6.5). A ratio of the adjusted "observed" numbers to the monthly average "expected" without seasonality is created by dividing the adjusted numbers by the monthly average, $\Sigma M/12$ (the 12 is multiplied by 100 to index it to 100).

$$\text{Index}_i = A_i \times \frac{\Sigma M}{1,200}$$

Vignette 6.1, Table 1, illustrates the results of monthly adjusting and indexing for cholera outbreaks in four countries at different latitudes (Emch, Feldacker, Islam, & Ali, 2008).

VIGNETTE 6.1, TABLE 1. Monthly Adjusting and Indexing

		Number of cholera outbreaks											
Month	Days	Kenya			Malaysia			Iran			Romania		
		M_i	A_i	Index	M_i	A_i	Index	M_i	A_i	Index	M_i	A_i	Index
January	31	7	6.87	104	13	12.76	103	6	5.89	102	0	0	0
February ^a	28	6	6.52	99	10	10.86	88	2	2.17	38	0	0	0
March	31	8	7.85	119	13	12.76	103	1	0.98	17	0	0	0
April	30	8	8.11	123	11	11.15	90	1	1.01	18	0	0	0
May	31	10	9.81	149	14	13.74	112	1	0.98	17	0	0	0
June	30	6	6.08	92	14	14.20	115	2	2.03	35	0	0	0
July	31	10	9.81	149	15	14.72	119	4	3.93	70	1	0.98	78
August	31	5	4.91	75	12	11.78	96	10	9.81	171	4	3.93	314
September	30	3	3.04	61	13	13.18	107	11	11.15	194	4	4.06	324
October	31	6	5.89	90	11	10.79	88	12	11.78	205	5	4.91	393
November	30	5	5.06	77	10	10.14	82	11	11.15	194	1	1.01	8
December	31	5	4.91	75	12	11.78	96	8	7.85	137	0	0	0
Total	365	79		100	148		100	69		100	15		100
Latitude		0°N			4°N			32°N			46°N		

Note. Data from Emch, Feldacker, Islam, and Ali (2008).

^aBeware of leap years.

VIGNETTE 6.2. Seasonality of Birth

The presence of seasonal patterns in the occurrence of human birth is ubiquitous on earth. The patterns of northern Europe are quite different from those of tropical Africa. In the United States the pattern has been consistent throughout the 20th century: a minor peak of births in February–March (spring conception), a great concentration of births in August–September (early winter conception), and a deep trough of births in April–May (hot summer conception) (see Figure 6.5). There has been spatial variation in the exact timing and sharpness of this pattern between the North and South, or between depression and growth decades, but the overall pattern has held for at least half a century—despite the facts that the U.S. population has become overwhelmingly contraceptive-using, has seen massive migration from South to North, has become three-quarters metropolitan instead of rural, and has seen the origins of immigration shift from eastern and southern Europe to Asia and Middle America. Only in the last 25 years has the pattern of seasonality begun to alter (Figure 6.5).

The fact that there is a seasonality to human birth that has a spatial pattern; varies in associations by scale of analysis; and reflects the complex interactions of biology, environment, socioeconomic and demographic structure, and behavior poses a challenge to interpretation and explanation in human ecology and medical geography. Those researching and publishing in biometeorology, anthropology, biological reproduction and fecundity, demographic sociology, or epidemiology do not read each other; they lack geographic integration. This vignette summarizes Meade's research on the subject.

METHODS

Since the efforts of geographer Ellsworth Huntington (1938), there has been remarkably little research on the etiology of seasonality of birth. Although his statistical methodology would not pass in an undergraduate term paper today (he did not even adjust for the number of days in a month; see Vignette 6.1), he asked some profound questions about the seasonal patterns he documented with data collected from all over the world. He ran latitudinal transects (straight lines along which data are collected) through the United States and Japan to examine climatic change in different cultures. In the former Soviet Union he compared the same ethnic groups in different locales and different ethnic groups in the same locale. He was interested in the effect of climate on people, and was the most famous and accomplished proponent of a school of thought known today as *environmental determinism*. Probably because of the subsequent discrediting of environmental determinism, geographers have ever since ignored the questions he raised.

The main method used to draw out the patterns was *harmonic analysis*, which has been used by climatologists to analyze such things as weekly or monthly precipitation. The most common way that medical researchers (especially in Britain) address seasonality of death, birth, or congenital malformation is to divide the year into four groupings of 3-month seasons for testing with chi-square. In a country as large as the

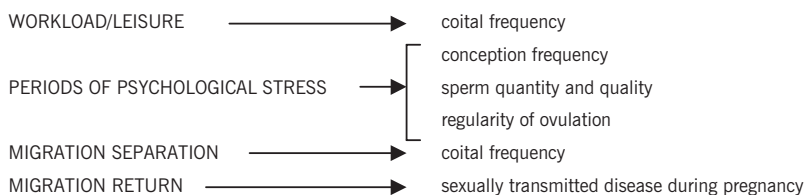
United States, however, these groupings are too arbitrary. March in Minnesota is not the same as March in California or Florida, for example. Given the bimodal pattern, furthermore, even regression by month would miss the pattern of the whole cycle. Imagine the monthly births as a snaking rope. If you push down the monthly births in February, those births are not uniformly redistributed. They may decrease the April trough or raise the September peak, or not. The purpose of harmonic analysis is to express any curve, however complex or irregular, as the algebraic sum of a series of simple sine curves. The first harmonic is one sine curve fitted to the monthly data, so that the sum of the squares of the difference between them is minimized. This adjusts the height of the curve (amplitude) and shifts the curve (horizontally, or through the calendar) until the height of the curve best fits the maxima in the data distribution, and the point (degree of a circle, day of the year) at which this occurs gives the phase angle of best fit. The second harmonic then fits two sine waves, which best catch a bimodal pattern. Altogether, six harmonics are considered to exhaust variation. When this was done to 1980 births in a transect from Connecticut to Florida, the amplitude of annual seasonality consistently decreased, and the phase angle shifted earlier in the year. This matched closely what Huntington had found.

MODELS OF COMPLEXITY

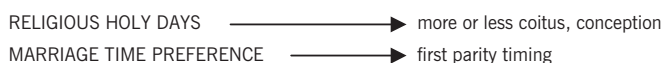
The first dimension of complexity is that influences affecting what is a long process are usually measured with only a single outcome, birth, for which data are collected. The first influences are on genesis and storage of the egg and on genesis of the sperm and maturation in the epididymis. Then, in sequence, influences act upon conception; egg division; zygote growth and movement through the fallopian tube; implantation in the uterus; maturation (with the risks of fetal loss in the first two trimesters and miscarriage in the third) as part of the mother's body, sharing most of her nutritional and immunological health and more; birth (stillbirth or live); and then variation in birth-weight, gender (sex ratio), congenital disease, neonatal death, and survival. Various influences of infection, diet, emotional stress, altitude change, activity, and much more act on this whole process. More than a third of the fertilized eggs will not survive to become 12-week fetuses (many of them because of errors in genetic replication; a few because of an incompatibility with their mothers' bodies that cannot be walled off). Yet, from the single measure of day of live birth, seasonal factors are implicated and analyzed through various disciplinary perspectives. The triangle of human ecology can help integrate them and focus the points of difference (Vignette 6.2, Figure 1).

Environmentally mediated sociocultural effects have received the most attention. Geographers and anthropologists working especially in developing countries with wet-dry climatic seasonality have noted the importance of separation for wage labor during the dry, nonfarming season. Alternatively, the religious and family celebrations that are usually held after harvest during the dry season provide leisure and raised libido to encourage conception then—an effect perhaps enhanced by the less frequent coitus during the intense labor periods of plowing, transplanting, or harvesting with

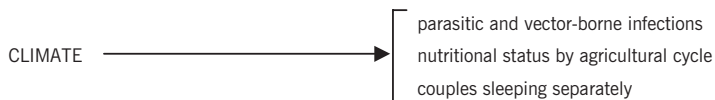
ENVIRONMENTALLY MEDIATED SOCIOCULTURAL EFFECTS



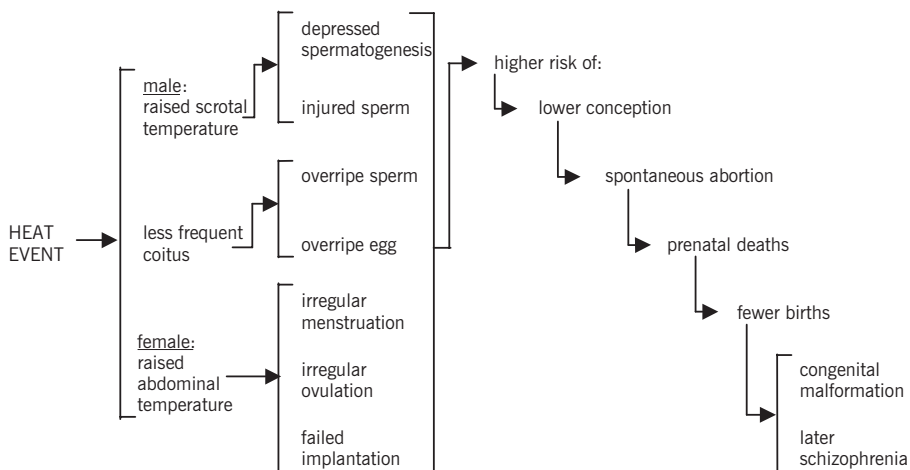
DIRECT CULTURAL EFFECTS



SOCIOCULTURALLY MEDIATED CLIMATIC EFFECTS



HEAT



VIGNETTE 6.2, FIGURE 1. Models of seasonality of birth.

their supposed exhaustion. The seasonal cycle of school year and vacation in the United States would fit this model.

An argument can be made for direct cultural effects over environment. It is clear from births that Ramadan, the Muslim month of fasting during daylight and self-denial of pleasures in general, is a period of low conception. Three religious/ethnic groups in Malaysia—Muslim Malays, Buddhist Chinese, and Christian Indians—all share the same uniformly hot equatorial environment, but have different birth periodicities that reflect Ramadan and Chinese New Year (as well as low births during astrological bad-luck years). The most common reaction when our students hear of the September peak of births in the United States is a giggle of “Christmas!”

Seasonal climatic patterns directly affect when mosquitoes and other vectors are active, when crops are grown and harvested, and when food is abundant or in short supply and low in nutrition. During the hot season in many places, it is common for men to sleep in the courtyard or on the roof, in the open air (or in old New York City, on the fire escapes). People may be more exposed to disease vectors. Couples often are separated.

There is some literature on these patterns, but as the reader can see, the influences are often both highly localized and often contradictory. There is a universal phenomenon involved. At high latitudes, it would be surprising if light periodicity did not affect human fecundity (see the discussion of biological rhythms earlier in this chapter). Animal breeding suggests, however, that the strongest influence is heat—a hypothesis not entertained by behaviorists in social science.

The scrotum is not enclosed in the safety of the abdominal cavity because it must be a little cooler. Experiments on monkeys, soldiers, and other subjects has clearly demonstrated the capacity of heat to lower sperm count and even totally destroy spermatogenesis. Indeed, sperm counts done in vasectomy clinics are lower in southern U.S. cities in the summer. Heat can affect every step in the process of birth, from menstrual cycle to implantation, as well as maternal activity pattern (including exposure to hazards and access to medical care).

TESTING HEAT

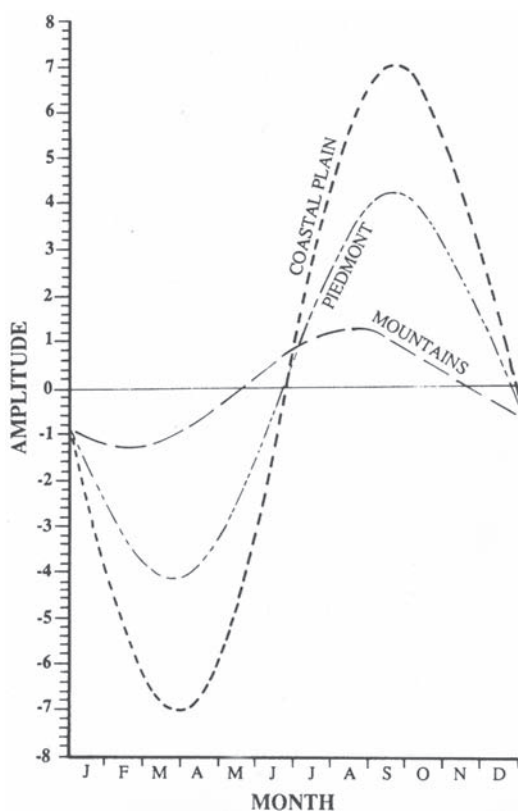
The trough of births in April–May in the U.S. pattern corresponds with conceptions in July and August. This is a national pattern. If heat is involved in lowering conception, variation across the nation and within states should show spatial variation associated with climatic variation. It is not known, however, whether the absolute temperature experienced (e.g., while jogging) or the length of hot weather to which people are exposed matters. What lags between temperature and birth, or what periods of temperature, should be compared; what extremes or averages?

In Vignette 6.2, Figure 2, birth indexes for the mountain, piedmont, and coastal plain regions of North Carolina have been converted into the first harmonic curves. These are certainly the patterns one would expect if the higher temperatures of the coastal plain, intermediate temperatures of the piedmont, and lower temperatures of

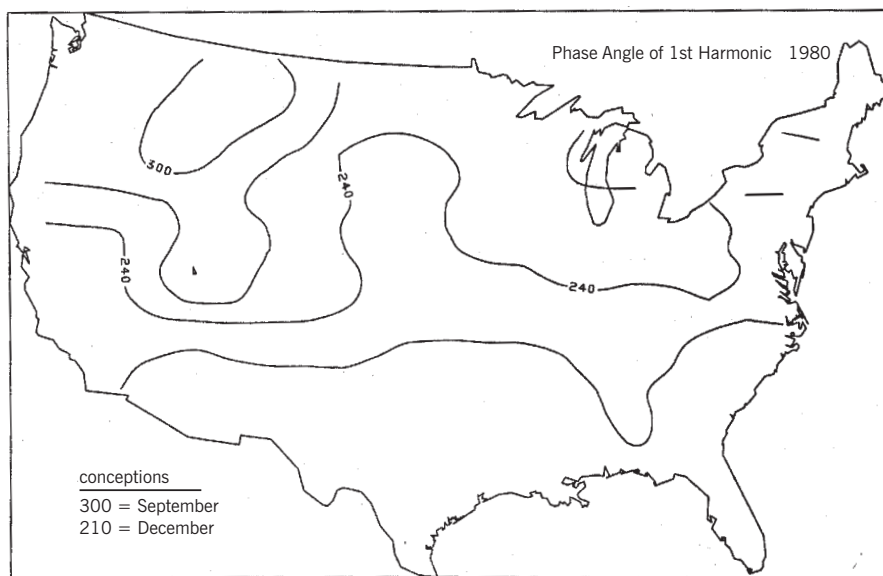
the mountains made any difference to conception. Temperature data averaged over weeks and regions were inversely correlated with births 38 weeks later. When the scale of analysis was shifted to states, the first harmonic curve was fitted to 12 months of births for each of the contiguous 48 states in 1980. The day of the year when the curve reached its peak amplitude, simply converted from the phase angle, was mapped with SURFER software (see Vignette 6.2, Figure 3). The resulting isoline map of human conception in the United States for 1980 shifts from early September in the North to December in the Deep South, with allowance for the mountains. It shows such strong climatic association that it resembles the U.S. Department of Agriculture's map of planting zones.

ASSOCIATIONS AND DEVIATIONS

The patterns of seasonality just discussed could probably not be mapped today. "Noise" and deviation began to become pronounced at the county level in the late 1980s. Within



VIGNETTE 6.2, FIGURE 2. Births by region in North Carolina, 1969–1971. The sine waves of the first harmonics show a greater amplitude of seasonality, but similar timing, that might be associated with heat as an etiological factor.



VIGNETTE 6.2, FIGURE 3. Map of U.S. seasonality of birth, 1980. The phase angle, changed to day of year, of the (first harmonic) annual pattern's peak has been mapped as an isoline map of U.S. conception, showing the time shift from south to north.

North Carolina, sociocultural factors were addressed at the county level and change was measured by comparing harmonic amplitudes for years around 1970, 1980, and 1990. The (annual) amplitude of the first harmonic was always highest for rural areas, but became increasingly important for urban and dramatically more important for metropolitan regions. Harmonic decomposition (i.e., using two, three, or four curves) looks at change in pattern and examines such factors as parity, legitimacy, race, occupation, and other information recorded on birth certificates. Across almost all groups, annual seasonality has increased and bimodal seasonality decreased: Annual seasonality has increased for married women, for example, but births out of wedlock have remained bimodal. Some relations were surprising: The more educated a woman, the higher the annual amplitude, for example. As can be seen in Figure 6.5, the bimodal pattern has partly disappeared.

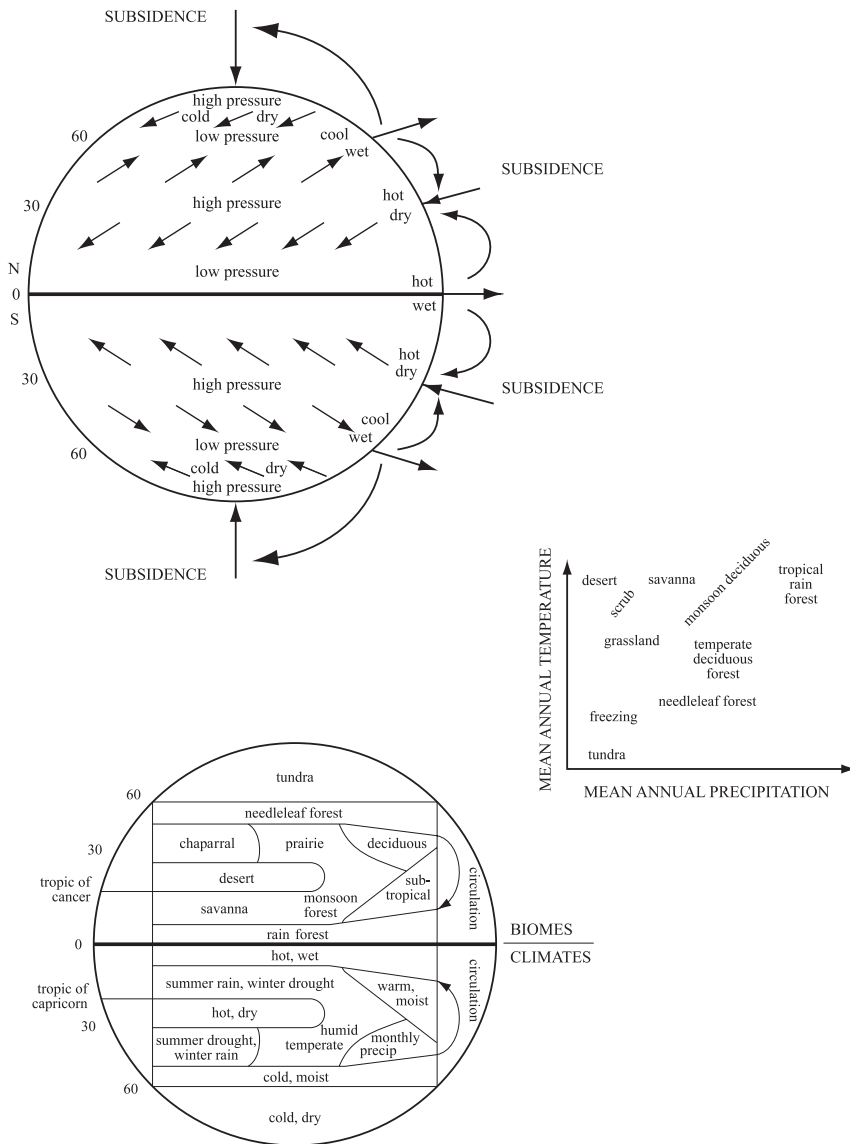
Both a bimodal seasonality of birth and a regular spatial pattern have persisted for decades across this land, but both seem to be declining in recent years. The diffusion of air conditioning may be responsible for ending the April–May trough of births associated with low summer conception, even as central heating has changed the pattern of mortality. This has been suggested by some regional-level studies. Almost universal contraception offers a likely hypothesis for the disappearance of the minor February peak, associated with May conceptions and perhaps the sap of spring. What remains most remarkable is the strong, and stronger, pattern of annual seasonality of birth (perhaps now July–September), which becomes stronger as the population becomes more urban and metropolitan, more educated, and even more contraceptive-using.

VIGNETTE 6.3. Physical Zonation of Climates and Biomes

The diurnal and annual variations in intensity of solar radiation, and the amount and timing of precipitation, temperature, and humidity, have been described, classified, and analyzed in dozens of ways. The variations are complicated by the tilt of the earth's axis, the physiography of mountains and depressions, the contrasts of land and sea, and the size of continents. The broad zones of vegetation, or *biomes*, that result are also zones of agricultural cropping patterns and of arthropod habitat. The schema in Vignette 6.3, Figure 1, of a prototype continent is simplified but provides a general framework for considering the study of health through landscape epidemiology and biometeorology.

Seasons occur as the tilt of the earth's axis alternately exposes the Northern and Southern Hemispheres to more direct rays from the sun. The vertical rays are limited to the latitudes between the Tropic of Cancer and the Tropic of Capricorn, with the highest annual amounts being received at the equator and the least at the poles. Because hot air rises and cold air sinks, the equatorial area is a zone of low pressure as the air rises, and the poles are zones of cold, subsiding, dry air (see Figure 1). Because cold air can contain less water vapor than an equal volume of warm air, water condenses as the air rises at the equator and cools at higher elevations, producing continually high levels of rainfall over the year. Similarly, the subsiding air at the poles is warmed as it approaches the earth's surface, and because its capacity to hold water vapor is increased, there is little precipitation. Due to processes of atmospheric circulation, air also subsides at about 30° north and south latitudes, forming a zone of little precipitation at each latitude. Because of the lack of cloud cover, this zone receives more radiation at the surface than does the equatorial zone. As subsiding air at each 30° latitude and subsiding air at the poles move over the earth's surface and converge, air rises along a broad frontal zone in the middle latitudes and creates a wet but cooler zone. There is less precipitation in this midlatitude zone than in the equatorial zone, but because there are also less evaporation and less transpiration from plants, there is high availability of water.

Different vegetational zones, known as biomes, extend over large areas. The biomes are adapted to differences in temperature and precipitation, as well as to seasonal variation. The complex and luxuriant rain forest needs continually wet conditions and high solar radiation, whereas lichen and other tundra vegetation can grow in cold, dry conditions with a short season of long days of solar radiation. As Figure 1 indicates, from the poles to the equator there is a transition of forest types: from needleleaf (coniferous, taiga, boreal) to temperate deciduous forest (which loses its leaves in the cold season), to subtropical forests (which include evergreen plants such as magnolias and palms), to the tropical rain forest. Desert may be hot or cold, but it is dry; types of desert vegetation such as creosote bush and cactus merge into types of grassland or evergreen scrub as water increases. The savanna biome is grassland (with scattered monsoon deciduous trees) that has adapted to conditions of summer rain and winter drought. At its dry margins the trees are sparse and the grass short, but as it extends toward the equator the trees increase in density until savanna merges into forest. *Chaparral* is the American name for the biome of evergreen brush that has adapted to the severe conditions of summer drought and winter rain. Grassland, known variously as *prairie*, *steppe*, *pampas*,



VIGNETTE 6.3, FIGURE 1. Physical zonation of climates and biomes.

and *veld*, extends from short, bunched grass on the arid margins to tall, dense grass on the woodland border.

Broad climatic patterns determine the general pattern of biome location. As can be seen from the prototype continent, the hot, wet equatorial zone coincides with the tropical rain forest biome, and the cold, dry polar zone with tundra. The subsiding zones of air at 30° latitudes North and South create two broad bands of desert, including the Sahara, Gobi (displaced northward by the Tibetan mountains), and North American deserts in the Northern Hemisphere, and the Kalahari, Australian, and Atacama des-

erts in the Southern. The rising-air, wet zone in each hemisphere is characterized by needleleaf vegetation at its colder latitudes and temperate deciduous at its warmer latitudes. As the climatic zones shift north and south with the sun's rays, they bring wet-dry seasonality to much of the tropics. The area between the equatorial wet and the 30° dry zone supports a savanna or monsoon deciduous vegetation. The area between the 30° subsiding zone and the midlatitude wet zone is characterized by chaparral. Finally, because of the rotation of the earth, air masses revolve and winds blow in ways that bring humid air from the sea into land on the eastern side of continents, but bring only overland, dry air to the western side. On the eastern side of continents, there is a transition from tundra to needleleaf to deciduous to subtropical to rain forest; on the western side of continents, the transition is from tundra to needleleaf to chaparral to desert to savanna to rain forest. Grassy biomes cover the transitional areas.

A last complication is introduced by altitude. The atmosphere cools with increasing altitude. Mountains, furthermore, form barriers so that surface air must go up on the windward side, causing precipitation, and come down on the leeward, forming a "rain shadow," or dry area. An altitudinal zonation of vegetation results that is very similar to the latitudinal zonation that was presented earlier. A tall mountain on the equator (as occurs in Kenya and Ecuador) has a vegetation zonation that proceeds from rain forest at its base to subtropical, deciduous, needleleaf, tundra, and ice on its windward side and from rain forest to savanna, thornbush scrub, needleleaf, and tundra on its leeward side. No matter in what climatic zone people are living, therefore, a great variety of vegetation, agricultural, and living conditions exist for human settlements.

The distribution of diseases transmitted by arthropods—ticks, mosquitoes, flies, and so forth—in different parts of the world can be partly understood from biome distribution. Certain nutritional deficiencies are associated with specific staple crops (see Chapter 2), and cultivation of these crops often follows environmental constraints of biome distribution. Conditions of hazard from air pollution (see Chapter 7) or solar radiation (this chapter) also largely coincide with biome distribution patterns. Understanding the location of major earth biomes can provide a useful framework for understanding the spatial patterns of the distribution of many health hazards.

CHAPTER 7

The Pollution Syndrome

Since the beginning of the Industrial Revolution, exponentially increasing kinds and amounts of metals, gases, and chemicals have been added to the air, water, and soil in earth's industrial regions. Many of the chemicals have never existed on earth before. Because biological processes to break them down have not yet evolved, some such chemicals persist for long periods. A major source of pressure on the environment today is the desire by newly industrializing countries (NICs) such as China for a lifestyle approximating that of developed nations. New transportation and economic activities are quickly consuming what remains of earth's fossil fuels. The purpose of this chapter is to survey these environmental processes and their spatial context as a background for research on diseases and health policies in developed countries and NICs. Dubos (1965) pointed out some time ago that humankind is now adapting, genetically and culturally, to the environments that humans have built. We spend most of our time in our homes and workplaces. In industrialized countries most people live in cities, but even in rural areas cultivated land and settlements create the environmental stimuli that surround people. Coping with these new stimuli is largely under the control of culture because genetically we humans as a species change very slowly. Dubos noted, for example, that there is no reason to think that our eardrums are any more able to withstand vibration than those of cave dwellers 200 generations ago. Yet noise levels have increased dramatically.

Normally audible sound consists of vibrations from 30 to 20,000 cycles per second. Loudness is measured in decibels (dB), a logarithmic scale based on sound pressure levels in which a doubling of the intensity of the sound is represented by an increase of 3 dB. Because the ear is more sensitive to sounds in the 4,000- to 6,000-cycle range, it takes less intensity to injure it in these ranges than at higher or lower frequencies. The "A scale" (dBA) was constructed to assess total noise by measuring four frequencies and adjusting for their different effects. Exposure to levels above 90 dBA can produce deafness. Legally, therefore, the maximum occupational

noise exposure for 8 hours in a day has been set at 90 dBA. Hearing loss, however, is thought to occur in susceptible individuals exposed to 75 dBA over an 8-hour daily exposure for a working lifetime, and in a considerable portion of those similarly exposed at 85 dBA. According to the World Resources Institute (1996), an estimated 100 million people in industrial nations are exposed to traffic noise in excess of 65 dbA.

Although this chapter focuses on chemicals and radiation, other serious health consequences, such as the effects of noise, are also associated with the environmental changes that accompany economic development. Few nations currently face more environmental degradation due to industrial development than China (Economy, 2004; World Resources Institute, 1994). China has changed more rapidly than any society in history, with an average annual growth rate in excess of 9% during the past two decades. Most of that nation's rivers, especially in urban areas, are seriously polluted. Dust and chemicals foul the air of many cities and cause widespread health problems. Chronic obstructive pulmonary disease (COPD)—which is linked to exposure to particulate matter, sulfur dioxide (SO₂), and cigarette smoke, among other factors—accounted for 26% of all deaths in China in 1988. Approximately 210 million people now have COPD worldwide. In 2005, more than 3 million people died of COPD, which was about 5% of all deaths globally that year (World Health Organization [WHO], 2008). Also, approximately 90% of deaths from COPD occur in low- and middle-income countries. Indoor exposure to emissions from poor-quality coal used for cooking and heating is also a major health risk, increasing the incidence of both pulmonary disease and stroke. In some areas of China, the uncontrolled irrigation of market crops with raw sewage may cause intestinal worm disease among children. There have been some reports of arsenic poisoning caused by drinking city water in Beijing. If the chemical and physical insults described in the discussion of Audy's definition of health (Chapter 2) were mapped, the surfaces of the industrialized countries would appear as jagged mountain ranges and stratospheric plateaus.

TOXIC HAZARDS OF NATURAL AND ECONOMIC ORIGINS

Toxic hazards to health and life are nothing new. In fact, they occur naturally in the environment. Poisonous and carcinogenic gases are emitted by swamps and volcanoes. Toxic chemicals are produced by parasitic fungi on rye, wheat, peanuts, and other crops, especially in wet years. One such fungal toxin that causes ergotism, for example, has been blamed by scientists for everything from the medieval Saint Vitus's dance to the Salem witch trials. Food plants, such as soybeans, cabbage, and wheat, produce an array of chemicals designed to protect themselves from fungi and arthropods; these chemicals can damage the liver, destroy red blood cells, block the absorption of protein or iodine, cause allergies, and generally poison livestock and humans. Nature produces many chlorine-containing chemicals (Abelson, 1994). The smoke of burning wood contains more than 100 organochlorine compounds, including the extremely toxic dioxin. Forest fires, domestic wood burning, and

slash-and-burn agriculture all create large amounts of these polychlorinated chemicals. Adding to these and other naturally occurring pollutants, however, humans are exponentially increasing pollutants of their own creation. Table 7.1 lists a few pollutants and health effects of current concern. Of greatest concern, perhaps, are the increasing numbers of chlorinated hydrocarbons and organophosphates, many of which are known to be carcinogenic or teratogenic (causing malformation of the fetus).

Many substances that are dangerous as pollutants are important economically. Lead is an important stabilizer, and is essential in many types of batteries. Polychlorinated biphenyls (PCBs) are noninflammable, have a high plasticizing ability, and have a high dielectric constant. They are therefore widely used in transformers and capacitors; as heat transfer and hydraulic fluids; as plasticizers in adhesives and sealants; and as anticorrosion coats for electric wires, lumber, and concrete. As PCBs leak, leach, and vaporize, however, they become air and water pollutants of great concern (and occasional accidental polluters of food and feed). They are stored in fatty tissue, pass along the food chain, and persist for long periods in the environment. In addition to disinfecting water, chlorine and its compounds are used in the manufacture of pharmaceuticals and in their content. The organophosphates have become so useful that the world's food supply relies heavily on them.

Lead offers an excellent example of how our cultural capacity to alter the environment is outpacing our biological ability to adapt. Lead is common in the rocks on the earth's surface and therefore in its waters. It usually occurs at low levels, and in some places it probably has always posed a health hazard. Lead is a systemic poison that interacts with a range of body chemicals. Mainly it interferes in blood formation by retarding the maturation of red blood cells in bone marrow, but it is implicated in chronic nerve disease and brain damage because it affects copper metabolism, among other things. Lead accumulates in the bones, where it replaces calcium. Normally about twice as much lead is derived from food as from either air or water, and about 75% is excreted in the feces, a little is passed in urine, and a little more than 10% is stored in the bones. Up to 50% of the lead inhaled reaches the blood, whereas less than 10% of ingested lead does. Treating for chronic lead poisoning is difficult because the process of *chelation*, by which a substance combines with the lead and causes it to be excreted from the body, can result in more lead being pulled out of the bone storage. This can precipitate an acute blood lead crisis that did not exist before treatment. Thus prevention has to be given priority over approaches that emphasize treatment.

OUTDOOR AIR POLLUTION

Air pollution may be loosely defined as the contamination of the atmosphere by a harmful level of toxic substances. The key words are "harmful level." The level of risk to any given individual is a product of both a substance's toxicity and the level of the individual's exposure. Virtually all substances are potentially harmful if present in sufficiently large quantities. The exact effects of air pollution on health

TABLE 7.1. Some Environmental Pollutants and Health Effects

Pollutant	Affected organs	Health effects	Source
Mercury	Brain, bowels, transplacental transmission	Minamata disease; liver and kidney disease; diarrhea; lack of coordination; numbness; convulsions; death	Chloralkali plants; pulp and paper processing; electrical industries; fungicides in water and food supply
Cadmium	Blood vessels; kidney	Hypertension; bone softening and fractures; kidney disease; cadmium emphysema	Mining many metals; electroplating; stabilizer for polyvinyl chloride; batteries, cigarettes, pigment
Chlorinated hydrocarbons	Fat tissue; liver	Hydrocarbon toxicity	Processing, storage and transfer of petroleum products; organic solvents; rubber, plastics, paints, lacquer; dry cleaning
Organophosphates	Nerve–muscle synapses	Dizziness; headache; muscular weakness; incoordination; liver and cardiovascular diseases; convulsions; bone marrow disease; blocks breakdown of acetylcholine, which transmits nerve impulses; accumulation leading to convulsions, blurred vision; diarrhea; stillbirth	Insecticide (DDT, lindane, dieldrin, aldrin, chlordane, toxaphene); hexachlorophene (shampoos, deodorants, insecticides, herbicides)
Polychlorinated biphenyls (PCBs)	Fat tissue; liver	Inhibition of cell growth and interference with enzymes; enhanced action of organophosphates; yusho disease; growth disturbance; fatigue; nausea; jaundice; diarrhea; cough; asthma; acne; loss of hair; numbness; nervous system disturbance; joint deformity at birth	Used as heat transfer media in transformers and capacitors; solvents in adhesives, sealants and anticorrosives; paints and rubber; ink; brake linings
Dusts of quartz, silica, carbon, asbestos, cobalt, and iron oxides	Respiratory interstitial tissue	Pneumoconiosis (scarring of lungs); silicosis; black lung	Mining; sandblasting; quarrying; pottery and ceramics; stone masonry
Hydrogen sulfide	Respiratory center in brain	Paralysis of respiration; consequent edema, hemorrhage, death	Oil wells and refineries; sulfur and protein; decay in sewers and mines
Fluoride	Bones; teeth	Binds to magnesium, manganese, and other metals to interfere with endocrine function and enzymes; damage to calcium metabolism and pituitary water balance; dental and skeletal fluorosis and osteomalacia if calcium intake inadequate	Food and water; aluminum and other smelting
Asbestos	Pleura and peritoneum	Mesothelioma, rapidly fatal once symptomatic	Mining; brake linings; fireproofing; talcum powder; cement; ceiling tiles; clothing
Beryllium	Lungs	Sarcoidosis	Metal alloys for heat stress and coal burning

Note. Data from the National Center for Environmental Health (1996).

are a function of the dose delivered to the receptor and the ability of the receptor to cope with the resultant stress. In humans, the stress experienced by a critical organ or receptor tissue from particle inhalation depends on the properties of the particles. The delivered dose is a function of the anatomical features of the receptor, as well as the manner of breathing, breathing rate, and the integrity of the body's defense systems. Some pollution-related diseases may take years or even decades to develop. It has proved difficult to connect specific types of air pollution with health hazards such as bronchitis or lung cancer, in part because of the difficulties of valid sampling and appropriate scale. Researchers have often had to resort to classifications of urban or rural, for example, as surrogate measures for air pollution. These classifications, unfortunately, bring with them many confounding factors.

Atmospheric pollution is caused by a variety of substances, some gaseous and some solid. Figure 7.1 summarizes the many different sources and types of pollutants and the chains of pollution, described in more detail below. We focus on a few pollutants for which the U.S. Environmental Protection Agency (EPA) has established national air quality standards: carbon monoxide (CO), lead, nitrogen dioxide (NO_2), ozone (O_3), particulate matter whose aerodynamic size is less than or equal to 10 micrometers (PM-10), and SO_2 . The EPA established limits for atmospheric concentrations of these pollutants in 1970, and these limits have been adjusted over time. Table 7.2 lists these pollutants and their quality standards as of 2008. According to the EPA, certain metropolitan regions in the United States have ambient levels of these criteria pollutants that exceed the established limits and thus may create health hazards where they are found (Figure 7.2). The EPA calculates the *air quality index*, which is a metric that ranges from 0 to 500. Anything over 100 is considered unhealthy for sensitive groups, and anything over 150 is considered unhealthy for anyone.

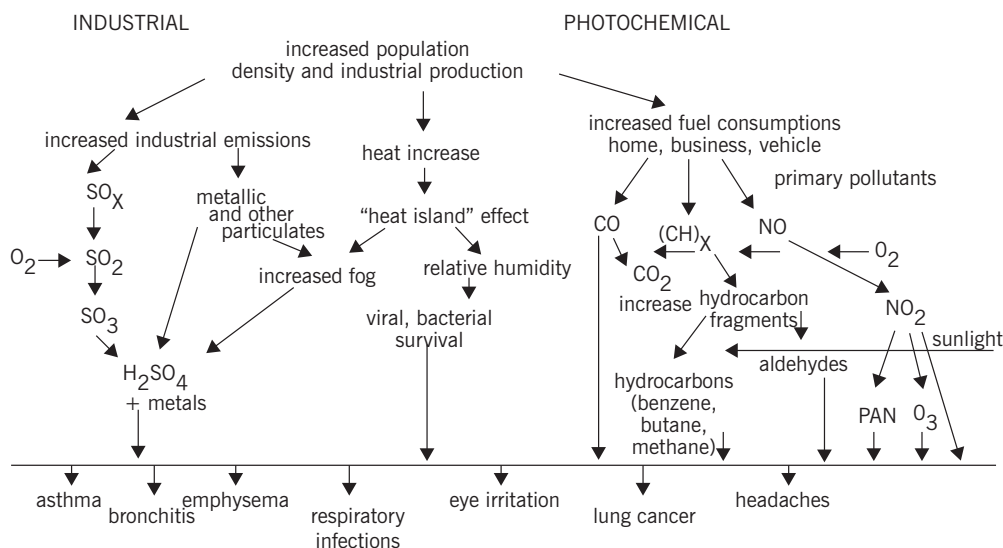


FIGURE 7.1. Chains of air pollution. PAN, peroxyacetyl nitrate.

TABLE 7.2. U.S. Environmental Protection Agency (EPA) Criteria for Air Pollutants and Health Effects

Pollutant	U.S. population potentially exposed (millions) ^a	Major sources ^b	Primary (health-related) standards ^b	Health effects
Carbon monoxide (CO)	89.1	Two-thirds of emissions from transportation sources	8-hour: 9 ppm (10 mg/m ³); 1-hour: 35 ppm (40 mg/m ³)	Reduces oxygen delivery to body's organs and tissues; major threat to sufferers of cardiovascular disease
Lead	N/A	Lead gasoline additives, nonferrous smelters, and battery plants	Maximum quarterly average: μ 1.5 g/m ³	May cause neurological impairment; fetuses and children may suffer from central nervous system damage
Nitrogen dioxide (NO ₂)	12.1	Two main sources: transportation and stationary fuel combustion from electric utilities and industrial boilers	Annual arithmetic mean: 0.053 ppm	Lowers resistance to respiratory infections, especially in children
Ozone (O ₃)	126.8	Transportation and industry	8-hour: 0.075 ppm; 1-hour: 0.12 ppm	Impairs lung function, especially among those with impaired respiratory systems
Particulates (diameter < 10 micrometers) Particulates (diameter < 2.5 micrometers)	54.6	Industry, power plants, transportation, construction, fires, agriculture, mining	24-hour average: 150 μ g/m ³ Annual arithmetic mean: 15.0 μ g/m ³ ; 24-hour average: 35 mg/m ³	Impairs breathing; aggravates existing respiratory and cardiovascular diseases; may damage lung tissue
Sulfur dioxide (SO ₂)	14.1	Coal and oil combustion; steel mills, refineries, pulp and paper mills; largest source: coal-burning electric power plants	Annual arithmetic mean: 0.030 ppm; 24-hour average: 0.14 ppm	Impairs breathing; aggravates respiratory and cardiovascular disease

Note. Ppm, parts per million; mg/m³, milligrams per cubic meter; μ g/m³, micrograms per cubic meter.

^aData from the American Lung Association (1996).

^bData from EPA (2008) (www.epa.gov/air/criteria.html).

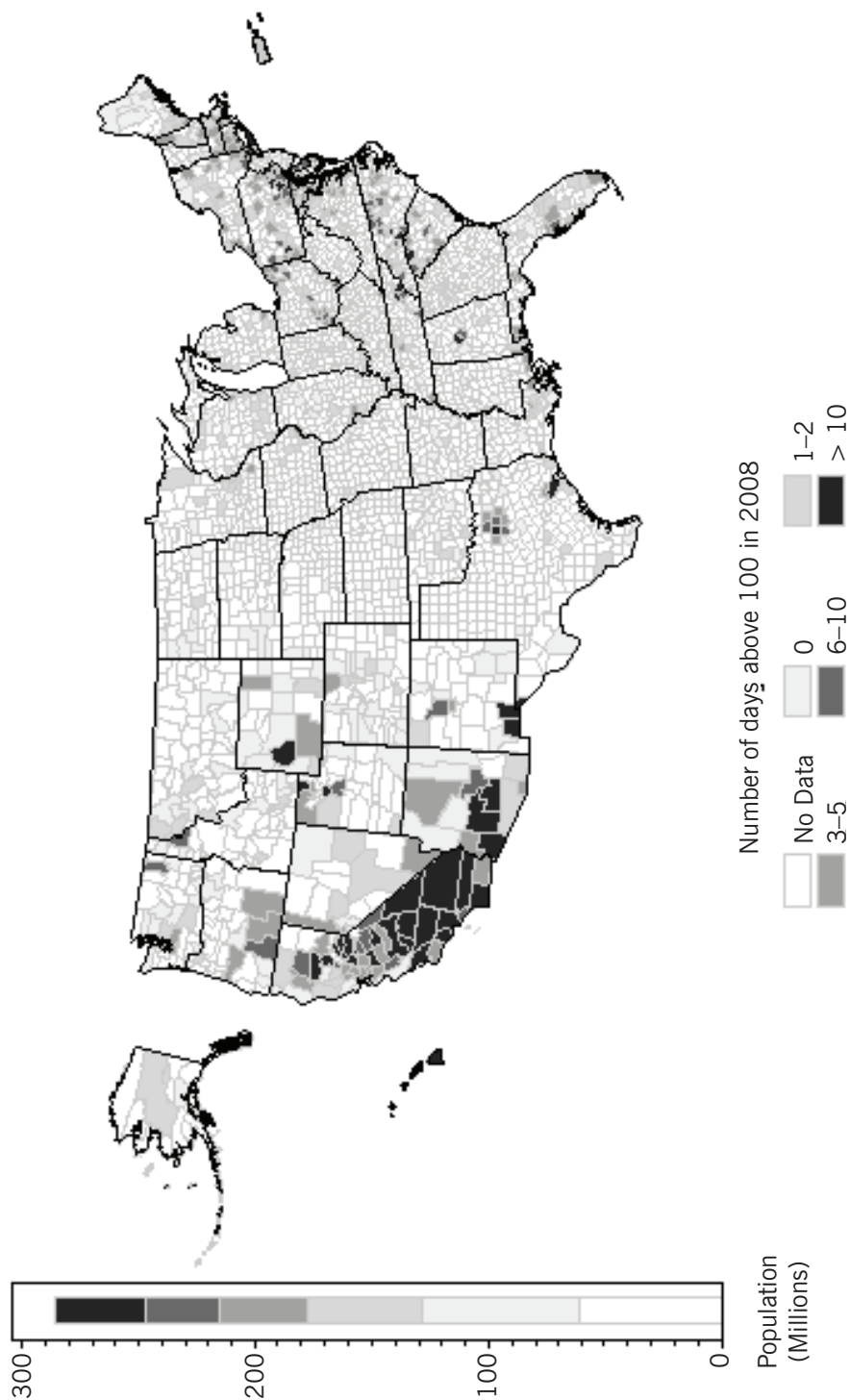


FIGURE 7.2. County air quality index map. The legend shows the number of days the air quality index was above 100 from January to October 2008 (100 is the threshold for the national air quality standard). From U.S. EPA (airnow.gov/index/cfm?action=static.qgi).

The major source of CO is automobile exhaust. Although it can be a regional phenomenon, CO pollution is often localized and occurs predominantly during cold weather. High concentrations may occur near areas of severe traffic congestion. In Tokyo police directing automobiles at street intersections take frequent breaks, during which they inhale pure oxygen to clear their lungs of pollutants. CO is the one pollutant that produces a change in human physiology that can be directly related to the concentration to which the subject was exposed. Death occurs in humans exposed to concentrations of about 1,000 parts per million, corresponding to blood levels of 60% blood carboxyhemoglobin. Because CO blocks the transport of oxygen in the bloodstream, and those with certain heart diseases require a high oxygen supply, it is not unreasonable to draw the same parallel as has been made with cigarette smoking and heart disease and to suggest that CO is a contributing factor (Koskela, Mutanen, Sorsa, & Klockars, 2000; Stern, Halperin, Hornung, Ringenberg, & McCammon, 1988; Strauss & Mainwaring, 1984).

Although lead is also toxic to adults and affects virtually all organ systems, adverse effects on cognitive development and behavior in children up to the age of 6 are of special concern. In 1992 the Centers for Disease Control (CDC) estimated that about 3 million, or 15% of the under-6 population, had blood lead levels exceeding 10 micrograms per deciliter of blood. This level is considered by the CDC as being high enough to trigger primary prevention campaigns in the local community. There are many sources of lead in the environment. Sewage sludge is commonly used as fertilizer in irrigation systems. Domestic sewage has almost everywhere been contaminated by industrial effluents, with the result that heavy metals such as lead and cadmium are present at relatively high levels. With repeated application of sewage, lead can quickly build up in the soil and thus pass into the human food chain.

The main source of lead exposure among lead-poisoned children in urban areas is lead-based paint in homes built before 1978. Although the use of lead-based paint in exterior and interior home surfaces was banned in 1978, the U.S. Department of Housing and Urban Development (1990) estimates that 74% (57 million) of all private housing units built before 1980 contain lead-based paint. Children may ingest lead directly from paint chips, but an important route of exposure is the normal mouthing of hands or objects such as toys, resulting in the ingestion of small amounts of lead-paint-contaminated house dust and soil.

Automobile emissions became the main focus of urban lead poisoning in the 1970s. Lead is used as a gasoline additive to improve the fuel's burning characteristics. It is emitted as a fine particulate in exhaust and settles in urban soils. There it is taken up by vegetation (including food crops in gardens), or children playing in or near streets are likely to ingest it as a result of breathing or handling food with soiled hands. The vertical (from the ground) and horizontal (from the road) gradients of lead levels in the air, as well as isotopic matching of local aerosol lead and local gasoline lead, provide irrefutable evidence of the importance of automobile emissions. The transportation network constitutes the broad geographical pattern of lead pollution. Levels of lead in the dirt of agricultural fields and in the dust of urban apartments decay sharply away from traffic. Since 1973, a reduction in tet-

raethyl lead in gasoline from over a gram per gallon to 0.1 grams per gallon was mandated in the United States. The incidence of elevated blood lead, particularly among inner-city minority (African American and Hispanic) children, has dropped dramatically as a result.

Hunter's (1977) research demonstrated that childhood lead poisoning was a "summer disease," due to the seasonal exposure to gutter dirt or automobile emissions. Earickson and Billick (1988) described elevated lead levels in children in Louisville and Detroit in the 1970s, demonstrating statistical association between airborne lead and other pollutants and poor neighborhoods. Bailey, Sargent, Goodman, Freeman, and Brown (1994) investigated the geographic variation in lead poisoning among children living in Massachusetts in 1990. Their statistical analysis corroborated the CDC's (1992) report that poor minority children living in female-headed households in deteriorating housing built before 1950 are at disproportionately high risk for excessive lead exposure.

One of the most pervasive forms of air pollution is commonly known as *smog* (smoke-fog). In discussing the policy and health implications of air pollution, one needs to distinguish between two quite different forms with different locational characteristics. The most serious air pollution episodes involve the *classical* or London smog so closely associated with England and the Appalachian region of the United States in the 19th and early 20th centuries. These episodes were characterized by high levels of SO₂ and smoke particulates that built up under stagnant weather conditions lasting 3 days or more. In many industrial processes, sulfur is released during the burning of large quantities of coal and other fossil fuels. Combining with oxygen and eventually with water vapor, sulfur can become dilute sulfuric acid and damage buildings and cloth as well as human lung tissue ("Our Dirty Air," 1989).

Photochemical smog was first recognized as a problem in Los Angeles during World War II and has since been observed in and around many cities of the world (Stephens, 1987). In those early times in southern California, there was widespread awareness of an eye-burning haze, cracks in the sidewalls of rubber tires, and "bronzing" of sensitive vegetation. At first, this phenomenon was blamed on industry, then on the outdoor burning of trash. Only after exhaustive research did scientists become aware that the major source of these gas emissions was motor vehicles. Tropospheric (ground-level) O₃ is the main constituent of Los Angeles smog. It is actually a *secondary pollutant* formed when hydrocarbons, volatile organic compounds, and oxides of nitrogen (NO_x), known as *primary pollutants*, react in the presence of sunlight. Fuel combustion sources such as electric utilities and industrial boilers, as well as transportation vehicles, contribute to atmospheric O₃. Unlike CO, O₃ is worse during warm, sunny weather. Respiratory irritation and breathing problems can occur with chronic exposures to NO₂, particularly in people with asthma and bronchitis, at exposures as low as 100 parts per billion. Children also appear to be susceptible to bronchitis and breathing difficulties in the presence of low levels of NO_x.

Both forms of smog can be common health hazards in places characterized by frequent *atmospheric inversions*. Normally the temperature decreases with altitude

up to 40,000 feet (12 kilometers). The regular rate of decrease is 3.5°F per 1,000 feet (6.38°C per kilometer) and is known as the *adiabatic lapse rate*. Because warm air rises, the polluted and heated air from the surface rises and is dispersed through atmospheric mixing. There are several ways, however, in which the air at lower altitude can become cooler than the air above it. Most frequently this can occur near a large body of water such as a large lake or the ocean. When the sun goes down in the evening and the land cools off, a light breeze can direct cool air off the lake or ocean onshore. This forms a cold layer under the warm one, which rises to cover the cold layer. This is called an *inversion layer* and will trap the body of air below it (Figure 7.3). In valleys between mountains with steep sides, inversions occur as the morning or afternoon sun warms the upper layers of the air, while at the bottom the air is sheltered and remains cold or is cooled by a river flowing through the valley. An inversion at higher elevation can occur when there is atmospheric subsidence and divergence. Such upper-air inversions can occur under stagnating anticyclones, and they tend to occur in certain latitudinal zones where subsidence is common at least seasonally (see Vignette 4.1). Los Angeles became notorious for its smog not only because it has so many automobiles, but because its location is characterized by onshore breezes; downhill air drainage; radiational cooling on clear, cloudless nights; lots of sunshine; and summertime atmospheric subsidence. Of all these factors, only the automobile emissions are subject to control.

Frequencies of inversions and the associated health hazards of smog vary greatly not only from place to place but seasonally. Industrial emissions and automobile emissions vary over the course of a week, as well as diurnally. Some pollutants persist for days, and others for seconds or less. The various cycling times of years, seasons, weeks, days, hours, and minutes form a sampling nightmare for the researcher. Some particulate matter, such as lead, has a steep distance decay

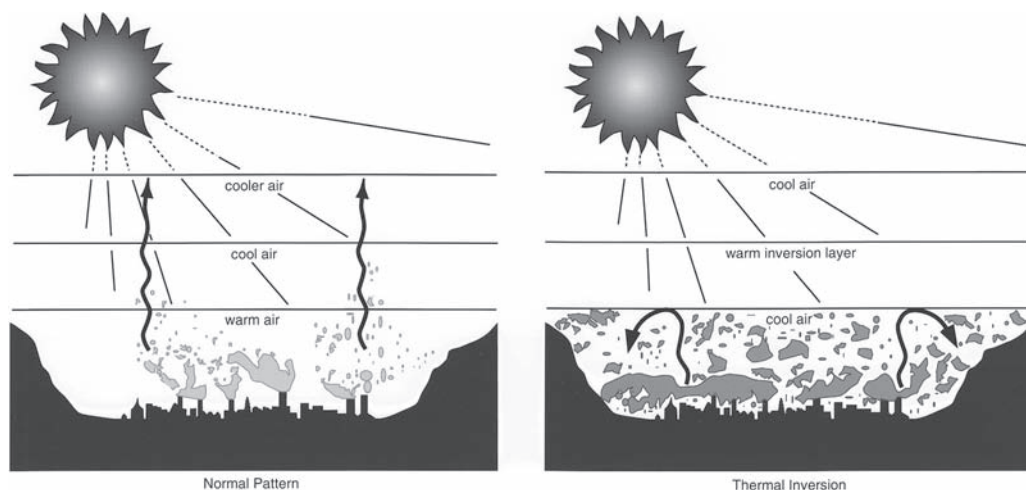


FIGURE 7.3. The effect of temperature inversion. The normal lapse rate may be reversed either near the ground or at high elevations.

that may move from dangerous to inconsequential within 200 yards of an intersection; some gases, such as SO_2 , may persist for thousands of miles. The several ways of modeling point sources of pollutants usually focus on mathematical models of plumes. Trigonometry can help predict where the pollutants will settle by calculating from the height of the ejection point (smokestack), wind speed, thermal stability, topography, and so forth. As numerous point sources merge into the areal base of a large metropolitan area, the dynamics of atmospheric mixing and movement become more relevant. The rate of emission is a major concern, of course, but so are wind speed and mixing height.

According to studies of U.S. metropolitan areas by Dockery and colleagues (1993), the finest particles (smaller than 2.5 micrometers) of dust, soot, smoke, or tiny droplets of acid are believed to cost tens of thousands of lives each year in the United States (see also Hiltz, 1995). The Dockery and colleagues research found statistically significant and robust associations between fine-particulate pollutants on the one hand and lung cancer and cardiopulmonary disease on the other. Fine particles may be more dangerous than large ones because people can breathe them deeply into the lungs, where tissue damage can lead to breathing difficulties or worse. This research was important because the research design controlled for cigarette smoking and other risk factors. Figure 7.2 demonstrates that in general, the worst fine-particle pollution is located in Southern California and a belt stretching from the Gulf Coast to the Great Lakes.

SO_2 is emitted largely from stationary coal and oil combustion sources, such as steel mills, refineries, pulp and paper mills, and coal-burning electric power plants. Human exposure to concentrations of SO_2 exceeding the national standard can affect breathing and aggravate existing respiratory and cardiovascular disease. Sulfur is especially hazardous to individuals with asthma, bronchitis, or emphysema, as well as to children and the elderly. Smelting of ores and other industrial processes may emit small particles of heavy metals such as lead, iron, manganese, and titanium, but most particulates are commonly carbon-based (i.e., soot). Mortality and respiratory illness increase significantly during heavy smog episodes. In London during a week in early December 1952, there were 3,500–4,000 more deaths than the expected average for that time of the year. The industrial valleys of Pennsylvania suffered a similar fate in 1948. Such dramatic events tend to raise public and official consciousness about air pollution. Recently it was estimated that air pollution in China kills 400,000 of its citizens each year (Watts, 2005).

There is another family of hazardous air pollutants, commonly referred to as *air toxics*; these have been shown to cause cancer, poisoning, and immediate illness in humans (U.S. EPA, 1996). Other less measurable effects include respiratory, immunological, neurological, reproductive, and developmental. Examples of air toxics include (but are not restricted to) dioxins, benzene, arsenic, beryllium, mercury, and vinyl chloride. In fact, U.S. clean air legislation lists 189 pollutants as hazardous air pollutants (HAPs) and targets them for regulation. Air toxics emanate from both stationary and mobile sources. In 1988, manufacturers in the United States emitted an estimated 2.4 billion pounds of toxic pollutants into the atmosphere (CDC, 1991). Toxins may be inhaled directly or deposited onto soil or into water bodies, thereby

indirectly affecting human health through ecological chains. Top consumers in the food web, usually consumers of fish, may accumulate chemical concentrations many millions of times greater than the concentrations present in the environment. As a result, fish consumption advisories have been issued for hundreds of water bodies in the United States, including the Great Lakes. The distribution of HAP emissions by state appears to correlate positively with the distribution of hazardous waste sites (U.S. Census Bureau, 1996, p. 238). The top 17 ranked states in terms of waste sites include 12 of the states in the highest HAP emissions category (more than 90,000 tons per year). The good news is that air pollutant emissions have on average decreased throughout the United States since 1970, despite growth in population, energy use, and vehicle activity (National Research Council [NRC], 2004, 2006). The bad news is that, as previously mentioned, NICs including China and India have shown the opposite trend.

The distribution within the United States of the release of toxic substances, including air toxics, is not uniform. Figure 7.4a is a map of Toxic Release Inventory (TRI) sites in 2006. The TRI is a publicly available database that contains information on toxic chemical releases reported annually. Figure 7.4b is a map of National Priorities List (NPL) sites, which are what the EPA considers the most serious uncontrolled or abandoned hazardous waste sites in the United States. As you can see from these maps, the entire eastern half of the United States is filled with TRI sites, and the Northeast, Midwest, and South have the most NPL sites.

Cutter and Solecki (1989) and Cutter and Tiefenbacher (1991) examined the spatial distribution of air-borne releases of acutely toxic materials in the metropolitan United States throughout the 1980s, and demonstrated through cartographic and statistical analyses that frequency of chemical firms and rail miles per state were good predictors of the number of incidents. Their study dispelled the notion that air-borne releases of acutely toxic chemicals occur only at fixed sites or from spectacular rail accidents. The highest likelihood of toxic chemical incidents was associated with high concentrations of chemical manufacture and distribution facilities.

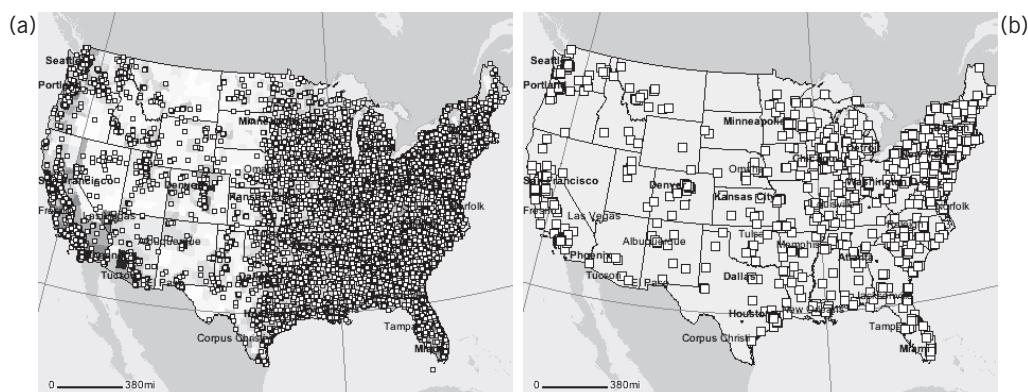


FIGURE 7.4. (a) Toxic Release Inventory (TRI). (b) National Priorities List (NPL) sites. From the U.S. National Library of Medicine (toxmap.nlm.nih.gov/toxmap/superfund/mapControls.do).

INDOOR AIR POLLUTION

Recent research has added a new dimension to the air pollution picture. A startling finding is that exposures to toxic chemicals are sometimes 70 times higher indoors than outdoors. This is partly because indoor air carries the double burden of outdoor pollutants brought inside and pollutants that are generated indoors. The CDC (1994) has classified indoor pollution as a factor of high environmental risk. A large burden of total indoor air pollutants, at least in the developed world, is probably attributable to environmental tobacco smoke in both homes and workplaces. The U.S. EPA (1992) estimates that this pollutant causes over 3,000 lung cancer deaths a year among nonsmokers and may be responsible for serious respiratory illness in hundreds of thousands of children. As public awareness of the hazards of exposure to environmental tobacco smoke increases, businesses and communities throughout the United States are taking actions to prevent involuntary exposure. These actions range from prohibiting smoking indoors to limiting smoking to specially designated, separately ventilated smoking rooms. (Our own university, which is in the tobacco state of North Carolina, recently banned smoking on the entire campus; smoking was also banned throughout the state in almost all restaurants and bars.)

Indoor air pollution in the developing world is largely caused by cooking and heating. For example, nearly all of China's rural residents use solid fuels (biomass and coal) for household cooking and heating, and such fuel use is responsible for approximately 420,000 premature deaths annually (Zhang & Smith, 2007). A recent meta-analysis showed that risk of pneumonia in young children is increased by exposure to unprocessed solid fuels by a factor of 1.8 (Dherani et al., 2008). In places such as Guatemala where open fires are used for cooking, one solution is the introduction of chimney woodstoves, which have been shown to reduce CO exposure and wheezing (Smith-Sivertsen et al., 2009).

A landmark EPA study (see Shabecoff, 1985, for a report) found it probable that such consumer products as paints, cleansers, propellants, plastics, and cosmetics, as well as adhesives, fixers, resins, and building insulation materials, were major sources of indoor pollution. This study, which measured the chemicals in the participants' bodies on a daily basis, found significant correlations between blood levels of these chemicals and visits to gas stations and dry cleaners, use of paint or solvents at home, and smoking. Important studies of air pollution in domestic space are sure to follow, but the habitat dimension of pollution hazards has been interlocked again with the behavioral dimensions of house type, clothing, hobbies, and customs of personal and domestic hygiene (see Chapter 2).

Modifications to the home environment in industrial countries have been linked with an increase in the prevalence of asthma (Jones, 1998). Modern homes are much better insulated than was previously the case. Many more homes now have central heating and sealed-unit double glazing. Fitted carpets have generally replaced loose rugs, and advances in construction technology have led to a greater use of synthetic building materials. As a result, houses are warmer and more humid, lacking fresh air. In this environment, air-borne contaminants are readily produced and build up to higher concentrations than are encountered out of

doors. Various indoor exposures have been related to asthma, including house dust mites, molds and fungus, NO₂ from gas appliances, volatile organic compounds, and formaldehyde.

Approximately 20 million Americans have asthma, including 6.3 million children. During the past two decades there have been large increases in the number of people with asthma, predominantly in children under 5. In 2000 there were almost 2.5 million hospital visits due to asthma, at a cost of almost \$2 billion. Although the asthma epidemic is clearly related to indoor air pollution, it is simultaneously related to the air quality of the community in which a person lives and works. For instance, Sahsuvaroglu and colleagues (2009) found that traffic-related pollutants, such as NO₂, were associated with childhood asthma in Hamilton, Ontario. In New York, Maantay (2007) found that people living near noxious land uses were 66% more likely to be hospitalized for asthma. This latter finding is also an issue of *environmental justice* (EJ), since the people in New York who were hospitalized with asthma were 30% more likely to be poor and 13% more likely to be minority group members. EJ is an area of inquiry that investigates the inequitable environmental burden borne by minorities and people living in economically disadvantaged areas. Asthma has been a common condition used in EJ studies because the disease burden disproportionately affects disadvantaged groups.

Researchers have also been paying increased attention to the radioactive gas known as radon. Radon is an odorless, colorless gas that arises from the natural decay of radioactive elements commonly found in rocks and many types of soil. In tightly sealed masonry structures, radon gas can build up to hazardous levels. According to EPA surveys, the average radon level in U.S. homes is 1.25 picocuries per liter of air. Almost all scientists agree that prolonged exposure to high levels (more than about 8 picocuries) of radon poses a genuine cancer hazard. The U.S. National Academy of Sciences estimates that radon is the cause of thousands of lung cancer deaths nationwide each year (National Research Council, 1999). Most of these deaths occur among people who smoke cigarettes. Other studies, however, show little evidence linking household exposure to cancer. In Winnipeg, for instance, researchers found that mean household exposures of 3.5 picocuries appeared to produce no increased risk for lung cancer. This raises questions about how much risk radon poses at low levels (Leary, 1994). In the United States, known radon concentrations are highest in Western mountain states, the Midwest, and Appalachia.

WATER POLLUTION

Several episodes have been responsible for attracting research attention and public interest to the hazards of water pollution. One of the first occurred in Minamata, Japan, during the 1950s, when more than 200 people died or suffered severe brain and nerve damage. The cause was methyl mercury polluting the water and being passed through the food chain in fish to the population.

In the 1970s the Love Canal incident, near Niagara Falls, New York, focused national attention on the problems of hazardous waste disposal and the dangers

of groundwater contamination. Hooker Chemicals and Plastics Corporation (now Occidental Chemical) purchased an uncompleted canal site, and between 1942 and 1953 disposed of about 22,000 tons of mixed chemical wastes into the ditch. Shortly after Hooker ceased use of the site, the land was sold to the Niagara Falls School Board for a price of \$1. In the late 1950s, residential development of the area attracted 900 families, and an elementary school was constructed on the property. Unusually heavy rain and snowfalls in 1975 and 1976 provided high groundwater levels in the Love Canal area. Portions of the Hooker landfill subsided and chemical residue surfaced, contaminating ponds and other surface waters. Basements began to ooze an oily residue; noxious chemical odors permeated the area; and chemical corrosion of subterranean pumps became apparent. In 1978 the New York State Department of Health declared the Love Canal area a threat to human health and ordered the fencing of the landfill site. Later the school was closed, and pregnant women and young children were evacuated from the site. Eventually, all the properties in Love Canal were purchased, and a Federal Disaster Assistance Agency was created to assist the city of Niagara Falls to remediate the Love Canal site. Since then, gasoline leaks from underground tanks, PCB dumps into rivers and drains, and leakage from municipal landfills have poisoned the water supplies of scores of communities nationwide and forced Congress to create a Superfund program for toxic waste cleanup (Zeigler, Johnson, & Brunn, 1983).

Ninety-five percent of all fresh water available on earth (exclusive of ice caps) comes from beneath the surface. Although most people in the world get their domestic water supply from surface water or shallow wells, about half the population in the United States and large cities elsewhere tap groundwater sources. The purity of groundwater has long been held as sacrosanct; however, untreated contaminated water can carry many infectious agents. The U.S. EPA (1994) reports the leading sources of water pollution and types of water pollutants in the nation's rivers, lakes, ponds, reservoirs, estuaries, and the Great Lakes. Agriculture is the most widespread source of pollution in the surveyed rivers. Bacteria from fecal contamination pollutes 34% of impaired river miles; excess nutrients, especially nitrogen and phosphorus compounds from agriculture, pollute 43% of lake acreage; and urban runoff and storm sewers are the most widespread source of pollution (nutrients and bacteria) in estuaries. The Great Lakes contain one-fifth of the world's fresh surface water and are polluted from both the air and surface runoff. In addition to nutrients from agriculture, toxic organic chemicals (primarily PCBs) have been found in 98% of the samples taken. Agricultural chemicals have led to feminization in cane toads in Florida because of high hormone levels (McCoy et al., 2008). Air pollution from industry and pesticides from agriculture have been implicated.

Millions of people living near the Aral Sea area (Uzbekistan and Kazakhstan) are experiencing one of the world's largest environmental disasters. Rivers feeding into this sea have been tapped for irrigation for cotton production for the past 40 years, and it is now only approximately 25% of its original size. This has led to many environmental health problems, not the least of which is poor water quality. Dissolved salt levels in drinking water are more than 20 times what they are in North America (Small, van der Meer, & Upshur, 2001).

Given the number of illnesses and deaths attributed globally to diarrhea, it seems fair to say that human feces remain among the world's most hazardous water pollutants (Fewtrell et al., 2005). Poor sanitation and a lack of access to clean water still constitute two of the world's most serious health problems. In the developed world, these problems have been mitigated by providing indoor piped water and flush toilets to virtually all urban residents. The same is true for the wealthy in developing countries. For most developing nations, however, financially strapped governments can ill afford anything like complete coverage with indoor plumbing. Table 7.3 lists 2004 data for those developing nations that lagged behind in terms of access to clean water and sanitation in the early 1990s. The table also includes the United States as a comparison.

The Joint Monitoring Programme for Water Supply and Sanitation (2006), which is maintained by both UNICEF and WHO, defines minimum standards of water and sanitation as follows. *Safe drinking water* includes a household connection to a water supply, a public standpipe, a borehole, a protected dug well, a protected spring, or rainwater collection. *Unsafe drinking water* includes surface water (river, dam, lake, pond, stream, canal, irrigation channel), an unprotected well, an unprotected spring, vendor-provided water, bottled water, or tanker truck water. *Improved sanitation* includes connection to a public sewer, connection to a septic system, a pour-flush latrine, a simple pit latrine, or a ventilated improved pit latrine. *Unimproved sanitation* includes a public or shared latrine, an open pit latrine, or a bucket latrine.

In cities of the developing world, households without indoor piping often obtain their water from one of many sources, such as overcrowded or distant communal standpipes; expensive private water vendors; or heavily polluted, shallow,

TABLE 7.3. Access to Safe Drinking Water and Improved Sanitation, Selected Low-Income Nations, 2004

Country	Access to safe drinking water (%)		Access to improved sanitation (%)	
	Urban	Rural	Urban	Rural
Bangladesh	82	72	51	35
Burkina Faso	94	54	42	6
Central African Republic	93	61	47	12
China	93	67	69	28
Ethiopia	81	11	44	7
Guinea-Bissau	79	49	57	23
India	95	83	59	22
Indonesia	87	69	73	40
Mali	78	36	59	39
Mozambique	72	26	53	19
Myanmar	80	77	88	72
United States	100	100	100	100

Note. Data from Joint Monitoring Programme for Water Supply and Sanitation (www.wssinfo.org/en/welcome.html).

hand-dug wells. Increasingly, developing countries are adopting borehole technology, where deep tube wells with engine-powered pumps are becoming important sources of uncontaminated domestic water (Emch, 1999; Fewtrell et al., 2005). In countries where vector-borne disease is prevalent, for example, tube wells are providing parasite-free village water sources for bathing and laundry. Those without flush toilets may end up using pit latrines or latrines located over ponds, streams, drains, or open sewers—all of which demand far more rigorous hygiene behavior than is required for the standard technologies of the wealthy.

Figure 7.5 illustrates the strata of an artesian water source. Surface water percolates into the soil, eventually reaching the water table, or a saturated zone of ground above an impervious stratum of rock. This groundwater may also intersect depressions and form springs or help fill lakes. Below the stratum of impervious rock there is often another layer of pervious rocks, such as sandstone, and a layer of impervious rock below that. Most rock strata are tilted. Where the pervious layer has contact with the surface, precipitation percolates into it to form an *aquifer*. The source area of this precipitation may be many hundreds of miles from a desired well site, so aquifer water may be under pressure if the rock strata are even slightly tilted. An artesian well drilled through the rock cap into the aquifer may, if the location is right, flow under its own pressure. Although aquifers are filled by precipitation, in some places the water that originally filled the aquifer fell in other climatic periods and other continental locations.

There may be several layers of aquifers. Often the lowest one is saline. Injection wells are sometimes used to dispose of toxic wastes in saline aquifers. Under

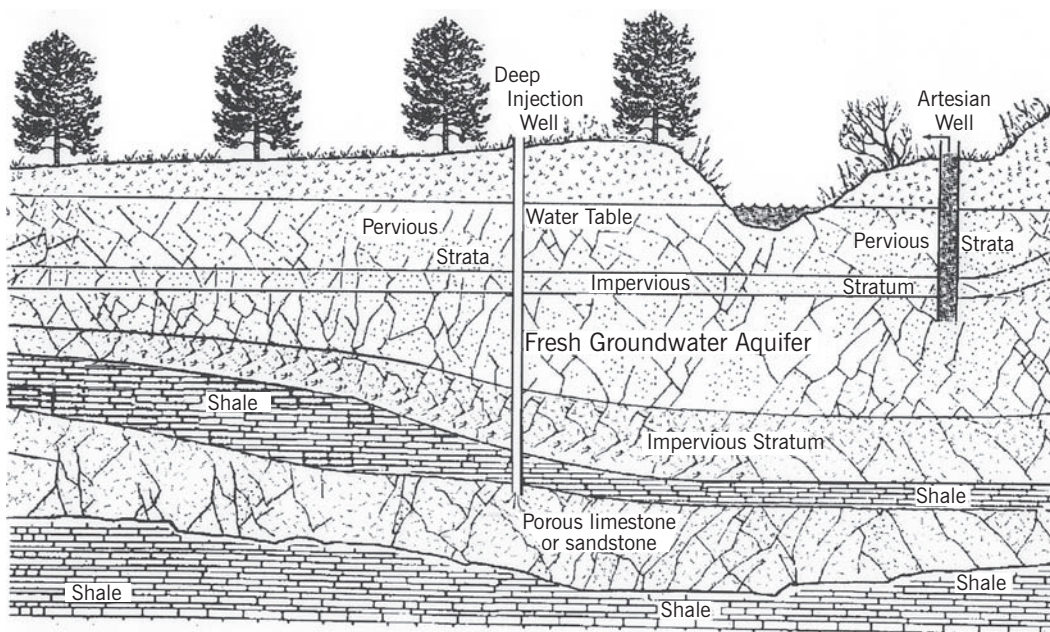


FIGURE 7.5. Groundwater strata.

great pressure, and through wells constructed in layers and sealed to prevent leaching into higher strata, liquid wastes are injected into the saline aquifer. Overall, U.S. groundwater quality is considered "good," but some regions have experienced "significant" groundwater contamination. The most frequent sources of this contamination are leaking underground storage tanks, agriculture (pesticides and herbicides), Superfund sites, and domestic septic tanks (U.S. EPA, 1994). Recently in several parts of the world, including Bangladesh and Vietnam, groundwater has been found to be contaminated with naturally occurring arsenic (Berg et al., 2001; Smith, Lingas, & Rahman, 2000).

One of the most comprehensive recent studies of water pollution and human health effects was carried out in Puerto Rico by Hunter and Arboña (1995). Because of its historical linkages to the West and subsequent industrialization, the Commonwealth of Puerto Rico has pushed consumerism to levels comparable to those found in the rest of the United States. The resultant outpouring of sewage and solid waste imperils this small, overcrowded tropical island that is dependent on groundwater. The sudden death of large numbers of fish has provided a sensitive biological indicator of environmental quality. The usual suspects include agricultural runoff and discharges from industry and municipal sewers. Here as elsewhere in the United States, water supplies are becoming increasingly polluted with toxins; yet water treatment techniques still primarily reflect concern with infectious diseases of the past, rather than the carcinogenic pollutants of today's environment. It has been several decades since the passage of the Safe Drinking Water Act, yet the gamut of toxic chemicals in the public water systems is broadening, and threats to human health have increased rather than diminished.

Concern about pollution of groundwater is growing. In the mid-1980s, the California Department of Health Service reported that one-fifth of the state's large drinking water wells fed by groundwater exceeded the state's pollution limits (Sun, 1986). Studies in two heavily populated counties in that state found high levels of trichloroethylene and dibromochloropropane, which pose reproductive and acute health risks, including stomach cancer and lymphatic leukemia (Kloos, 1995). Other studies, however, have found no relationship between contaminants and disease. This is often the case when there is a failure to control for confounding factors or when conclusions about a population are based on samples obtained in small areas. This can lead to ecological fallacies, which have been mentioned elsewhere in this book. Pesticides have been detected in half of Iowa's city wells. More than 1,000 wells in Florida have been shut down because of contamination with EDB, a chemical used to kill soil nematodes. The common beliefs that pesticides would decompose in the soil and that nature would cleanse itself through percolation can no longer be accepted.

As with any pollution, the critical parameters are amount, mixing/dispersal, cleansing, and decay. The fate of chemicals in surface water or groundwater is very different in these respects. Surface waters can be easily polluted, and the pollution can spread rapidly, be focused into high concentrations in time, and be flushed and removed. Large lakes and estuaries require a longer time to accumulate the pollutant, and perhaps decades to turn over and be cleansed. In contrast to these,

groundwater quantities are enormous and spread over entire basins and even continents. The water is so slow moving that sometimes it is difficult to determine the direction of flow. It may take years for a pollutant to move more than a few feet from its source. It is difficult for pollutants to get access to groundwater because of filtration by the earth; once it is contaminated, however, groundwater is nearly impossible to cleanse, and long periods (up to hundreds of thousands of years) may be required for its renewal. The actual size of the groundwater supply in the United States is not known because extensive mapping of the nation's aquifers has not been completed.

RADIOACTIVE POLLUTION

Both nuclear and biological terrorism have become plausible contemporary threats to most nations. Besides the major nuclear powers, many smaller and less politically stable nations now possess the capacity to build nuclear devices. It is generally agreed not only that outright nuclear war would devastate specific target areas, but that even a limited nuclear action could drastically affect the climate, ecosystem, and human genotype everywhere on earth (NRC, 1975). In this chapter, we address only the hazards of peaceful nuclear power.

Radiation Hazards

Naturally occurring radiation hazards are cosmic radiation and decay of radioactive elements. Cosmic sources provide a continuous background of eternal radiation with which all life has evolved. Today about half of the radiation exposure to the average person in the United States is from the natural background. Another 40% is attributable to medicine in the form of diagnostic X-rays and radiation therapy (Upton, 1982). Much of the natural radiation emanates from radioactive materials that have been around for some 4 billion years. Substances such as uranium-235 (U-235), uranium-238 (U-238), and thorium-232 have extremely long *half-lives* (the time required for a material's radioactivity to decrease by half). Knowledge of some of their decay sequences and by-products helps in understanding the nature of certain environmental hazards.

In April 1986 a tremendous explosion from the reactor at the Chernobyl nuclear power plant in the former Soviet Union (present-day Ukraine) sent a radioactive "plume" of radionuclides into the upper atmosphere, where it began to drift into northern and western Europe. Substantial areas throughout the continent were bombarded with different kinds of radioactive isotopes. (Figure 7.6 shows a map of the distribution of cesium-137 released after the Chernobyl accident.) From April 26 to 28, high atmospheric pressure over northeastern Europe established a northerly airflow that carried the plume—at first affecting the former Soviet Union, then later northeast Poland and Scandinavia, where radiation monitors in Sweden and Denmark indicated abnormally high readings. The triggering of these monitors was the first indication in western Europe that a significant nuclear accident had occurred

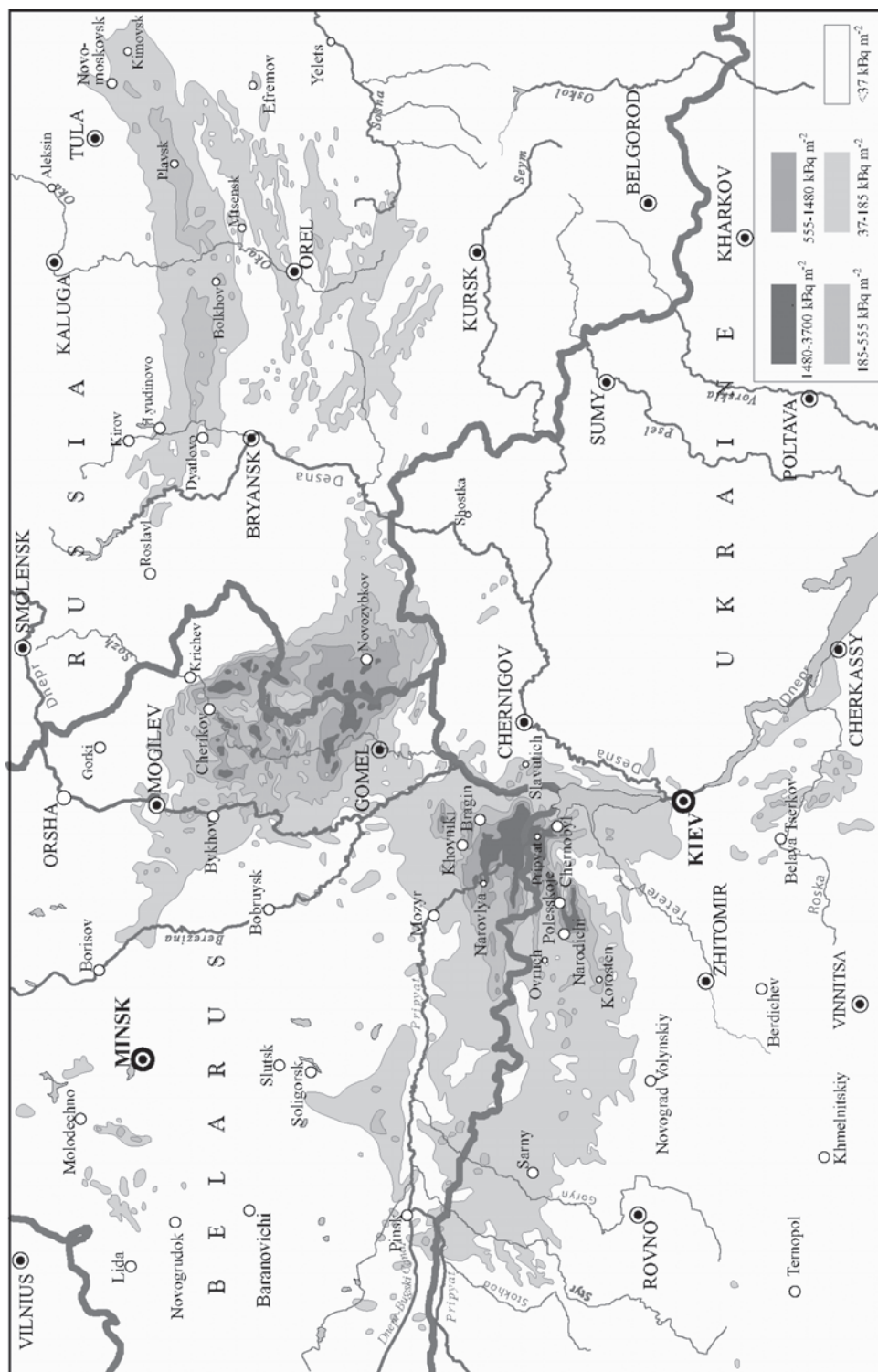


FIGURE 7.6. Distribution of cesium-137 released in the Chernobyl accident. From the United Nations (www.unscear.org/unscear/en/chernobyl.html).

(Gould, 1990). All territory within an 18-mile (30-kilometer) radius of the destroyed reactor had to be evacuated and remains uninhabited today. Radiation may have affected unborn fetuses near the site, but by the time the plume reached England, increases in perinatal mortality were insignificant (Bentham, 1991).

Chernobyl used, and other nuclear power plants still use, the fission power of U-235 and U-238. As U-238 decays it produces radon-222, an inert gas that diffuses into the atmosphere. Radon-222 is the source in turn of lead-210, a global constituent of atmospheric fallout. It decays to polonium-210. Among the radioactive daughter nuclei that result, cesium-137, iodine-131, and strontium-90 are particularly important because they accumulate within organisms. Cesium-137 has a half-life of 30 years and attacks the entire human body. Iodine-131, with a half-life of 8.1 days, collects in the thyroid gland. Strontium-90 is similar to calcium and becomes part of the bones, where it affects the production of red blood cells in the marrow. Bone marrow is essential for producing materials needed by the body's immune system.

Several radioactive gases are also produced by current nuclear technology and released into the atmosphere to become part of the background, or ambient, radioactivity. Principally these are the inert gas tritium and the isotopes krypton-85, xenon-133, and xenon-135. These do not enter into biological or chemical activity and so are not accumulated by organisms. The release of radioactive gases and other products is regulated to stay within limits of normal variation in background radiation. One often hears the reassurance after an unscheduled release of radiation that there is no more exposure for the population than getting a dental X-ray. (The exposure is, of course, in addition to the dental X-ray.)

Radiation can cause cancer and other cell mutation. Damage to DNA may take generations to be expressed. As with any disease agent, some people will be affected at lower levels of radiation than others who remain apparently untouched. Little is known, however, of dose relationships either at the individual or the population level. Is the relationship linear—the more radiation, the more mutation? Is there a threshold level below which the body usually can cope, and above which damage begins? Are there different levels of tolerance, depending on whether the radiation is inhaled, ingested, or contacted?

There seems to be considerable variation among species in resistance to radiation. Much of our knowledge about what radiation does to the body has been provided by studying the effects of the atomic bombs dropped on Hiroshima and Nagasaki in 1945. Accidental and occupational exposures provide some data, but generally about acute effects. The soldiers exposed to watching nuclear explosions in the early days of atmospheric testing have been claiming compensation for their cancers and reproductive problems, but it has proven difficult for all the experts in court to separate these problems from those occurring in the general population. If humans are exposed to a dose of between 100 and 250 rads, they will develop fatigue, nausea, vomiting, diarrhea, and some loss of hair. However, most victims recover completely from the immediate illness. Doses of 400–500 rads, however, like strontium-90, impair bone marrow function. Fifty percent of those exposed in this dose range will die. A dose of 2,000 rads is almost always fatal within a few weeks of exposure (Turk, 1985).

Johnson (1981) investigated the case of the radiation hazard around the Rocky Flats nuclear weapons plant near Denver, Colorado. He demarcated a region of hazard, using the plutonium content of the soil for miles around the plant to measure cumulative exposure from 1953 to 1971. Figure 7.7 shows the results of calculating the relative risk of excess cancer deaths from 1969 to 1971. The cancers involved were those associated with Japanese survivors of the atomic bombs: leukemia; lymphoma; myeloma; and cancers of the lung, thyroid, breast, esophagus, stomach, and colon. The distance decay effect associated with plutonium released through the plant's smokestack (note the westerly wind direction configuration) clearly creates a hazardous region. The hazard zone was not perceived for two decades; otherwise, the release of radionuclides might have been reduced or the population warned. Populations may be exposed to hazards without knowing it, and thus may experience symptoms and not relate them to environmental conditions.

Since World War II, the dramatic growth in the use of electric and electronic devices has been accompanied by a parallel increase in human exposures to the electromagnetic fields (EMFs) that these devices emit. Electric power and telecommunication systems, electric and electronic appliances, broadcast facilities, and radar systems have all contributed to rising background levels of EMFs in living and work spaces in industrialized countries. Although we cannot deny the benefits

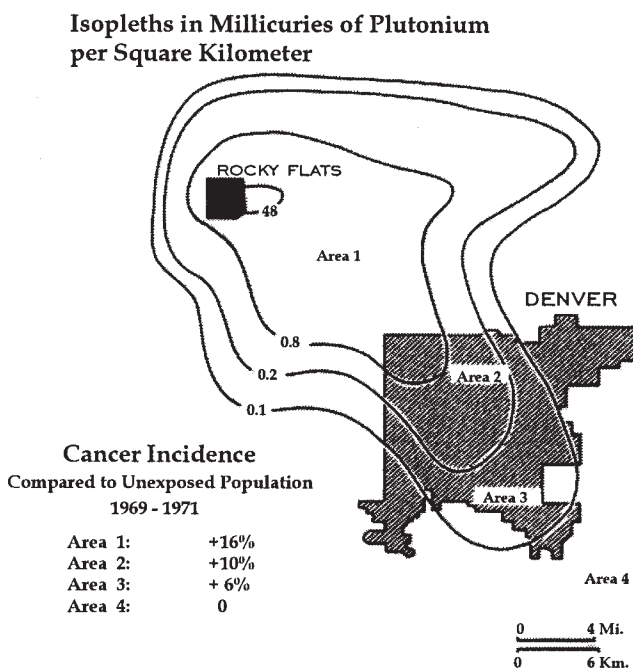


FIGURE 7.7. Cancer incidence around a nuclear weapons plant. An isoline map of plutonium that has accumulated in the soil provides a regionalization for exposure and correlation with cancer incidence. From Zeigler, Johnson, and Brunn (1983, p. 48). Copyright 1983 by Association of American Geographers. Reprinted by permission.

of these technologies, they have raised concerns about whether exposure to typical levels of EMFs in the environment are harmful to health. In several compelling articles on the subject in *The New Yorker* magazine, Brodeur (1989, 1990) presented detailed descriptive accounts as circumstantial evidence that some cancer clusters were associated with chronic exposure to the EMFs given off by power lines carrying high current and high voltage to electric power substations, which are not uncommonly located near residential areas. To support his case, he cited some two dozen epidemiological studies that had been conducted and published in the medical literature in the United States and elsewhere, showing that children and workers exposed to power-line-generated EMFs were developing leukemia, lymphoma, melanoma, brain tumors, and other cancers of the central nervous system. Such claims have been hotly contested by other epidemiologists. Florig (1992) concedes that the scientific evidence about the health risks of exposure to EMFs is "suggestive" of injurious health effects, but he notes that this evidence has thus far not been compelling enough to trigger regulatory action by federal authorities. Because of possible confounders and biases in most of the epidemiological studies available at this writing, scientific opinion about whether the evidence from these studies represents a real EMF health effect varies enormously.

The geographic scale of a research scheme can confuse resultant statistics. One example is provided in a study of child and young adult cancers in New Jersey using surveillance at several scales—urban, suburban, rural county, minor civil division, and individual communities (Schneider, Greenberg, Donaldson, & Choi, 1993). If one compared New Jersey's pediatric cancer incidence with that of other states, that state's cancer rate would have appeared to be on the increase. Surveillance at a finer scale, however, showed no statistically significant pattern of excess cases. Yet another level of surveillance suggested that pediatric cancer did cluster statewide, and that the clustering was strongest among white cases in specific age groups with specific cancer types. A final examination of individual cases yielded information that clusters were exemplified by very few cases and very few small political divisions. Thus, although clustering might exist at the state scale, few smaller jurisdictions had enough cases to support expensive epidemiological studies that might yield fruitful results. Knowing this information at the outset might counterbalance political pressures exerted by concerned local populations for government action.

Political responses to community health concerns have had the effect of delaying the siting and licensing of new power lines, radar systems, and communications antennas, and numerous court cases have arisen in which plaintiffs claimed impaired health because of exposure to EMFs. Hundreds of millions of dollars have been spent in the United States alone on behalf of government and industry to investigate the biological effects of EMFs. Because the scientific evidence on EMF bioeffects is both complicated and contradictory, regulatory bodies and organizations concerned with scientific standards have been unable to reach a consensus on prescriptive approaches to EMF risk management (Flynn, Kasperson, Kunreuther, & Slovic, 1997). Rerouting power lines, repacking wires, moving office spaces to distance workers from power sources, and increasing the shielding of emissions from electrical appliances are some of the possible ways to ameliorate exposures. All of

this is obviously expensive and raises economic issues, such as whether mitigation costs balance out potential health benefits, or whether any party bears an unfair burden of potential risk. Like other environmental issues, the EMF controversy is as much about sociopolitical and ethical concerns as it is about health risk. There is a fertile field here for geographers to investigate not only the deleterious health effects in proximity to EMF sources, but also the perceptual context (Sooman & Macintyre, 1995).

Other Sources of Radiation Hazards

The hazards associated with nuclear power plants and nuclear weapons have been emphasized previously, but there are other sources of radiation hazards. One is the use of radiation in hospitals and research laboratories, and the disposal of the low-level radioactive wastes they produce. Another is the mining of uranium and the manufacture of nuclear fuel. Yet another is the transport and storage of all the used material.

Uranium ore typically contains 2–5 pounds of uranium oxide per ton. After uranium is mined, therefore, large piles of waste and tailings remain. Efforts have been made to grade the slag heaps and stabilize them with vegetation, but in the Colorado River basin alone they cover several thousand acres and contain tens of millions of tons. Thorium-232, U-238, and radon can be leached from these heaps by rainfall and eroded into water systems. The occupational exposure of the miners and subsequent processors has to be carefully controlled. The ore is crushed, ground, and leached with acid, and uranium is recovered from the leached liquid by such procedures as ion exchange and solvent extraction. Radioactive wastes are produced at each step.

The need to transport radioactive material creates linear hazard zones. Uranium from the mines must be transported to enrichment and fuel fabrication facilities. The fuel must be transported to nuclear reactor sites. Spent fuel must be transported to storage sites. Most nuclear power plants in the United States are east of the 100th meridian, in the more densely populated parts of the country. The linear hazards of transport are not going to diminish as the amount of spent fuel from later-generation plants increases and as the older plants themselves begin to be dismantled and transported for storage. Dry-cask storage enables these wastes to be stored for a century or more at existing power plants and storage sites; however, permanent repositories (ones that may be considered “safe” for more than 10,000 years) are needed. U.S. energy agencies began a search in 1975 for such a site in the states of Michigan, Ohio, New York, Utah, Texas, Louisiana, Mississippi, Washington, and Nevada. In 1982 Congress passed the Nuclear Waste Policy Act, which designated that two repositories would be constructed—one in the western half of the nation, the other in the eastern half. Considerable political opposition from targeted states ensued; the act was amended in 1987, and Yucca Mountain, Nevada, was named as the only repository site to be studied. This project has been stalled by unexpected scientific, management, intergovernmental, cost, public opposition, schedule, and regulatory compliance problems. As of this writing, the state of Nevada remains

opposed to this repository, and although development of the site proceeds, it is not a certainty that all high-level nuclear waste in the nation will ultimately reside there (Flynn et al., 1997).

The health consequences of nuclear power are difficult to assess and easy to exaggerate. Clearly, however, the spread of nuclear power for the generation of electricity to low-income countries that, because of lack of money and skilled personnel, are less able to regulate it and maintain the highest standards of safety practice is a matter of concern. It necessarily involves greater international shipment of nuclear fuel and—because of efforts to control the general availability of plutonium, which can be used for nuclear weapons—the return shipment of spent nuclear fuel. We are now well into the nuclear age.

RISK ASSESSMENT AND PREVENTION

With more than 3 million chemicals registered, 70,000 chemicals in general use, and new ones appearing at a rate of more than 1,000 a year, it is not surprising that the environment in the industrialized countries has become contaminated. Exposure to pollutants can be either acute or chronic. A person may be exposed once for a few minutes at work to a leakage or spill, or may consume over a lifetime undetectable levels of contaminant in food or water. Toxic waste from a dump may contaminate a nearby residential water supply to very high levels, or low levels of benzene may be breathed throughout a lifetime in the ambient air of an urban area. Even for most of the chemicals thought to be hazardous, little is known about threshold levels or dosage effects. Threshold limit values for major chemicals have been announced by government agencies responsible for protecting workers. These values establish time-weighted average concentrations that are acceptable for occupational exposure, in the same way that standards have been set for exposure to radiation, noise, and air-borne dust. Few standards have been estimated for nonoccupational exposures.

Research is complicated because reliable, sub-national-scale data do not exist for etiological study. There are, for example, no micro-area data for industrial exposures, cigarette or alcohol consumption, or dietary composition. One must work between small local surveys of specific points and broad socioeconomic classifications of national associations and trends. It is not surprising that point and area data are often improperly mixed, that researchers generalize across scale, or that spatial autocorrelation is ignored (see Vignette 12.3 in Chapter 12 for an explanation of this). The personal monitors, currently being tested, that identified indoor pollution may in the future provide reliable data for researchers interested in studying micro-scale exposure through daily and weekly activity patterns. For now, most studies rely on surrogate but measurable variables, such as substituting ethnicity for diet.

One of the clearest and most concise explanations of the measurement and uncertainties associated with risk estimation is that by Talcott (1992), who claims that sizable uncertainties are associated with quantitative environmental risk

assessments. Risk estimates incorporate the following questions: What is the probability that release of a substance will occur? What quantity of this substance will be released? How will the concentration of this substance change as it disperses from the point of release? How many people or other organisms in the environment will be exposed to this substance? How much of the delivered dose of this substance will be taken up by organisms? And what will be the relationship between a particular dose of this substance and an organism's response; that is, how will the toxic effects of this substance increase with increased dose? From the point of view of environmental justice, this last question has deadlocked many a jury. The toxicity of a particular dose of a substance varies not only across species, but among individuals of the same species. Sex, age, size, diet, and the geography of exposure to a substance, among other confounders, affect how toxic increased doses of a substance are to organisms.

One way the EPA assesses populations at risk is by using the tools of geographic information systems to identify the buffer zones around hazardous sources such as dumps or nuclear power plants. Complicating factors include prevailing wind direction; changes in humidity; local turbulence and mixing; air drainage; frequency of inversion; and differences in soil, rock, and topography. Underground movement of water causes a population situated upflow to be at higher risk from the hazard location than a population situated downflow a little farther away. The transportation and routing system, labor circulation, or seasonal changes in population density or industrial production also affect risk. Studies by political scientists, sociologists, and geographers have shown that often the most critical locational determinant for siting a hazardous facility is the political weakness of the local population (Greenberg & Schneider, 1996).

Geographers can help in the detection of old waste disposal sites, the planning of new sites, and the etiological analysis of diseases in developed economies. Processes of siting, transporting, storing, processing, concentrating, or dispersing occur in a differentiated environment of short-term micrometeorology, local geomorphology, transportation systems, settlement patterns, and economic and social activity. Risk assessment of population exposure to known and unknown hazards is a multiscale and complicated business of intrinsic geographic interest, but there has been limited geographic contribution thus far.

GLOBALIZATION AND THE PERCEPTION OF HEALTH HAZARDS

Pollution flows without regard to international boundaries. Acid rain, the greenhouse effect of increasing levels of carbon dioxide, and the fallout of strontium-90 as a result of atmospheric nuclear testing are problems shared by humans in general, regardless of the sources. There has been a less generally recognized but concomitant diffusion of useful technology that produces hazardous by-products. Among the most sought-after industrial capacities are petroleum refineries, the manufacture of plastics and synthetic materials, electronics, pharmaceuticals, and the pro-

duction of fertilizers and insecticides. Mining and smelting are being extended into previously undeveloped areas. Agriculture is being intensified through application of more chemicals. Research is progressing into such useful areas as recombinant DNA and the capacity to produce strains of food crops resistant to fungi and insects—without consideration of what effects the “naturally occurring” plant chemicals that produce resistance will have on human health.

Chlorofluorocarbons (CFCs) emitted into the atmosphere, because of their extraordinary chemical stability, reach the stratosphere where they release chlorine by photolysis. This free chlorine scavenges O_3 and destroys it (Lindley, 1988). Release of CFCs occurs through industrial activity, equipment leakage, or the disposal of old refrigeration and air-conditioning units, as well as by use of aerosol cans using CFCs as propellants. Even some members of the health care professions use CFCs as a diluent for ethylene oxide in cold sterilization procedures in hospitals, clinics, and the manufacture of health-related devices. Whereas CFCs are the most voluminous O_3 depleters, there are several other O_3 -depleting substances in use, including carbon tetrachloride, hydrochlorofluorocarbons, halons, methyl bromide, and methyl chloroform.

The O_3 layer was observed to be thinning over Antarctica in the late 1970s. Repeated observations have confirmed the attenuation and charted its progress (Stolarsky, 1988). There is some attenuation in the Northern Hemisphere as well, but at about 60° north, not over the pole. On average, the global stratospheric O_3 layer declined by 2–4% between 1980 and 1988 (Last & Guidotti, 1991). Between 1988 and 1995, global production of CFCs dropped 76%, due mainly to a reduction in use of O_3 -depleting substances in industrial nations.

There has been noticeably increased human exposure to ultraviolet (UV) light in recent years. O_3 in the stratosphere absorbs much of the UV light that would otherwise reach the earth's surface. It has been estimated that a 1% reduction in the O_3 shield would result in a 2% increase in the UV light reaching the surface. The human health effects of increased UV irradiation due to O_3 depletion include higher risks of nonmelanoma skin cancer and malignant melanoma; cataract and retinal degeneration; and possible impaired immunological responses that increase risks of other conditions, including systemic malignancies. Even at high northern latitudes, exposure to sunlight is the principal determining factor for nonmelanoma skin cancer (MacKie & Rycroft, 1988). Oddly, the amplification factor appears to decrease with increasing latitude, which means that residents in the very highest latitudes of North America may be at less increased risk for the carcinogenic effect of increased UV penetration than those at the latitudes where most of the Canadian and some of the U.S. population lives. There have been successful attempts to curb CFC generation and release (French, 1997). On January 1, 1996, production of CFCs was to have ceased in developed countries under the provisions of the Montreal Protocol of the Vienna Framework Convention on Ozone-Degrading Substances. However, given that there will not be a freeze of CFC production until 2013, and given the long half-life of CFCs (75 years or more), O_3 -depleting activity is expected to persist at significant levels well into the 21st century. Consumption is rising rapidly in some countries with large populations, including China, India, and the Philippines. Also,

there is a black market in smuggled CFCs in Eastern Europe, which threatens to undermine the phaseout in industrial countries.

The experience gained by industrial nations as they passed through their technological transition should have been of great preventive value to poor developing nations, but there is little evidence of its application. Modern health hazards are diffusing rapidly. Most developing countries perceive the immediate benefits of technology more acutely than they see the distant, nebulous hazards. Their national budgets are stretched thinly to cover the needs of education, health, and infrastructural development. Civil servants with the education and skills required to inspect and regulate plants that use advanced chemical technology are in limited supply. The general populace is usually uninformed and naive about modern chemicals. Farmers, for example, commonly contract urticaria and more severe poisoning from handling insecticides and herbicides, and they sometimes rinse their equipment in streams that provide water for drinking and washing. If they are illiterate, they cannot read labels. Farmers in Southeast Asia have been known to dump insecticide in streams to kill fish for harvest, as they used to do with native plant toxins. Chemical and drug manufacturers have been known to export to developing nations poisons and other substances that have been banned for use in developed countries. Herbicides similar to Agent Orange, used in the Vietnam War, have been marketed to countries such as Colombia, resulting in reports of miscarriages and birth defects. The attitude in developed countries is often that other countries must regulate themselves, but the expert knowledge and resources are seldom available in the developing world.

THE GEOMETRY OF HAZARDS, POWER, AND POLICY

There are many spatial dimensions to environmental health hazards. Atmospheric and water pollution constitute areal hazards; mines, manufacturing sites, refineries, and waste and storage dumps constitute point hazards; and the roads and sea lanes of transportation constitute linear hazards. Each is of concern at a different scale, and each involves different kinds of regulatory and preventive policies and different levels of government. Technology control and alternatives must be carefully considered in terms of social justice and compensation, private investment, and tax costs. Geographically, the costs and benefits may or may not coincide in one place. In the case of automobile accidents or pharmaceutical side effects, for example, the zone of maximum benefits from the technology and the zone of maximum risk of its hazards coincide.

Zones at risk from atmospheric or water pollution usually contain, but extend well beyond, the place that economically benefits from production. In contrast, the hazards of agricultural technology are usually felt most by farm workers, but the benefits extend to a much wider area. Toxic wastes, radioactive wastes of all kinds, and their transportation form a class by themselves because the maximum risk is usually borne by a place far removed from the zone of maximum benefit. For example, malfunctions at the Chernobyl nuclear power plant endangered the

local populations as far north as Scandinavia, and most of the chemical industry that fills South Carolina's major toxic waste dump is located in and benefits other states.

Whether the noxious facility at issue is a permanent, national, high-level radioactive waste dump or a mental hospital, such facilities must be located somewhere (Elliott & Taylor, 1996). One research question is this: How do people perceive the risk of noxious facilities and react to them? A team of geographers and others at McMaster University in Canada (Elliott et al., 1993; Eyles, Taylor, Johnson, & Baxter, 1993) discovered that Canadians residing in southern Ontario took for granted proximate solid waste disposal facilities ("just a part of the landscape") because these sites had existed for an extended time without incident. It is not that people were unconcerned about risks to safety and property value. Indeed, across the three sites studied by this research team, levels of reported concern ranged from a low of 28% (existing solid waste incinerator) to 67% (existing solid waste landfill) to 74% (proposed solid waste landfill). However, those interviewed perceived that risks were beyond their control: "uncertain, perhaps invisible, sometimes stigmatizing impacts of social and environmental change" (Eyles et al., 1993, p. 811).

A longitudinal analysis of one of the original sites studied by the Canadian research team confirmed these earlier findings (Elliott et al., 1997). Although levels of reported concern about a *proposed* solid waste disposal facility were relatively high (74% of residents within 4.5 kilometers of the site), levels of reported concern decreased significantly over time (measured at the time of landfill construction and 2½ years after operation began), while levels of unsolicited "dislike" increased significantly. It would appear, therefore, that residents in close proximity to the site have learned how to cope with the existing landfill in their midst. This in no way implies that area residents are happy about the situation. As one respondent put it, "It's not because you like it any better; when there's nothing you can do about it, you just accept it and go on" (quoted in Elliott et al., 1997, p. 236).

At what scale should democratic control of technology be exercised? Noxious facilities are usually located where people have the least political power. For urban facilities, this usually means areas where ethnic minorities or the poor live, rather than upper-class suburbs. It makes sense to locate hazardous facilities in areas of sparse population, if the geology is compatible. The sparseness of population, however, also means that there are fewer people to oppose the will of the numerous people elsewhere who have reaped the benefits.

Eyles (1997) sets out a critical and theoretical environmental health research agenda, which includes the language and perception of environmental risk in everyday life; concepts of structure and power; consensus and conflict in environmental health policy; and questions about definitions of environmental quality. This agenda has a sociopolitical dimension because it pits property owners, workers, consumers, and others concerned about health risk against economic entities and government agencies. It also has an ethical dimension because it involves balancing individual desires to eliminate involuntarily imposed risks (however small) with society's need to have reliable technology at an affordable price. Geographic research cannot afford to ignore these sociopolitical and ethical concerns.

CONCLUSION

The hazards faced by urban and rural areas, or by old industrial countries and NICs, are converging as technologies and pollution spread. As many industries decentralize and spread to other countries, so do certain hazardous occupational exposures. As trucks and automobiles, insecticides and herbicides, radios and air conditioners become universal, so do the products of their manufacture. Less than a decade after the 1978 declaration in the United States of the first national emergency due to technological rather than natural causes (at Love Canal) came the disasters of thousands of deaths from the release of poisonous gas by a pesticide plant in Bhopal, India, in 1985, and the catastrophe at Chernobyl in 1986. Agricultural intensification; proliferation of industrial occupations; production of toxic and radioactive wastes; and consumption of petrochemicals, pharmaceuticals, synthetic textiles, electronics, glues, and paints are now almost universal. The scale of environmental change does not match the scale of contemporary regulation and control.

The ultimate causes of our environmental problems are those of population numbers, inequalities, and diminishing natural resources. As humans multiply, they seem to be oblivious to the fact that resources are finite, and that everything on earth has not been placed here explicitly for their benefit. Human gains have been achieved at great costs: depletion of material resources and extinction of many other forms of life. The greatest "gains" have accrued disproportionately to the wealthy nations and industrial organizations, who have been best able to marshal technology. Tropical populations, though not wealthy by Western standards, have also exacted their cost on the environment. Over the past few decades, intensive logging and burning of forests worldwide, partly to open up land for agriculture, have contributed to air pollution and micro-climatic change. Unfortunately, technology has not been employed as effectively as it might have been to reduce pollution of the environment. To effectively reduce emissions, nations must adopt policy measures such as energy or fuel efficiency standards for appliances and transport vehicles; they must also eliminate market distortions such as low, subsidized energy prices that reduce the economic benefits to consumers of saving energy (World Resources Institute, 1996, pp. 315–338).

REFERENCES

- Abelson, P. (1994). Chlorine and organochlorine compounds. *Science*, 265, 1155.
- American Lung Association. (1996). *Health effects of outdoor air pollution*. Washington, DC: Author.
- Bailey, A. J., Sargent, J. D., Goodman, D. C., Freeman, J., & Brown, M. J. (1994). Poisoned landscapes: The epidemiology of environmental lead exposure in Massachusetts children 1990–1991. *Social Science and Medicine*, 39, 757–766.
- Bentham, G. (1991). Chernobyl fallout and perinatal mortality in England and Wales. *Social Science and Medicine*, 33, 429–434.
- Berg, M., Tran, H. C., Nguyen, T., Viet Pham, H., Schertenleib, R., & Giger, W. (2001). Arsenic contamination of groundwater and drinking water in Vietnam: A human health threat. *Environmental Science and Technology*, 35(13), 2621–2626.

- Blumenthal, D. S., & Ruttenber, A. J. (1995). *Introduction to environmental health* (2nd ed.). New York: Springer.
- Brodeur, P. (1989, June 12, 19, 26). Annals of radiation: The hazards of electromagnetic fields. *The New Yorker*, pp. 51–89, 47–73, 39–68.
- Brodeur, P. (1990, November 19). Department of amplification. *The New Yorker*, pp. 135–150.
- Centers for Disease Control and Prevention. (CDC). (1991). Perspectives in disease prevention and health promotion: Toxic air pollutants and noncancer health risks—United States. *Morbidity and Mortality Weekly Report*, 40(17), 278–279.
- Centers for Disease Control and Prevention. (CDC). (1992). Blood lead levels among children in high-risk areas—California, 1987–1990. *Morbidity and Mortality Weekly Report*, 41(17), 291–294.
- Centers for Disease Control and Prevention. (CDC). (1994). *National Health and Nutrition Examination Survey (NHANES III)*. Atlanta, GA: Author.
- Cutter, S. L., & Solecki, W. D. (1989). The national pattern of airborne toxic releases. *Professional Geographer*, 41, 149–161.
- Cutter, S. L., & Tiefenbacher, J. (1991). Chemical hazards in urban America. *Urban Geography*, 12, 417–430.
- Dherani, D., Pope, D., Mascarenhas, M., Smith, K. R., Weber, M., & Bruce, N. (2008). Indoor air pollution from unprocessed solid fuel use and pneumonia risk in children aged under five years: A systematic review and meta-analysis. *Bulletin of the World Health Organization*, 86, 390–398.
- Dockery, D. W., Arden Pope, C., III, Xu, X., Spengler, J. D., Ware, J. H., Fay, M. C., et al. (1993). An association between air pollution and mortality in six U.S. cities. *New England Journal of Medicine*, 329, 1753–1759.
- Doyle, R. (1997). Air pollution in the United States. *Scientific American*, 276, 27.
- Dubos, R. (1965). *Man adapting*. New Haven, CT: Yale University Press.
- Earickson, R. J., & Billick, I. H. (1988). The areal association of urban air pollutants and residential characteristics: Louisville and Detroit. *Applied Geography*, 8, 5–23.
- Economy, E. C. (2004). *The river runs black: The environmental challenge to China's future*. Ithaca, NY: Cornell University Press.
- Elliott, S. J., & Taylor, S. M. (1996). Worrying about waste: Diagnosis and prescription. In D. Muntton (Ed.), *Hazardous waste siting and democratic choice: The NIMBY phenomenon and approaches to facility siting* (pp. 290–318). Washington, DC: Georgetown University Press.
- Elliott, S. J., Taylor, S. M., Hampson, C., Dunn, J., Eyles, J., Walter, S., et al. (1997). “It’s not because you like it any better . . .”: Residents’ reappraisal of a landfill site. *Journal of Environmental Psychology*, 17, 229–241.
- Elliott, S. J., Taylor, S. M., Walter, S., Stieb, D., Frank, J., & Eyles, J. (1993). Modelling psychosocial effects of exposure to solid waste facilities. *Social Science and Medicine*, 37, 791–804.
- Emch, M. E. (1999). Diarrheal disease risk in Matlab, Bangladesh. *Social Science and Medicine*, 49, 519–530.
- Eyles, J. (1997). Environmental health research: Setting an agenda by spinning our wheels or climbing the mountain. *Health and Place*, 3, 1–13.
- Eyles, J., Taylor, S. M., Johnson, N., & Baxter, J. (1993). Worrying about waste: Living close to solid waste disposal facilities in southern Ontario. *Social Science and Medicine*, 37, 805–812.
- Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L., & Colford, J. M. (2005). Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: A systematic review and meta-analysis. *Lancet Infectious Diseases*, 5(1), 42–52.
- Florig, H. K. (1992). The potential health effects of electric and magnetic fields. *Science*, 257, 468.
- Flynn, J., Kasperson, R. E., Kunreuther, H., & Slovic, P. (1997). Overcoming tunnel vision: Redirecting the U.S. high-level nuclear waste program. *Environment*, 39, 6.
- French, H. F. (1997). Learning from the ozone experience. In L. R. Brown et al. (Eds.), *State of the world, 1997* (pp. 151–171). New York: Norton.

- Gould, P. R. (1990). *Fire in the rain*. Baltimore: Johns Hopkins University Press.
- Greenberg, M. R., & Schneider, D. (1996). *Environmentally devastated neighborhoods*. New Brunswick, NJ: Rutgers University Press.
- Hilts, P. J. (1995, March 10). Deadly toll is found in dirty-air cities. *The New York Times*, p. A1.
- Hunter, J. M. (1977). The summer disease: An integrative model of the seasonality aspects of childhood lead poisoning. *Social Science and Medicine*, 11, 691–703.
- Hunter, J. M., & Arboña, S. (1995). Paradise lost: An introduction to the geography of water pollution in Puerto Rico. *Social Science and Medicine*, 40, 1331–1355.
- Johnson, C. J. (1981). Cancer incidence in an area contaminated with radionuclides near a nuclear installation. *Ambio*, 10, 176–182.
- Joint Monitoring Programme for Water Supply and Sanitation. (2006). *Meeting the MDG drinking water and sanitation targets: The urban and rural challenge of the decade*. Geneva: World Health Organization/UNICEF.
- Jones, A. P. (1998). Asthma and domestic air quality. *Social Science and Medicine*, 47, 755–764.
- Kloos, H. (1995). Chemical contaminants in public drinking water wells in California. In S. K. Majumdar, E. W. Miller, & F. J. Brenner (Eds.), *Environmental contaminants, ecosystems and human health* (pp. 30–43). University Park: Pennsylvania Academy of Science.
- Koskela, R., Mutanen, P., Sorsa, J., & Klockars, M. (2000). Factors predictive of ischemic heart disease mortality in foundry workers exposed to carbon monoxide. *American Journal of Epidemiology*, 152(7), 628–632.
- Last, J., & Guidotti, T. L. (1991). Implications for human health of global ecological changes. *Public Health Review*, 18, 49–67.
- Leary, W. E. (1994, September 6). Studies raise doubts about need to lower home radon levels. *The New York Times*, p. B-7.
- Lindley, D. (1988). CFCs cause part of global ozone decline. *Nature*, 323, 293.
- Maantay, J. (2007). Asthma and air pollution in the Bronx: Methodological and data considerations in using GIS for environmental justice and health research. *Health and Place*, 13, 32–56.
- MacKie, R., & Rycroft, M. J. (1988). Health and the ozone layer. *British Medical Journal*, 297, 369–370.
- McCoy, K., Bortnick, L. J., Campbell, C. M., Hamlin, H. J., Guillette, L. J., Jr., & St. Mary, C. M. (2008). Agriculture alters gonadal form and function in the toad *Bufo marinus*. *Environmental Health Perspectives*, 116, 1526–1532.
- National Center for Environmental Health. (1996). *1995 year in review*. Washington, DC: U.S. Department of Health and Human Services.
- National Research Council (NRC). (1975). *Long-term worldwide effects of multiple nuclear-weapons detonations*. Washington, DC: National Academy of Sciences.
- National Research Council (NRC). (1999). *Health effects of exposure to radon*. Committee on Health Risks of Exposure to Radon (BEIR VI). Board on Radiation Effects Research. Commission on Life Sciences. Washington, DC: National Academies Press.
- National Research Council (NRC). (2004). *Air quality management in the United States*. Washington, DC: National Academies Press.
- National Research Council (NRC). (2006). *State and federal standards for mobile-source emissions*. Washington, DC: National Academies Press.
- Our dirty air. (1989, June 12). *Newsweek*, p. 49.
- Sahsuaroglu, T., Jerrett, M., Sears, M. R., McConnell, R., Finkelstein, N., Arain, A., et al. (2009). Spatial analysis of air pollution and childhood asthma in Hamilton, Canada: Comparing exposure methods in sensitive subgroups. *Environmental Health*, 8, 14.
- Schneider, D., Greenberg, M. R., Donaldson, M. H., & Choi, D. (1993). Cancer clusters: The importance of monitoring multiple geographic scales. *Social Science and Medicine*, 37, 753–759.
- Shabecoff, P. (1985, June 11). U.S. calls eleven toxic air pollutants bigger threat indoors than out. *The New York Times*, p. A1.
- Small, S., van der Meer, J., & Upshur, R. E. G. (2001). Acting on an environmental health disaster: The case of the Aral Sea. *Environmental Health Perspectives*, 109(6), 547–549.

- Smith, A., Lingas, E., & Rahman, M. (2000). Contamination of drinking-water by arsenic in Bangladesh: A public health emergency. *Bulletin of the World Health Organization*, 78(9), 1093–1103.
- Smith-Sivertsen, T., Diaz, E., Pope, D., Lie, R. T., Diaz, A., McCracken, J., et al. (2009). Effect of reducing indoor air pollution on women's respiratory symptoms and lung function: The RESPIRE randomized trial, Guatemala. *American Journal of Epidemiology*, 170, 211–220.
- Sooman, A., & Macintyre, S. (1995). Health and perceptions of the local environment in socially contrasting neighbourhoods in Glasgow. *Health and Place*, 1, 15–26.
- Stephens, E. R. (1987, February 17). Smog studies of the 1950s. *Eos*, pp. 1–5.
- Stern, F. B., Halperin, W. E., Hornung, R. W., Ringenburt, V. L., & McCammon, C. S. (1988). Heart disease mortality among bridge and tunnel officers exposed to carbon monoxide. *American Journal of Epidemiology*, 128(6), 1276–1288.
- Stolarsky, R. (1988). The Antarctic ozone hole. *Scientific American*, 258, 30–36.
- Strauss, W., & Mainwaring, S. J. (1984). *Air pollution*. London: Arnold.
- Sun, M. (1986). Ground water ills: Many diagnoses, few remedies. *Science*, 232, 1490–1493.
- Talcott, F. W. (1992, Winter). How certain is that environmental risk estimate? *Resources*, 107, 10–15.
- Turk, J. (1985). *Introduction to environmental studies* (2nd ed.). Philadelphia: Saunders.
- Upton, A. (1982). The biological effects of low-level ionizing radiation. *Scientific American*, 246, 41–49.
- U.S. Census Bureau. (1996). *Statistical abstract of the United States: 1996* (116th ed.). Washington, DC: U.S. Government Printing Office.
- U.S. Department of Housing and Urban Development. (1990). *Comprehensive and workable plan for the abatement of lead-based paint in privately owned housing: Report to Congress*. Washington, DC: Author.
- U.S. Environmental Protection Agency (EPA). (1992). *Respiratory health effects of passive smoking*. Washington, DC: Author.
- U.S. Environmental Protection Agency (EPA). (1994). *National water quality inventory, 1994 report to Congress*. Research Triangle Park, NC: Author.
- U.S. Environmental Protection Agency (EPA). (1996). *National air quality and emissions trends report, 1995*. Research Triangle Park, NC: Author.
- Watts, J. (2005). China: The air pollution capital of the world. *Lancet*, 366(9499), 1761–1762.
- World Bank. (1994). *World development report*. New York: Oxford University Press.
- World Health Organization (WHO). (2008). Chronic respiratory diseases program. Retrieved from www.who.int/respiratory/en
- World Resources Institute. (1994). *World resources 1994–95*. New York: Oxford University Press.
- World Resources Institute. (1996). *World resources 1996–97*. New York: Oxford University Press.
- Zeigler, D. J., Johnson, J. H., & Brunn, S. D. (1983). *Technological hazards*. Washington, DC: Association of American Geographers.
- Zhang, J., & Smith, K. R. (2007). Household air pollution from coal and biomass fuels in China: Measurements, health impacts, and interventions. *Environmental Health Perspectives*, 115, 848–855.

FURTHER READING

- Briggs, D. J., & Elliott, P. (1995). The use of geographical information systems in studies on environment and health. *World Health Statistics Quarterly*, 48, 85–94.
- Brown, L. R., et al. (Eds.). (1999). *State of the world*. Washington, DC: Worldwatch Institute.
- Burton, I., Kates, R. W., & White, G. F. (1993). *The environment as hazard* (2nd ed.). New York: Guilford Press.
- Croner, C. M., Sperling, J., & Broome, F. (1996). *Geographic information systems: New perspectives*. New York: Guilford Press.

- tives in understanding human health and environmental relationships. *Statistics in Medicine*, 15, 1961–1977.
- Hall, B., & Kerr, M. L. (1992). *1991–1992 green index: A state-by-state guide to the nation's environmental health*. Washington, DC: Island Press.
- Hester, R. E., & Harrison, R. M. (1998). *Air pollution and health*. London: Royal Society of Chemistry.
- Monmonier, M. (1997). *Cartographies of danger: Mapping hazards in America*. Chicago: University of Chicago Press.
- Platts-Mills, T. (Ed.). (1999). *Asthma: Causes and mechanisms of this epidemic inflammatory disease*. Boca Raton, FL: CRC Press.
- Pollock, S. T. (1993). *The atlas of endangered places (environmental atlas)*. New York: Facts on File.
- Waller, L. A. (1996). Geographic information systems and environmental health. *Health and Environment Digest*, 9, 85–88.

CHAPTER 8

Political Ecology of Noncommunicable Diseases

The United States, and the world, are getting older. As we have seen in Chapter 5, with the demographic transition and the age structure changes that result from falling fertility (fewer children), the primary causes of death and the important chronic diseases have changed from infectious to noncommunicable. A few years ago the leading cause of death in the world shifted from diarrhea to heart disease. This was because China, which lowered its fertility 30 years ago, is now experiencing the shift in age structure as its enormous population grows older.

Infectious agents of disease are continuing to emerge (or reemerge) nevertheless. Most convincingly, a species of bacteria, *Helicobacter pylori*, has been shown to be the agent of chronic gastritis and ulcers. Thriving in the acid of the stomach, this bacterium infects 20–50% of people in developed countries and higher proportions in developing ones; yet most people remain without symptoms even when infected from childhood. It would appear that the stresses that have been associated with ulcers again act through the endocrine and immune systems to reduce health (see Chapter 2). It also now seems likely that bacteria in the circulatory system determine where the plaques of cholesterol that narrow arteries and cause atherosclerosis will form. It seems that we have more viral and bacterial DNA in our flesh and blood than we do human! The functions and impacts of this DNA are still unknown. A major new center combining the medical schools of the University of North Carolina and Duke University has just been established to investigate the role of dental ecology and hygiene in the introduction of such systemic viruses and bacteria (e.g., through gum disease). Paul Ewald (2000) makes a strong argument that microorganisms are involved in many of our chronic and degenerative diseases, and that they constitute a major (if not *the* major) cause of inflammation reactions in many organs of our bodies as we age. Finally, even helminths are now being considered in relation to

chronic diseases. One new research direction is developing the hypothesis that the intestinal helminths that have been part of us for hundreds of thousands of years may now cause imbalance in our immune systems by their absence—an imbalance expressed as autoimmune diseases such as inflammatory bowel disease or Crohn's disease. (This is part of the hygiene hypothesis; again, see Chapter 2.) The present chapter does not pursue these infectious leads, but they need to be noted.

Many things make the geographic study of noncommunicable diseases difficult. The older age structure and long-standing mobility of the population mean that exposures to hazards are difficult to trace or determine. The latency periods of these diseases may last decades. Often exposure has occurred in a different place from where the disease becomes manifest. Most of the diseases are manifested a long time after the susceptible persons are exposed to the risk (insult). Trying to understand the spatial pattern of chronic and degenerative diseases is thus very difficult. Careful research design, based on knowledge of the biological disease process as well as the spatial relationships of the units of observation, is absolutely essential in geographic research on these complex disease patterns. Nonsense correlations occur frequently with multiple variables and large data sets; there must be a plausible biological pathway for cause to have effect. Categories of chronic and degenerative diseases are often hard to define precisely. Sophisticated radiology and laboratory procedures may not be available for diagnosis. Terminology fads mean that definitions change. Cause is also often difficult to determine because we do not know whether there is a dose or a threshold relationship between disease expression and original hazards. The greatest impediment to research of all is that most of our reliable data are for mortality, not for chronic incidence or for morbidity that has been overcome. Prostate or breast cancer that is successfully treated is not a cause of death. A person killed in a car accident does not die from his or her advanced arterial occlusion, rheumatoid arthritis, or other chronic condition, and so it usually is not recorded that the person had the condition. There are many epistemological and methodological problems with scale of analysis, units used as observations, spatial autocorrelation, confounding, and other issues discussed in other chapters. This chapter also looks (too briefly) at epidemiological design, causal reasoning, and diseases of unknown etiology.

Chronic, degenerative, and noncommunicable diseases are *multifactorial*; that is, they have many causes, and the agents of disease have many effects. These do not occur randomly in a population. In Chapter 1, we have stated that *political ecology* is concerned with how higher-level socioeconomic and political structures and processes create the context in which local cultural and individual behavior interacts with the environment. Noncommunicable diseases take decades to develop and are strongly influenced by the social structures and economies that are the context of population and cultural interactions with the environment.

There is a remarkable geography to noncommunicable diseases. This chapter looks at how age, sex, race, poverty, and class, interacting at different spatial scales, give pattern to the outcome of individual cultural ecologies. It uses the perspective of political ecology to look at patterns of unequal mortality and their causes

at the world, national, and state scales. The *poverty syndrome* is discussed, as are risks of race and of female gender. Then a section examines epidemiological design and methodology. With reasoning clarified, the political ecologies of cancers and cardiovascular disease (CVD) are discussed, and several studies of these diseases with different spatial methodologies are examined. Finally, analytic approaches to diseases of unknown etiology and their difficulties are explained, using Parkinson's disease as an example. A discussion of the *precautionary principle* and its relevance to putting environmental factors back into studies of noncommunicable diseases ends this chapter.

THE DIMENSION OF MORTALITY

World Level: Europe versus Africa

There are over 57 million human deaths in this world each year. Table 8.1 lays out the importance of CVD (heart diseases, stroke), malignant neoplasm (cancer), chronic obstructive pulmonary disease (COPD; e.g., emphysema, bronchitis), and diabetes in populations like Europe's, late in the demographic transition. As fertility rates fall and scores of countries approach the end of the demographic transition, these diseases are becoming epidemic in them. In the table, the two poles of the transition are represented by "early" Africa and "late" Europe. Noncommunicable diseases now kill almost 60% of the world's dead, but in countries with an older age structure, their toll is almost 90%. Of those noncommunicable diseases, CVD and cancer account for more than 70%. Any change in behavior or risk that affects those diseases changes world mortality.

TABLE 8.1. Deaths by Cause for World, Africa, and Low-Mortality Europe

	Low-mortality stratum of			
	World	% total	Europe	Africa
Total deaths	57,029,000	100.0	3,920,000	10,664,000
			<u>% total</u>	<u>% total</u>
Noncommunicable disease	33,537,000	58.8	89.0	21.1
Cardiovascular disease	16,733,000	29.3	41.1	7.8
Ischemic heart disease	7,206,000	12.6	17.1	3.1
Cerebrovascular disease	6,509,000	9.7	10.6	3.4
Malignant neoplasm (cancer)	7,121,000	12.5	26.5	3.8
COPD (lung disease)	2,748,000	4.8	3.6	1.1
Diabetes	968,000	1.7	2.3	0.8
Unintentional injury	3,551,000	6.2	3.5	4.6
Road accident	1,192,000	2.1	1.2	1.8

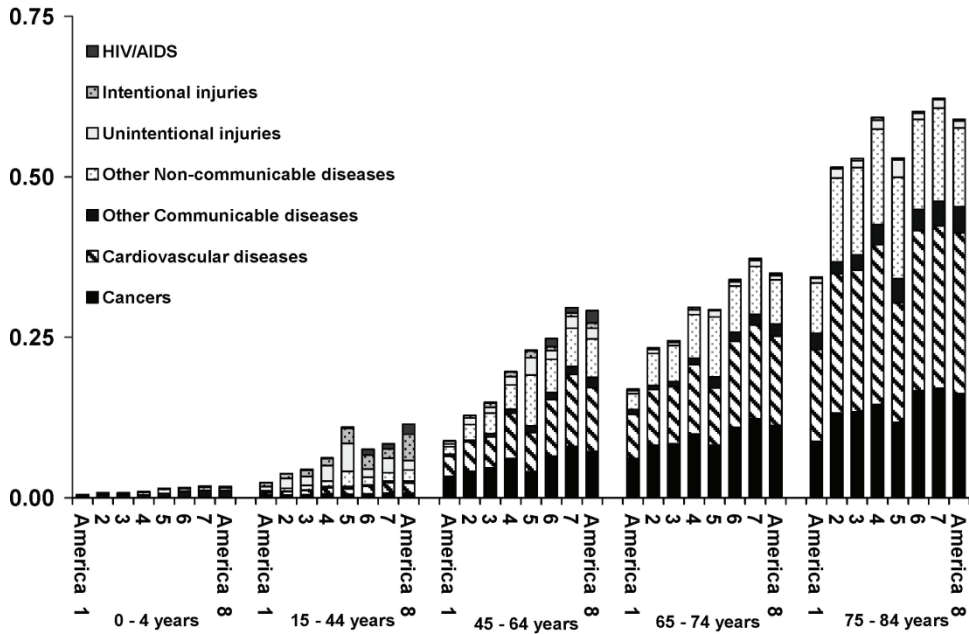
Note. Data from WHO (2004b). The data are from 2002, when the world population was 6,224,985,000.

National Level: The United States

In 2005 the U.S. population had a life expectancy at birth of 77.8 years. This reflected an infant mortality rate down to only 6.87 (per 1,000 births), and a death rate of 825.9 per 100,000 (or, age-adjusted on the U.S. standard population because the age structure is becoming older year by year, 798.8; this was a record low). In the simplest terms, 2,448,017 people died that year. Heart disease killed 26.6%, and cancer killed 22.8%, for a combined total of almost half of all deaths. Stroke and chronic obstructive upper respiratory disease (COPD) each killed a little over 5%, and accidents almost 5%. So these five causes of death claimed about two-thirds. The next five causes claimed only 14%: diabetes, Alzheimer's disease, influenza/pneumonia, kidney disease, and septicemia, in order. These and other causes of death were far from equal among races, sexes, classes, and geographic regions, however. A long-standing pattern continued, based on age-adjusted rates, of male death rates being twice those of females for the five leading causes. Firearms also killed 30,694 people, mostly male. The most remarkable trends in U.S. death rates over the last 20 years have been the remarkable decline in CVD death rates, and the growing disparities at the county level. In a detailed analysis of county-level mortality data from 1961 to 1999, Ezzati, Friedman, Kulkarni, and Murray (2008) found that mortality because of CVD declined everywhere. But there was a decline in life expectancy (i.e., increased mortality) after 1983 in 11 counties for men and 180 counties for women, caused by increased mortality rates from lung cancer, COPD, diabetes, and other noncommunicable diseases that were no longer compensated for by the decline in CVD. Higher HIV/AIDS and homicide deaths also contributed to a decline in life expectancy for men, but not women. Ezzati and colleagues concluded that there was a steady increase in mortality inequality from 1983 to 1999 at the county level, and that this resulted from increased mortality among the worst-off segment of the population.

A later study, also at the county level, took another approach and examined causes of death by age, sex, race, ethnicity, location, and county-specific measures of population density, per capita income, and cumulative homicide rate (Murray et al., 2006). For each of 3,141 counties, the researchers estimated life expectancy and the risk of mortality from specific diseases, as well as considering health insurance and health care utilization. They found a gap of 20.7 years in life expectancy at birth between 3.4 million high-risk urban black males and 5.6 million Asian females in 2001! The order of life expectancy among sex and racial groups (discussed later) remained unchanged from 1982 to 2001. The largest disparities by age were for young (15–44) and middle-aged (45–50) adults. These were caused primarily by chronic diseases and injuries with well-established risk factors. Murray and colleagues (2006) concluded that “disparities among tens of millions of Americans are enormous by all international standards,” and that the “observed disparities cannot be explained by race, income, or basic health care access and utilization alone.” Instead, they called for public health strategies and national policies aimed at reducing fundamental socioeconomic inequalities—policies they found practically absent in the United States. A graphic representation from their article of probability of dying by cause, by age, and by sex is presented in Figure 8.1. The figure clearly

Males



Females

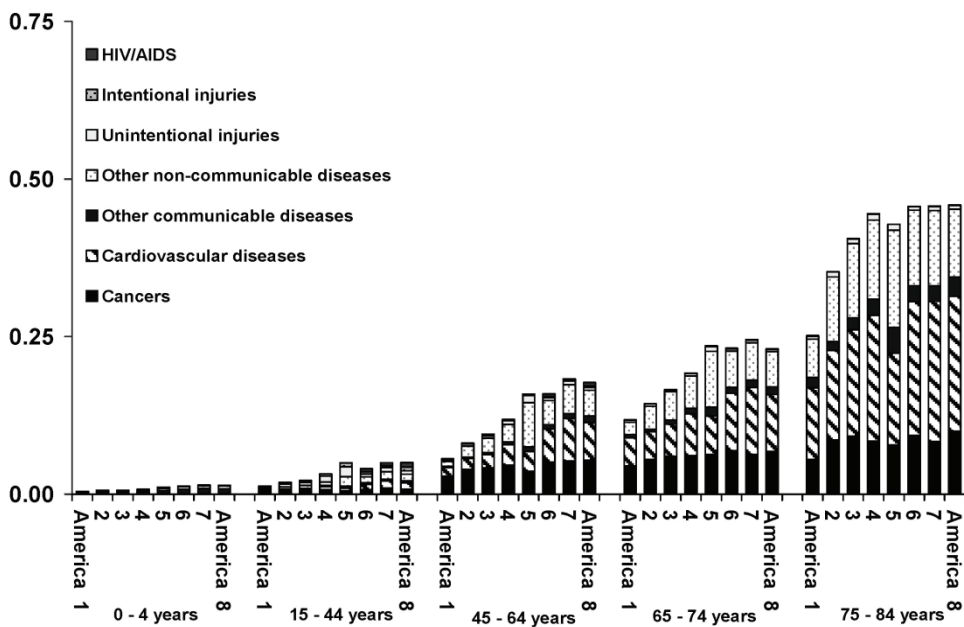


FIGURE 8.1. “Eight Americas”: U.S. probability of dying by age, sex, and disease group. From Murray et al. (2006). Reprinted by courtesy of *PloS Medicine*.

shows the impact of AIDS and injury on young men, and the impact of suicide and pneumonia on the oldest groups. Notice especially the divergence among the eight bars graphed for each age group. These bars are for the eight subgroups of the population identified in the study, which were based on the 8,288 race–county units the researchers constructed and merged into 72 units (to avoid small numbers and boundary changes).

Murray and colleagues (2006) called these eight subgroups of the U.S. population “eight Americas.” They are summarized in Table 8.2. Asians had sustained over two decades a life expectancy >5 years higher than that of rural whites in the north, and about 7 years higher than that of middle America (of which 98% was white). Native Americans had a life expectancy comparable to or higher than that of middle America unless they lived on or near Western reservations in counties of the lowest income class. Probably the most interesting finding was the wide range of mortality within the white population that was not primarily related to income. Whites in “America 2” lived in the rural Dakotas, Montana, Nebraska, Iowa, and Minnesota and had a life expectancy of 76.2 years for males and 81.8 for females. Despite similar income levels, whites in Appalachia and the Mississippi Valley could expect to live 4.2 (males) and 3.8 (females) years less, respectively (Murray et al. note that life expectancy is “equal to that of Mexico”); this is rather comparable to the gap of 6.4 (males) and 4.6 (females) years between whites and blacks in the United States as a whole.

State Level: North Carolina

As one goes down in scale, greater detail and more intricate connection of information usually become possible. This is the case in the state of North Carolina, in which a special report on racial and ethnic differences (North Carolina Department of Health and Human Services, 2004) and a random telephone survey first conducted over 2 years (2001–2003) and updated annually (North Carolina State Center for Health Statistics, 2003, 2007) concentrated on risk factors. In Table 8.3, the 10

TABLE 8.2. The “Eight Americas”

American grouping by life expectancy ranking	Population (millions)	\$ income per capita	% completing high school
1. Asian (in counties <40% Pacific Islanders)	10.4	21,566	80
2. Northland low-income rural whites	3.6	17,758	83
3. Middle America ^a	214.0	24,640	84
4. Low-income whites in Appalachia and the Mississippi Valley	16.6	16,390	72
5. Western Native Americans	1.0	10,029	69
6. Black middle America	23.4	15,412	75
7. Southern low-income rural blacks	5.8	10,463	61
8. High-risk urban blacks	7.5	14,800	72

Note. From Murray et al. (2006). Adapted courtesy of *PloS Medicine*.

^aWhites not in Americas 2 and 4; Asians not in America 1; Native Americans not in America 5.

TABLE 8.3. North Carolina's Leading Causes of Death

Cause ^a	Deaths	Rank	Whites' AADR	Rank	Blacks' AADR
Heart disease	75,969	1	240.1	1	295.9
Cancer	63,709	2	194.8	2	245.7
Female breast ^b		1	148.9	2	130.6
Prostate ^b		2	135.4	1	216.8
Lung ^b		3	71.2	3	65.2
Colon/rectal ^b		4	48.1	4	52.6
Stroke	21,921	3	68.3	3	96.5
Chronic lung disease	14,402	4	50.2	6	31.8
Pneumonia/influenza	7,481	5	25.2	7	25.2
Diabetes	8,510	6	21.5	4	55.6
Motor vehicle injury	6,404	7	19.2	8	20.7
Kidney disease	5,204	8	13.4	5	34.9

Note. AADR, age-adjusted death rates.

^aData from North Carolina Department of Health and Human Service (2004, Table 5), except as noted.

^bData from North Carolina Department of Health and Human Service (2004, Table 3).

major causes of death are ranked and age-adjusted death rates given for whites and blacks. Although the Hispanic/Latino population of the state increased between 1990 and 2000 from 77,000 to 379,000, and for 2007 was estimated by the U.S. Census Bureau at 600,000 (6.7% of the state's 9 million population), that population is strongly concentrated in the young adult ages; there are too few numbers in all age and sex categories to analyze here because of the small-number problem. The leading cause of death for the Hispanic population is motor vehicle accidents. Native Americans in North Carolina live in two areas: The Cherokees are concentrated in the mountains near the Great Smokies (their pre-"Trail of Tears" homeland), in Cherokee County; and the Lumbee Indians, who have struggled for decades to get federal recognition as an official tribe, form an identifiable population in two southeastern counties in the state. Not only do Native Americans in age and sex divisions suffer from the small-number problem, but there have been discrepancies in recording race on cause of death and hospital records. So our analysis is again focused on blacks and whites.

The first three causes of death, heart disease, cancer, and stroke, are the same as for the national population; again, age-adjusted death rates for blacks are much higher. More specifically, breast cancer and lung disease mortality rates are higher for whites, but prostate cancer and colon cancer mortality rates are higher for blacks. After that, whites (70% of the state population) and blacks (21.4%) diverge. Emphysema is much more important for whites, and diabetes and kidney disease for blacks.

An increasingly important subdivision of the North Carolina population is characterized by presence of a disability (Center for Women's Health Research,

2007). Higher percentages of women ages 18+ with disabilities than without have high blood pressure (46% vs. 24%), are obese (36.7% vs. 22.3%), have high cholesterol (47% vs. 32%), and diabetes (17.7% vs. 4.7%). There are also strong regional patterns, which show up in every map. Four risk factors for which the coastal plain as a region has higher rates than the rest of the state are as follows: knowledge that folic acid helps prevent birth defects; no leisure activity; current cigarette smoker (30% vs. 24% in the piedmont); and never had a colonoscopy. Two risk factors for which the mountain region has higher rates are these: currently uses a pipe or cigar; and keeps firearms around the house.

All the causes of death listed above and discussed in the rest of this chapter are diseases with multiple causes and multiple outcomes. Their complex ecologies are discussed later. Complications resulting from, and involvement with, composition of the population by class, race, and gender need to be addressed first.

THE POVERTY SYNDROME

Many diseases seem to occur consistently more often among the poor than among the affluent. There seems to be a *sociopathological complex*, made up in large part of stress, lifestyle, diet, housing, polluted air, old paint, and old pipes. Physical, social, and mental diseases have similar patterns within urban areas. Some diseases, such as tuberculosis, are traditionally associated with the poor. Others, such as breast cancer, have more often afflicted the affluent. And some, such as lung cancer, have recently decreased at high socioeconomic status (SES) and increased at lower SES.

Many of the associations of health problems with poverty occur throughout the world. While noting the major improvements in health in the latter part of the 20th century, the World Health Organization's (WHO's) Commission on Social Determinants of Health (WHO, 2007) concluded that country averages conceal persistent and significant differences according to "social status" (defined as consisting of education, gender, and occupational hierarchy). As Harpham, Lusty, and Vaughan (1988) modeled them, a wide range of the diseases of the poor have universal associations characterized as direct, environmental, or psychosocial. Their direct factor included unemployment, low income, limited education, inadequate diet, and prostitution. Their environmental factor was composed of water quality, sanitation, overcrowding, poor housing, rubbish accumulation, lack of garden land, traffic, industrial hazards, pollution, and accidents. Their psychosocial one included stress, alienation, insecurity, depression, smoking, alcoholism, drugs, and abandoned children. Harpham et al. found these characteristic associations occurring from the slums of Bombay (Mumbai) to London.

There is a higher-order causality to these associations, a political economy, which in geography includes study of the relations of societal process with spatial form. Usually in the United States such studies address the inner city; rural poverty is rarely addressed. Dear and Wolch (1987) have described how "landscapes of despair" were formed when mentally ill people and others were "deinstitutionalized" in the late 1970s and released into unprepared and largely unfunded commu-

nities for service and support. At the turn of the last century, human services policy was spatially expressed in the location of public asylums in major cities. As the cities grew into industrial centers, the growth of downtown areas and suburbanization of people working there led to a clear identity of central business districts and their fringe zones as areas that housed noninstitutionalized, service-dependent populations (e.g., blocks of single-room-occupancy hotels in which elderly, poor, alcoholic, transient, and other populations congregated). The general philosophy of treatment and care of such people had become a “specifically urban phenomenon” across North America, a social construction of “the service-dependent ghetto.” In the United States in the early 1980s, community building, housing subsidy, and welfare funds were substantially reduced. People who were deinstitutionalized and did not return home to supportive families, churches, or other social groups gravitated toward “inner-city landscapes where increasing demands for assistance were met by diminished capacity to supply both shelter and services” (Dear & Wolch, 1987, p. 199). Over 20 years later, new issues for mental health are demanding attention: the impact of sprawl, the emergence of an independent scatter-site housing model for poor urban households, and in general the impact of place on community integration and trust (Yanos, 2007).

Homelessness—people’s not even having shelter—is a condition affected by economic, social, and political forces that get localized in specific places. People commonly become homeless because they lose a job and thereby the ability to pay rent. They may also flee physical abuse; lose a supportive relative (of their disability) to death or imprisonment; become incapacitated by substance abuse; or become susceptible for many other reasons. Some people are homeless for only short periods; some cyclically find and lose jobs/apartments/families; some become chronically homeless. Wolch and Dear (1993) describe the following dualism that results in poverty, with overwhelming health risks around the world. On one hand, there is rising economic marginality (of persons or groups) that results from economic restructuring, welfare state restructuring, and demographic change. On the other side, there is a decrease in affordable shelter because of national housing policy, urban housing markets, and loss of affordable units. The U.S. financial crisis and housing foreclosures in fall 2008 overwhelmed shelters (both human and pet) even before bitter cold set in.

Housing itself is probably the easiest part of the sociopathological complex to understand. The poor are often overcrowded. Older residences have often been subdivided into apartments for numerous families and individuals. Larger families mean that small children add to the high-room-density measure frequently associated with influenza, bronchitis, and tuberculosis. Old pipes may have lead solder and occasionally are composed of lead; they tend to be made of metals whose cadmium and other trace elements are easily eroded. Although lead is no longer allowed in house paint in the United States, older buildings still have lead paint dust.

Usually considered a Northern inner-city problem, in the rapidly growing South lead may be more of a rural risk (Hanchette, 1998). Figure 8.2 shows Hanchette’s spatial modeling of the risk of lead for children. In her research she identified the

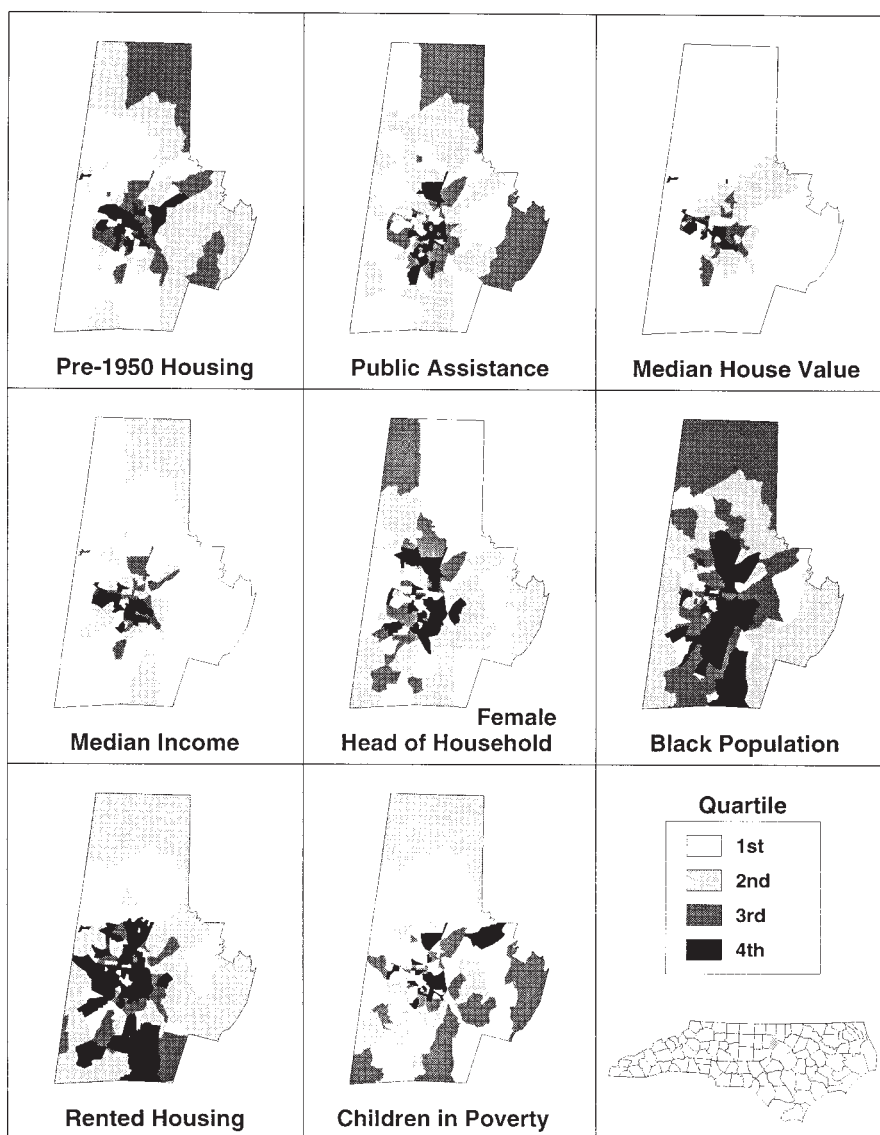


FIGURE 8.2. Lead poisoning risk and risk factors for Durham County block groups, classified by quartiles, and the predicted lead poisoning risk with overlay of lead poisoning incidents. From Hanchette (1998). Used by permission.

census variables that had been repeatedly statistically associated with high blood lead levels in schoolchildren. They were somewhat redundant: low income, rental housing, having a single mother, being black, neighborhood percentage of children in poverty, low median household value, public assistance (welfare), and housing older than 1950. In what is known (from physical geography) as *suitability mapping*, she mapped each of the variables in quartiles and used a geographic information system (GIS) to overlay them, so that each census tract could be classified as low-,

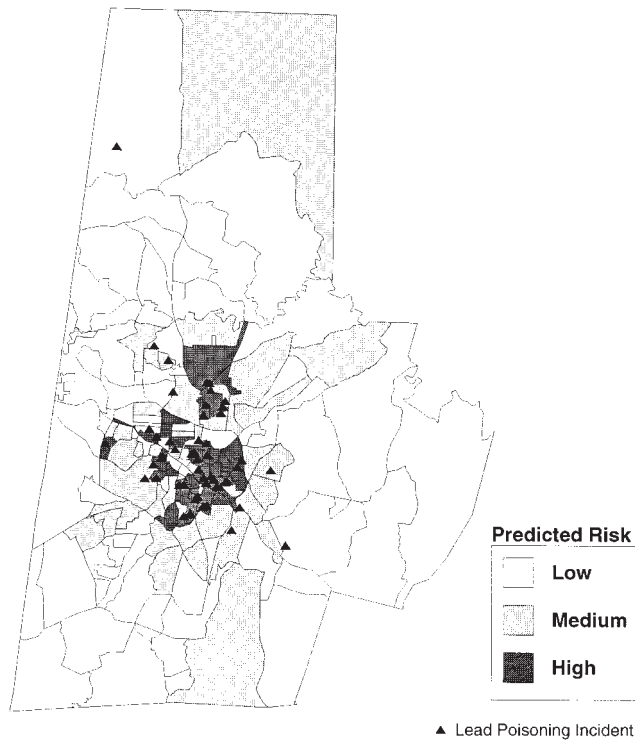


FIGURE 8.2. (*cont.*)

medium-, or high-risk according to the number of highest-quartile risk factors in it. The known incidents of high blood lead were then overlaid on the map of risk factors. All but a few incidents occurred in the highest-risk blocks; those few occurred in children who were going to school outside of their neighborhoods.

In mapping Durham, Raleigh, and other North Carolina cities, however, Hanchette came to realize how little pre-1950 housing (the main “paint chip” risk nationally) was present in North Carolina. Most of the in-migration, and the great increase in population, has taken place in the last 30 years in the cities. When she broadened her scope, Hanchette realized that the old housing with lead risk was in small towns, especially “down East” on the coastal plain. When the state looked there, North Carolina found high blood lead levels in black children living in the old, often historic houses of the small towns on the coast, where slavery had been strongest as an institution. Pursuing the political ecology, Hanchette (2008) found that in the 1960s as schools were integrated, whites fled to new suburbs and blacks moved into the old historic housing. Poverty prevented the painting, renovation, and other needed upkeep, so that black children came to be exposed to paint chips and lead. Exposed asbestos is a more recently recognized hazard that is still being removed from old buildings as each is renovated. Central heating systems are frequently obsolete or in poor repair and, besides failing to counteract drafts, contribute to mold, dust, and other allergens.

A diet high in starchy foods, sugar, and saturated fat seems in international comparisons to contribute to a greater tendency toward diabetes and coronary heart disease. Obesity, a major risk factor for both, is even more frequent among black females than among other groups composing the U.S. population (see the North Carolina mortality rates in Table 8.3) and more frequent among poor whites than among rich ones. In the past decade, however, it has been increasing in all population groups, including the college-educated. The role of diet is closely tied to lifestyle factors. When people in the United States started jogging and aerobic dancing for health reasons, these cultural practices, like most others, diffused down the socioeconomic scale from people with higher levels of access to information to those more tied to traditional ways. In Figure 8.1, look at the disproportionately higher contribution of cardiovascular disease to mortality among poor rural southern whites, poor inner-city urban blacks, and poor rural southern blacks.

Usually an innovation, or fad, diffuses from metropolitan areas in California and New York to other urban centers and out to small towns and rural areas (see Chapter 10). Diffusion also occurs among neighborhoods of a city and among blocks within neighborhoods. Knowledge about the benefits of steaming vegetables instead of overboiling them; the hazards of bacon fat, whole milk, salt, and beef; or the uses of exotic protein sources such as tofu spreads slowly. A new definition of a healthy diet seemed to appear among the educated just when dietary behavioral patterns once associated with the well-off became widely practiced among the poorly educated. There are also socioeconomic impediments to lifestyle change. Health spas, gyms, and safe jogging paths are not readily available to the poor. Night shifts, double jobs, and lack of child care impede many changes. Crime and fear of it can lead to deprivation of social life, especially for the elderly (see the discussion of heat waves in Chapter 6). Stress also results from anxiety over jobs, fear of inscrutable institutional policies, insecurity over Social Security or Aid to Dependent Children payments and medical care/insurance coverage, and hassles with landlords.

One way of viewing the whole sociopathological complex is in terms of the accumulation of insults to health (Chapter 2). Some people with high levels of health can cope with the insults and emerge stronger and more creative for the struggle. Others, their health depressed by psychological and social insults, are unable to rally from infectious, chemical, or further mental insults. They become part of the small percentage of the population that accounts for a large percentage of illness and repeated health care needs.

RACE IN THE STUDY OF HEALTH RISKS

As a genetic classification system, race has limited relevance to disease etiology and today is not found to be biologically accurate (see Chapter 2). However, a few genetically caused diseases, such as sickle cell anemia or Huntington's chorea, do seem to be linked with specific racial groups. In some other diseases, such as "essential hypertension" in blacks, a special susceptibility of one group or another seems to

be involved. In still others, such as prostate cancer, genetic susceptibility may be part of the etiological puzzle. Too often, however, automatic separation into racial categories masks a superficial socioeconomic or class-based, analysis that could be more etiologically significant.

African Americans are a heterogeneous population, and they constitute a minority (less than 13%) of the U.S. population. This means that although a higher *proportion* of blacks live in poverty, receive welfare, are single mothers, are unemployed teenagers, are high school dropouts, and so forth, substantially larger *numbers* of people who are in these categories are white. Millions of African Americans are college-educated professionals leading the lifestyles of the affluent. However, if even college-educated black women have lower-weight babies, is there a role of genetics aside from lower access to adequate prenatal care? Race is indubitably important as a social construction, a category with ascribed characteristics and meaning. If race is taken as a social construction that affects the behavior both of people within the group and of those outside toward it, is there another kind of cause? The work of Evans, Barer, and Marmor (1994) and those who have followed them, for example, strongly indicates that it is not absolute deprivation but relative status that affects health (as discussed in Chapter 2).

In the United States a higher proportion of minority group members are poor, but the majority of poor people are white. The main groups of Americans living in poverty are single mothers, the mentally and physically disabled, the aged, and children. There is a much smaller group of those afflicted with substance abuse. There has been a long-term trend of deterioration in the relative health circumstances of the poor versus the rich, the uneducated versus the educated, and blacks versus whites, expressed as increasing socioeconomic inequalities in mortality (Figure 8.1). Demographic projections make it obvious that there will be future increases in the proportions of old persons, disabled people, minority group members (not just blacks), and probably single mothers (not just never married, but divorced, abandoned, and widowed) living in poverty, as well as in the proportion of children doing so. The structure and pattern of statistical relations are complicated by their strong spatial variation, with concentrations differing among states, counties within states, and neighborhoods within cities and counties. Scales of analysis addressing “population” without place can easily miss significant associations or opportunities for intervention.

The well-recognized inverse relationship of SES and mortality has many underlying factors: causation by and risk from environment or behavior; public health prevention (or lack thereof); and health services (or lack therefore)—early detection, access to care, aggressive treatment, and social support. Besides genetic susceptibilities and socioeconomic associations, there are cultural differences in such things as perception of symptoms, compliance with medical regime, and family networking. Not only patients but also their health care providers can be influenced by belief systems and perceptions in ways that affect early detection, aggressive treatment, or simply recognition of problems. Separating the racial, socioeconomic, cultural, and institutional/policy components of the mortality differential is not simple. Never-

theless, as noted earlier in this chapter, Ezzati and colleagues (2008) found that an increase in mortality inequality across U.S. counties from 1983 to 1999 resulted from an increase in mortality among the worst-off segment of the population.

The differences in this area are not all socioeconomic or racial. As Figure 8.1 and the discussion of mortality in the “eight Americas” have shown, whites in the rural northland and in Appalachia/the Mississippi delta had the same race and comparable incomes, but divergent mortality rates. An earlier study (Geronimus, Bound, Waidmann, Hillemeier, & Burns, 1996) featured in the second edition of this text used standardized mortality ratios (see Chapter 5) to compare poorest and nonpoor areas and towns (e.g., Harlem vs. Queens, Watts vs. Los Angeles, central vs. northwest Detroit, and “black belt” Alabama vs. northern Alabama for blacks, and the Lower East Side of New York City vs. Queens, Detroit vs. Sterling Heights, northern vs. southwest Alabama, and Appalachia vs. western Kentucky for whites) from 1989 to 1991 and found shocking dimensions of inequality: The probability of a 15-year-old black male in Harlem surviving (in 1990) to age 65 was only 37%, whereas that of a white female in Sterling Heights, near Detroit, was 93%. For both sexes the death rate in each poverty area was excessive. But, as the study authors noted, the “black belt” in Alabama had the highest rate of poverty but the lowest excess mortality; Harlem had the lowest rate of poverty but the highest excess mortality. The social, economic, cultural, racial, and regional geography of mortality in the United States is *very* complex.

GENDER: WOMEN'S HEALTH

Male mortality is higher for all racial and ethnic groups and at every age than female mortality. Human ecology recognizes that the universal human sex ratio at birth (104–105 males for every 100 females) probably has evolved to counter the higher youthful male mortality and so to approach an equal sex ratio by the reproductive years. In earlier times, this was countered later by the high female mortality in pregnancy and childbirth. With the low fertility rates described (in Chapter 5) for the end of the demographic transition, women have been largely relieved of that stress and now have longer life expectancies than men—so much so that the sex ratios (males per hundred females) in nursing homes can approach 30 for populations over 80. The higher mortality rates of males have also, through numerous studies, been associated with greater resistance to seeing a doctor and admitting sickness or weakness. Women in turn have been described as having higher morbidity but less mortality because they do readily seek medical attention. In fact, in one public health survey after another, women rate their health worse than men from adolescence to late middle age, and they do visit doctors and even hospitals more often; yet at every age, women have lower mortality than men.

Women's health has become a focus of interest because in the past it was buried in the totality of health and disease, so that little was learned about it. Sometimes *men's* health was considered synonymous with *human* health. Clinical trials of the efficacy of pharmaceutical drugs were conducted only on men but assumed

to be valid for all. Sometimes studies of causation, prevention, and treatment of a major disease, such as heart disease, were focused on men because they had higher rates of that disease and the economy suffered greater loss as a result. Although such studies were cost-effective and the results were clearer and stronger precisely because the variability and confounding posed by including women's differences were avoided, nevertheless the results were considered applicable to the general population across all differences. Women are both the same as and different from men. Differences in health or treatment of disease are not simply due to money, power, and social construction of interpretation, but also due to gender. Examples involving women's health have been used throughout this text. This section focuses on how the concepts and approaches of this text are all relevant to women through the integrative framework of the triangle of health ecology (Figure 2.2), with its three vertices of (and interactions among) population, habitat, and behavior.

Population

Gender, genetics, age, immunological status, and nutritional status constitute the basic vertex of population. The genetics of being female result in a few gender-specific diseases, especially those of the reproductive organs. Female hormones are genetically given, of course, as well as the elaborately orchestrated capacity for pregnancy and lactation. Several genetic conditions are gender-linked at the chromosome level. For example, genes that cause a serious condition (hemophilia) and those that cause a mild condition (color blindness) both occur predominantly in males because their Y chromosome contains no counter gene to repress such recessive characteristics when they are carried on the longer X chromosome. The full expression of female, however, is gradually revealed through the life cycle.

Historically, as noted above, women died at high rates from the rigors and cumulative risks of pregnancy and childbirth. The recent reduction of fertility has done a great deal to spare their lives. When, as in India and China, young girls die at a higher rate than boys, it is usually taken as *prima facie* evidence of severe discrimination in nourishment, nurturance, and care—discrimination that on occasion even results in female infanticide.

Between menarche and menopause, women are biologically vulnerable to a range of insults associated with their reproductive role. These include such things as infection of the kidneys/urinary tract, scarring, and infertility resulting from sexually transmitted disease; nutritional shortages of iron, vitamin A, or calcium; and so on (Koblinsky, Timyan, & Gay, 1993). Besides blood lost to menstruation, malaria, and intestinal parasites, nutritional deficiencies are made worse by pregnancy, and greatly exacerbated by multiple ones. More than half the pregnant women in tropical Africa and Asia are anemic (which, in addition to the usual energy and health complications, can result in oxygen crisis for a mother and child at birth). Advanced vitamin A deficiency is usually monitored as xerophthalmia, but in women depleted by recurrent pregnancy, lower levels of deficiency may pass unnoticed while they damage the immune system. Even in developed countries, despite good nutrition and prenatal care, most embryos do not survive to become fetuses. Nevertheless, in

the developing world, crude and septic induced abortions are a significant cause of maternal mortality for women without means of birth control.

The high rates of maternal mortality evident in Table 8.4 illustrate the heavy burden that such mortality still exacts. As described in Chapter 4, the summary measures of *disability-adjusted life years* (DALYs) is used by the WHO to best estimate and portray the burden of disease. It is based on death and chronic disability compared to healthy years of life lost. In this case, DALYs are also reminders of the large number of children rendered motherless by the lack of birth control. The lifetime chances of dying from maternal mortality are given as proportions of women of childbearing age: For Afghanistan, the estimate is 1 in 8; for Japan, it is 1 in 11,600. The table portrays the 10 largest countries in the world, in various stages of the demographic transition.

Cyclic exposure to estrogen seems to affect cholesterol metabolism and act protectively against heart disease. On the other hand, cumulative estrogen exposure is one of the clearest insults increasing the risk of breast cancer. Hormonal fluctuation can affect mental health, especially postpartum depression, and can be a risk factor for suicide.

In the developed world, two-fifths of a woman's life is lived after menopause. Fertility rates of fewer than two children per couple also mean that most of a woman's life is not engaged in reproduction. Medicine and health research in general have been so preoccupied with women's reproductive processes, reproductive organs, and infections and diseases that little has been learned about their health in other respects. Even clinical trials of drugs were carried out only on men, as noted earlier; one reason for this was the expressed fear that a fetus might be harmed (some would say that this fear was a construction of the value of a woman's life). It has recently been learned that women's hearts, bones/calcium metabolism, and patterns of fat deposition are different from men's, and that so are the effects of relevant

TABLE 8.4. Impact Measures of Maternal Mortality, 2005 (Lifetime)

Country	DALYs ^a	Chance: 1 in . . . ^b
Bangladesh	728	51
Brazil	728	370
China	177	1,300
India	824	70
Indonesia	398	97
Japan	41	11,600
Nigeria	1,589	18
Pakistan	1,019	74
Russian Federation	117	2,700
United States	67	4,800

^aData from WHO (2004b).

^bData from Population Reference Bureau (2008 Data sheet) (in turn estimated by WHO, UNICEF, United Nations Fund for Population Activity, and World Bank).

drugs. Several large longitudinal studies, such as the Nurses' Health Study and the Women's Health Initiative, were undertaken in the late 20th century to research the effects on women of estrogen supplementation, calcium supplementation, various exercise regimens and diets, and common drugs such as aspirin. The Women's Health Initiative study was stopped several years early because of clear evidence that the estrogen supplementation then being recommended by doctors not only for menopausal symptoms, but in order to reduce heart attacks (as presumably the women's own estrogen previously did, since men had higher rates), actually increased the chance of stroke.

There are notable interactions of population with habitat and behavior. Customs about the seclusion of women and definitions of modesty affect diffusion of contagion, exposure to vectored disease, and even biometeorological influences. Rickets in Saudi Arabia, for example, is purely a cultural artifact of *purdah*. (On the other side of the world, more open advertisement about new products for menstruation created and spread a new toxic shock syndrome almost overnight.) Cultural beliefs also determine at what age marriage is suitable, and such things as whether the degree of inbreeding to cousins or outcrossing to other ethnic groups is tolerated. Values about desired family size, use of contraception, provision of medical services, accessibility of education to girls, job opportunities, and status/financial contribution from outside the home for women have all changed enormously, as expressed at the 1994 International Population and Development Conference in Cairo and the follow-up International Women's Conference in Beijing.

Increases in income, trade, and agricultural production, as well as the technology of food preservation, storage, shipment, and marketing, have changed nutrition enormously. One consequence in the developed world of continually abundant, high-protein, high-fat food has been the lowering of the age of menarche. It used to be 16 or more rather universally, but now in places it occurs at 11 and even younger (extremes possibly due to pseudoestrogens in food and water today), with consequences for reproductive behavior and attendant risks even as the age of marriage has been culturally raised to more than 20 in much of the world. Probably the most notorious cultural control on nutrition is the widespread custom of the husband's eating first and then the children. When at last the woman eats the leftovers, the most nutritious foods are gone, and adequate calories may not remain. In the developing world, when a mother's milk is changed to a bottled substitute, a baby's antibodies to disease and exposure to diarrheal agents also may change.

Habitat

The natural, built, and social components of human habitat affect women's health in ways already described for people in general, and yet somewhat differently than they do men's. Of course, this is the least studied and least understood aspect of women's health. There is different exposure to insults because of different roles. In the study of the Malaysian land development's impact on disease ecology, for example (see Vignette 4.1), women were exposed to scrub typhus from rat mites on settlement pathways, whereas men were out tending their tree crops before dawn, when

Anopheles mosquitoes were still biting. Mobility, jobs, and relationships are among the behavioral components that structure the exposure to habitat conditions.

Radon and other radiation occur naturally in many environments but may be a greater hazard for fetuses. Women suffer more from hay fever and chronic sinusitis caused by natural allergens. The chlorinated hydrocarbons and polychlorinated biphenyls (PCBs) distributed in the environment have an affinity for storage in fatty tissue, which puts women's bodily deposits, especially in breast tissue, especially at risk (Calabrese, 1985).

The greatest risk factor of the built environment for women is simply that they generally spend much time at home. The exposure to infectious agents of children occurs predominantly at home. Whatever the insults from cleaning agents, formaldehyde, or household allergens, they especially affect women. To the extent that women have different types of factory or office jobs than men do, the insults of the workplace environment would also differ, but this has been little studied. The built environment includes the presence or absence of such facilities as health clinics and family planning centers, day care facilities, schools, nursing homes, and hospitals.

The most obvious habitat differences for women are social. Activity spaces, mobility patterns, and isolation versus social contact mediate exposures to such things as air pollution, cigarette smoke, motor vehicle accidents, contagious disease, fresh air, and social support, to name a few.

Behavior

The most complex and yet clearest interactions for women's health fall into the behavioral vertex of beliefs, social organization, and technology. Indeed, most of the approaches to health promotion informed by social theory focus on this vertex alone (Doyal, 1995; Koblinsky et al., 1993). A woman's multiple social roles—as child nurturer, caregiver, homemaker, wage earner, career professional, community organizer, lifelong learner—lie here. Should the woman deal with the infant's toilet needs (and get worms), tend the child's fever (and catch it), deal emotionally with the elder's deterioration and needs, and wash the cadaver (and get cholera)? Has the woman any control over whether to have five children or two or none? Must she cope with full-time job demands, whether law clerking or gutting chickens on a conveyor belt, as well as with domestic chores, parenting, and a connubial relationship? What are these relationships and duties, and how are they determined, and by whom anyway?

Cultures have strong belief systems about everything from what foods are taboo for pregnant women to how thin or plump women should be; from what sick role is allowed women to what home remedies mothers should use; from what exercise is appropriate for women to how much domestic violence must be tolerated. Should menstruating women be shut away in separate buildings? Should postpartum women be "roasted" daily in overheated rooms, as is done in East and Southeast Asia, to reduce the womb? Should female genitals be mutilated so that husbands can know their spouses will be pure and uninterested in sex? Should women burn themselves alive on their husbands' funeral pyres, or are widows allowed to

remarry, if only to be able to care for their children? Is it appropriate for women to expose themselves to the sun? Is it appropriate for them to smoke cigarettes? The behaviors that result from such cultural beliefs have fundamental implications for health. The beliefs can sometimes, perhaps unexpectedly, be strong cultural buffers against harm.

Social organization that affects women's health can be as personal as customs of polygamy or monogamy sanctioned by religious systems, or as bureaucratic as welfare aid to homeless women with dependent children. Broad social structures and institutions affect women's health: child care services; programs of higher education and of job training; and programs for health care, vaccination, life enrichment for senior centers, and even nursing home pet visitation. Legal structures—work laws, minimum wage, safety regulations, and antidiscrimination/affirmative action laws—also channel health effects. There is a lifetime of social science research to be done in political economy about such relations.

The complexities of interactions even within just this behavioral vertex are inscrutable when the context of major concerns, such as mental health, is addressed. Consider simply the dimension of substance abuse. Women have the same range of alcohol metabolism that men do, although their generally lower body weight gives them generally less capacity for absorption. Under the influence of alcohol, women are especially susceptible to rape and to physical abuse.

But women tend to drink for different reasons than men do; they do it to meet different emotional needs and to escape different problems. More importantly, they generally drink alone—probably because society has in the past thought women drinking alcohol at all to be unacceptable, let alone in public. Because they are not social drinkers, they are hidden drinkers. They are less prone to drunk driving manslaughter, and they are less subject to social intervention or likely to receive the concern of others. Programs such as Alcoholics Anonymous are specialized on the needs of men, focused on the reasons they drink, and occasionally even thick in some of the abusive attitudes that led to the women's drinking. Government programs for women are seldom available at all.

Technology can involve the full workplace panoply of computers and chemicals, or it can involve the labor-saving gadgets and chemical innovations of homes in developed countries. Medical technology, however, is of special importance. The removal of birth from homes to hospitals and the development of intensive care for premature infants have greatly reduced mortality and transformed the birthing experience. The development of infant formulas to substitute for mothers' milk has removed the mortal crisis of inadequate lactation; has freed infant care from the sole proprietorship of a mother; and has transformed the experience of bonding, as well as that of work. Such formulas, aggressively promoted for profit, have also resulted in infant death from diarrhea and malnutrition where the human resources of adequate money, potable water, fuel, and cleanliness did not comport with safe usage.

Heart disease continues to be a greater cause of both mortality and morbidity among men than women, but after menopause it becomes the major cause of death and the most common life-threatening morbidity for women, too. Numerous other differences between the health status of men and women continue. What compo-

nent of the differences in arthritis is due to different experience of occupation and what to hormones or diet? Is men's greater hearing impairment due to occupational hazard, or to cultural behavior regarding music, guns, or power motors? What could be the cultural ecology of more hay fever, sinusitis, and bronchitis among women? Or are these differences all due to women's crying to the medical system at the least sniffle while men endure in silent agony?

CAUSAL REASONING AND EPIDEMIOLOGICAL DESIGN

It is very difficult to get at the differences among ethnic groups and between races as etiology that matters. It is even harder to identify causal exposures to the environment or socioeconomic relationships as such. Age standardization of rates (see Chapter 5), and often sex or race standardization as well, are usually the first steps for comparison and pattern discernment among differences in risk and exposure. Etiological analysis, however, often involves analyzing developments over long periods or among people with diverse behavior and habitats. Determining the cause of an acute infectious disease can often be helped by interviews and by the small numbers and specific nature of the cases involved. Determining the cause of a disease with a long latency period can be even more difficult, especially given the mobility of population and changing exposure over time.

Four of the logical canons developed by John Stuart Mill (1856) underlie most causal reasoning in health studies.

1. *Difference*. When all conditions among the study populations are alike except for one, that one condition is implicated as either causal or preventive of the disease. This is the classic logic of laboratory experimental studies. Rats with the same inheritance are kept under identical conditions, except for exposure to the chemical that is being studied. The best application of such an experimental design to study of human populations can be found in the clinical trial, in which one group of people is given a new drug and another group is given a placebo, such as a valueless sugar pill. In this way psychological attitude, as well as the healing or exposure risks of time, are kept the same for both groups except for effects of the drug being studied.

2. *Agreement*. When all circumstances are different except for the variable being studied, that variable is implicated as causal.

3. *Concomitant variation*. When a factor varies systematically with the frequency of the disease, that factor is implicated as being causal. When more or less of the variable is associated with more or less of the disease, it is varying systematically.

4. *Residue*. When the effect of the known causal factor is removed in order to isolate and measure the variation remaining, successful explanation of the remainder supports the factor's causality. This is the method that geographers use when they map the residuals of a regression or the factor scores of a factor analysis, in order to see whether the pattern of unexplained variation remaining elicits any further hypotheses.

When repeated studies support these logical canons, causation is gradually established. Studies of cigarette smoking, for example, have found that when groups matched for age, sex, ethnicity, income, education, occupation, personality, and activity differ only in smoking cigarettes or not, there is a great difference in the incidence of cancer between them (i.e., the difference canon is supported). If people are studied who live in totally different cultures and environments—if they are Muslims and Buddhists and Christians; if they are poor and rich; if they are literate and illiterate; if they do and do not eat meat; if they live in cold, dry places and hot, wet places, or in urban places and rural places—the smokers always have a higher incidence of lung cancer than do the nonsmokers (i.e., the agreement canon is supported). When populations of smokers are subdivided according to how many cigarettes they smoke, at every increment there is associated greater risk of getting cancer (i.e., the canon of concomitant variation is supported). When the effect of cigarette smoking is statistically removed, it is possible to identify patterns of exposure to air pollution (i.e., the residue canon is supported). Thus, although statistics cannot prove anything, the accumulated and consistent evidence along different logical paths makes any doubt that cigarette smoking causes lung cancer unreasonable.

It is difficult to be sure, however, that all other factors are the same. Many early animal studies of the dietary effects of water hardness, for example, had to be redone when scientists learned how to measure trace elements. The water had not only differed in calcium carbonate, as intended and measured, but also in the trace amounts of molybdenum, cadmium, selenium, and other elements. The amounts of trace elements had not been held constant and plausibly were causally related to the outcome.

A *confounding variable* is one that varies in a systematic way with the hypothesized causal relationship being studied. Although *A* seems to cause *B*, in fact another variable, *C*, is affecting both *A* and *B*; the relationship between *A* and *B* is therefore spurious. Sometimes one is aware of a confusing interaction. For example, soft water has been associated with higher risk of stroke. In the United States, soft water occurs in the Southeast on the coastal plain, which of course is a little above sea level. When researchers study stroke in soft-water regions and hard-water regions, they also find strong associations with altitude and with the range of temperature changes. Which is truly causal? Or do these characteristics (variables) affect each other and merely happen to vary in the same way as stroke, perhaps because of still another variable that has yet to be identified? One of the most difficult tasks in social science research on health is identifying which variables need to be controlled so that they will not confound the relationship being studied. This, of course, is the role of theory, but at early levels of understanding of disease etiology, theory may be inadequate.

Confounding can be controlled both by analysis and by research design. The most confounding variable of all, as discussed earlier, is age. Whether one studies the prevalence of antibodies to a virus, life stresses, activity pattern, cholesterol deposition, or public health knowledge, one needs to know the age of those involved. We have seen how analytical technique can control for age by standardizing data.

The devilish thing about confounding factors, however, is that the researcher may not know they exist. One of the major purposes of epidemiological research strategy is to control for confounding factors even when they are not clearly specified.

The ideal epidemiological evidence is to find that the different disease frequencies in two populations are dependent on a difference in a certain factor, and that furthermore, within each of the populations, this factor is more common among those with the disease than those without it. Two broad strategies address these questions: *cohort studies* and *case-control studies*, both diagrammed in Figure 8.3. In a cohort study, a population is studied. It is divided into those with and without a particular exposure, and the frequency of the disease outcome is noted. When this study is started before the exposure and people are followed forward in time, it is called a *prospective study*. When the disease outcome has already happened and the history is reconstructed through interviews and records, it is a *retrospective study*. A case-control study starts with people who have the disease or who died of it, and compares their behavior and exposures with those of the rest of the population. Although all the nondiseased population may be used, usually it is sampled. The

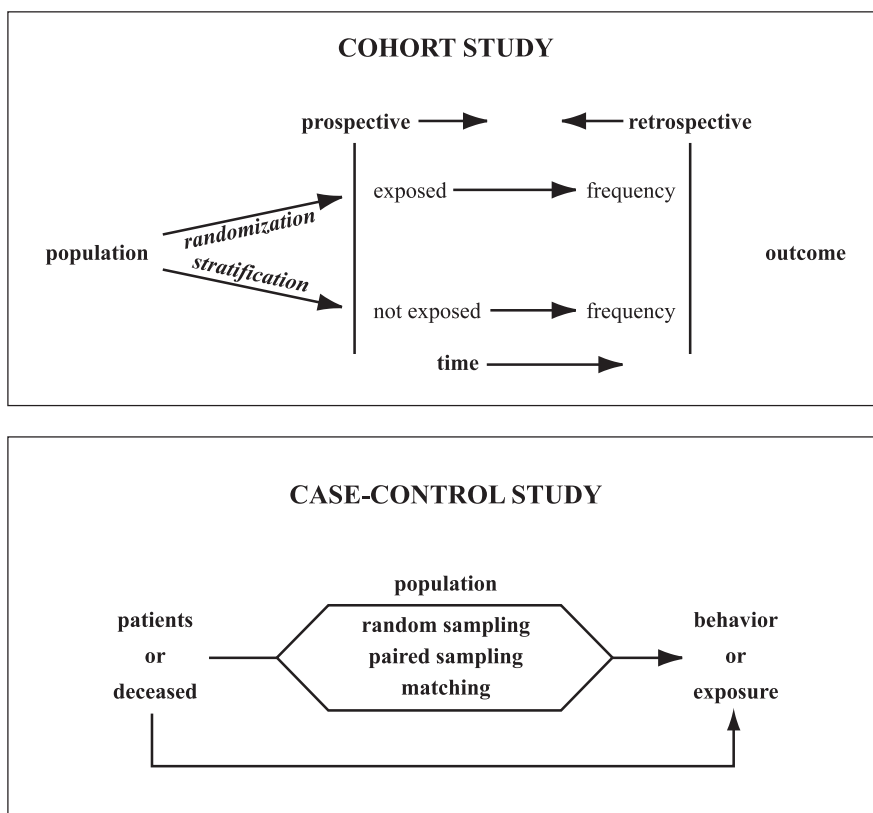


FIGURE 8.3. Epidemiological design.

whole population may be sampled randomly, or cases may be paired in a systematic way with people from the general population. One may interview the neighbor next door to the right, for example, or the next patient admitted to the hospital for a non-related reason. Alternatively, one may pair the sample deliberately by matching the patient on a series of criteria (age, sex, ethnicity, income, education, etc.).

Each of these research strategies has advantages and disadvantages. A prospective cohort study has the great value of allowing direct estimation of the risk associated with causal factors. As one follows the population through time, for example, one can measure how many times they ate a certain food or drank a certain source of water and how long they spent in a certain occupation. There are no spurious relationships added by the means of collecting data, such as a bias introduced when people who are suffering are motivated to remember events more carefully. The great disadvantages of such a study, however, are the costs in time, skilled personnel, and money. Following a population is laborious, time-consuming, and expensive, especially if the disease is not too common and large numbers of people must be involved in order to get enough data for a definitive analysis.

A case-control study is relatively quick and inexpensive, and these characteristics make it easily repeatable and able to include large numbers of people. This approach can introduce certain biases into data collection, however, and researchers are less likely to find things they did not set out to find. Questions are formed by researchers, and people recall their specific experiences in reaction to the questions.

In both these research strategies, randomization and stratification are used to convert confounding variables into control variables. When a population is randomly assigned to two groups or individuals are randomly chosen for comparison, the effects of the confounding variable are converted into residual variation equally distributed between groups, so there is no need for analytic control of it. For example, consider a test of a new vaccine. Some children are naturally resistant to the disease, and some may have had it already. Some children are in more contact with other children, some have better diets, and some have more psychological stress. All these things may affect whether or not a child gets the disease. A researcher wants to compare the proportion of children given the vaccine who contract the disease with the proportion of unvaccinated children (given a placebo) who contract the disease, in order to assess the vaccine's effectiveness. By assigning children randomly, the researcher converts the difference in contracting the disease that is due to relative natural resistance or to differences in exposure into unexplained variation common to both groups equally, and so it is removed as a confounding factor. Matching cases or stratifying the population with regard to certain characteristics has the same effect of making sure that those characteristics do not introduce confounding variance. Stratifying or matching by ethnicity or education, for example, eliminates variance due not only to those characteristics but to every variable associated with them, such as diet, home environment, and neighborhood influences. These epidemiological methods of analysis have been developed and standardized over decades.

Geographers have implicitly used such causal logic in several of their methodologies (Amedeo & Colledge, 1975). *Analogue area analysis*, for example, matches characteristics of regions so that they are as similar as possible except for the variable being studied (see later discussion). The method of *agreement* is also frequently used. Japanese and Icelanders are different in almost every respect, except that both populations have a high incidence of stomach cancer and both populations consume large quantities of smoked fish. Therefore, other populations that consume large quantities of smoked fish should have a high incidence of stomach cancer. The research question is posed.

DISEASE ECOLOGY: CANCER

Cancer is a family of diseases. There is no one cause or one cure for cancer; there are only multiple effects of multiple treatments and outcomes. Our knowledge and classification of cancer are analogous to our knowledge and classification of infectious disease in the 19th century before germ theory. Then cholera was not differentiated from yellow fever or typhoid from diphtheria. There were only fever, "flux," vomiting, rash, flatulence, cramps, and so forth. The best doctors of the day could write of malaria "becoming typhous in its course." It was only after germ theory gave us a microbe for each disease that the separate entities were classified, and causation and treatment were defined appropriately. Classification of cancer by the site of occurrence is similar to classification of "fever" in the 19th century. For example, if stomach cancer and cancer of the pancreas are both caused by a certain chemical, they are classified as different diseases; however, pancreatic cancers that are caused (for all we know) by minerals or viruses are lumped together by organ, in spite of their different disease agents. As we learn more about the different causes and about appropriate prevention and treatment, our classification scheme is bound to change.

The long latency of these diseases has bedeviled etiological research because of the many changes over time and space that occur. Geographers, whose study usually encompasses areas large enough to have significant population mobility, need to appreciate these time-space processes. Figure 8.4 illustrates a model of cancer that explains why it takes cancer so long to develop. When a carcinogen enters the body, it is usually detoxified by enzymes, broken down, and excreted. Occasionally these enzymes and related processes can activate the carcinogen and enable it to enter a cell. If it binds to anything in the cell except the DNA, it can only affect that individual cell. If it binds to the DNA, many repair mechanisms attack it. The bonds between the carcinogen and the DNA are broken, and the carcinogen is broken up, transported, and excreted. Only if the DNA is replicated before the carcinogen is removed can the carcinogen affect the new cell information, creating a cancer gene. The evidence clearly indicates that one altered gene is usually not enough to cause cancer. The process must be repeated, seemingly against all odds, by another carcinogen in a cell that already has the altered genetic material. The odds are changed if substances known as *promoters* (sometimes metal elements, fractions of hydrocar-

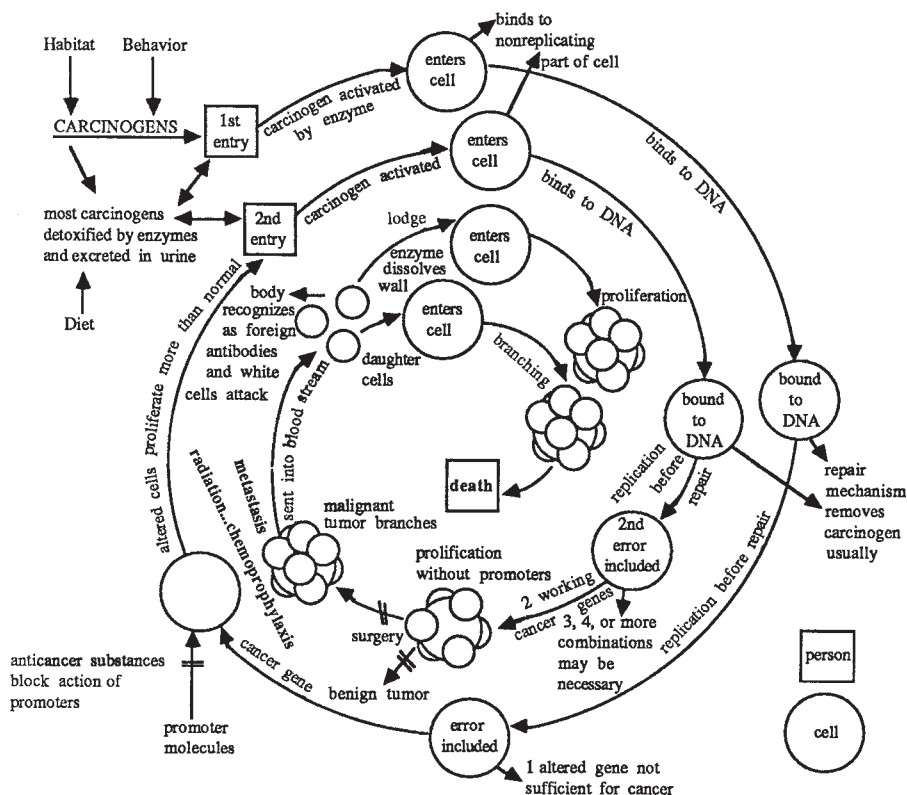


FIGURE 8.4. Cancer causation. Repeated alteration of genetic material is usually required to cause cancer, and the generally low probability that the necessary alteration will be repeated is affected by the presence of promoter substances and the dosage and duration of carcinogens, as well as the presence of protective nutrients for repair and buffering mechanisms.

bon chains, fatty acids, or substances such as saccharin) encourage the altered cells to proliferate faster than normal cells. This process creates an increasing number of targets for another carcinogen to enter. The rare chance that such a contact will occur, among all the normal cells of the body, helps to explain why decades may elapse before expression of the disease. It also explains why cancer is most common in tissues with normally high rates of proliferation, such as skin or the lining of the intestines and uterus, and rare in nonproliferating tissues such as nerves. Proliferation increases the number of cells containing the altered genetic information. Every time a carcinogen enters the body, the whole chain of chance events must repeat itself.

In 1999, scientists succeeded in turning normal human cells into cancerous ones by altering four genes. They turned on an *oncogene* to make the cell multiply quickly, shut off a tumor suppressor gene that limited replication, disabled a gene that caused abnormal cells to self-destruct, and turned on a gene to rebuild the telomere caps on chromosome ends that normally shorten and age with each division. It took these four changes in one cell to create a cancer.

As we continue to follow Figure 8.4, we see that a benign, noninvasive tumor can be removed by surgery. If the tumor sends out branches and envelops other tissue, it is cancer. The malignant tumor can send “daughter” cells into the bloodstream—an event known as *metastasis*. The daughter cells drift with the bloodstream, in danger of being recognized as foreign and attacked by the body’s antibodies and white cells, until they lodge. They produce an enzyme that dissolves the blood vessel wall, which allows them to enter normal tissue and begin proliferation and invasive branching in the new site. Eventually vital tissues and processes are involved, and death results.

Cancer’s need for repeated defiance of the odds in an orderly but improbable sequence of events explains the long latency period. The need for separate carcinogens, promoters, and overcoming of at least four defense mechanisms of the body explains the often baffling etiology and outcome. The adequacy of vitamins (and other antioxidants) in the diet is important for the body’s ability to detoxify a carcinogen, especially after it has bonded within a cell. The status of the immune system is critical because it must identify cells with altered genetic information as foreign and destroy them. All the interactions and balances of this complex process change over time.

The Genetic Base

Some kinds of cancer have long been suspected of being genetically linked. There is evidence that some cancers affect men more than women when there has been equal exposure. Others are common in some population groups (e.g., nasopharyngeal cancer among the southern Chinese) and rare in others. Some cancers are more common in a family than in the general population. It is for these *familial* cancers that genetic markers are most likely to be found. The specific genes involved in familial disease may not be involved at all in the *sporadics*, or cases that occur without an apparent hereditary role.

Cells must divide and replicate if tissue is to grow and maintain health. Each replication involves the exact reproduction of their DNA. Accidents happen, however. The four bases of the cells’ DNA may spell out G, C, A, T, for example. In this example, G matches with C, and A with T, to form a three-dimensional molecular structure that locks with a specific enzyme. If during replication somehow G matches with T, an error of mismatch has occurred. Alternatively, radiation may knock a proton off, changing G and causing an error of excision. Genetic excision or damage may eliminate the ability to receive and regulate a chemical, as when a damaged cell without ability to respond by negating a cell growth signal turns cancerous. In either case, proteins patrolling the DNA have to recognize the damage and call in repairs. Figure 8.4 implies several mechanisms of genetic involvement.

The importance of damage recognition and enzyme combinations in DNA repair and carcinogen detoxification is genetically given. Differences may make people more or less susceptible. A variant of the p450 gene enzyme that is poor at metabolizing a certain carcinogen, for example, may increase a person’s susceptibility to the toxic effects of cigarette smoke. Men may be more susceptible to many

mutagenic agents because they bind more carcinogen to their chromosomes. This has been most studied with regard to lung cancer and the carcinogens and promoters in cigarette smoke. Men have been found to have greater stimulation of DNA repair synthesis, resulting from the increasing amount of damage. Damage and repair activities increase with age for both sexes, but are higher at all ages for men.

The best single example of the processes above lies with *Science's* 1993 "molecule of the year," p53, which is resident on chromosome 17. When it was first found that p53 was involved in 60% of cancers, and even 80% of colon cancers, it was implicated as a carcinogen binder. Advancing research, however, demonstrated that p53's role is to halt the replication of any cell with DNA damage, thus stopping abnormal growth processes. It is a tumor killer. Damage to p53 itself, however, can sabotage this critical control and defense system. It is now the gene scientists turn off when they are studying the process of carcinogenesis discussed previously.

The most notorious gene in the class of tumor suppressor genes is BRCA1 (a BRCA2 has also been recently discovered). It makes a protein that binds to DNA and is directly involved in repairing DNA damage. It interacts with many other genes for purposes that are as yet uncertain, but include development of the embryo. More than 600 mutations of this gene have been identified. The mutated gene makes proteins too short to fix the damage to DNA of other genes, as shown in Figure 8.4. Damaged BRCA1 causes greatly increased risk of early-onset breast cancer (breast cancer is bimodal in incidence; most occur in women over 60, but damaged BRCA is associated with breast cancer in 30-year-olds). It also increases the risks of ovarian, fallopian tube, and prostate cancer.

One of the most active areas of cancer research concerns *oncogenes* and the breaking of chromosomes. Molecular biologists have identified more than 20 of these mysterious genes, which normally are inactive. They are found in such diverse organisms as flies, fish, mammals, and yeast. Thus, although their functions are unknown, they are presumed to be vital because they have been so carefully conserved throughout evolution. Oncogenes seem to be found consistently in tumorous cells, often in multiple copies. Sometimes they are involved in turning on growth genes, and possibly in chromosome breaking. It seems clear that chromosomes have a propensity to break in certain places. Genetic control of these weak places may be involved in susceptibility to cancer.

There have been many developments and breakthroughs in understanding since the human genome was mapped. It is now realized that genes can have multiple expressions, depending on the contexts of other genes, and that genes work in groups and combinations to create those contexts. In the DNA example used above, G may match with C; however, sometimes G is expressed via G, C, and A's acting as one and involving an entirely different enzyme system—one that previously was not identified with G. Body stem cells, converted to tumor cells, may be involved in the ability to migrate through other body tissues and thus metastasize. Possible new pathways for controlling cancer by shutting its growth down genetically have opened; whether these pathways become real roads to effective treatment or merely cul-de-sacs will probably be known by the time the next edition of this book is published!

The Behavioral Base

Cultural risk factors for cancer involve a variety of customs and practices, as well as economic goods and occupations. Given that solar radiation and genetic susceptibility by population group have not changed, the propensity of modern youth to sunbathe aggravates risk of skin cancer, compared to earlier practices of wearing long sleeves and broad hats and otherwise modestly covering the body. Tobacco smoking, or chewing, is the most obvious endangering behavior. Diet can both help and harm. A diet poor in fiber means that undigested food passes too slowly through the intestines; it is thought that this allows toxic bacterial products to become concentrated and increase risk for colon cancer. The hydrocarbons and nitrosamines of smoked fish (and presumably other smoked products) and charred meat seem to be risk factors for stomach cancer. Diet may also be preventive: Substances in green, yellow, or purple vegetables serve as antioxidants to protect cell walls. The vitamin E in fish skin or the antioxidants in green tea may do the same. The human diet contains a variety of natural carcinogens, promoters, antimutagens, and anticarcinogens. Science has barely begun to address the relative risk or protection of the enormous variety of dietary habits around the world.

The Habitat Base

The behavioral base in economic structure, occupation, industrial processes, and regulatory control has dominated risk assessment—but it is in the habitat, especially the built environment, that carcinogens lurk. Workers exposed to asbestos, to the manufacture or use of industrial chemicals, to agricultural sprays, or to the other multiform dusts and radiations connected with earning a livelihood in an industrialized society have long served as the guinea pigs for most of our threshold and effect data. The larger-scale behavior of locating certain enterprises (e.g., oil refineries, nuclear power plants, or toxic waste incinerators) in certain populous areas and not others has rarely been evaluated from such an experimental perspective. An increasing concern for issues of environmental racism, highlighted by geographers and other social scientists, has demanded such research by the U.S. Environmental Protection Agency (EPA).

Mostly because of the high incidence of breast cancer in the Northeast, including Long Island, New Jersey, and Cape Cod, major studies (both large-scale and local) have been carried out in recent years. As the map of breast cancer in Plate 12A shows, the industrial shores of the Great Lakes (Chicago, Detroit, Buffalo) and the Northeast coast have concentrations of the most significantly high rates in the country. Ethnically, these areas include many people of eastern and southern European descent. But they also include concentrations of many chemicals, both present and past. The Silent Spring Institute on Cape Cod has organized at a grassroots level and worked to develop GIS and other analytical systems and environmental monitoring and analysis to include studies of local settlement in research on local environments. One of these projects was an exhaustive analytical study of existing PubMed databases and publications on the connection of chemical pollutants identified as

important in toxicological (laboratory, animal) studies and breast cancer (Brody et al., 2007). Many environmental pollutants cause mammary gland tumors in animals and are hormonally active, especially mimicking estrogen. A small number of studies with confirmed evidence were found to support an association between breast cancer on the one hand and polycyclic aromatic hydrocarbons and PCBs (the latter, though not allowed since the 1970s, have accumulated in fish to become women's main carcinogen exposure) on the other. The evidence of a connection with dioxins and organic solvents is suggestive but methodologically awkward, including disasters of mass exposure of children (not yet reproducing) in Russia and Italy (following accidents in chemical plants making herbicides). Methodological problems abound, especially centered on inadequacy of exposure assessment, lack of access to highly exposed (usually occupational) populations, and the almost total lack of investigation of chemicals in humans that have been found toxic and carcinogenic in animals. Associations of breast cancer with organochlorine pesticides (aldrin, DDT, dieldrin, heptachlor, etc.) have also been suggestive but not conclusive. Researchers urge greater use of toxicology data for control and prevention.

Patterns and Change

There is amazing geographic variation in the occurrence of cancer. For decades it has been recognized that differences in incidence rates, not only between men and women but also between urban and rural populations and among countries, hold etiological clues. The highest female cancer rate by country is far lower than the top 10 male rates. Japan has the highest stomach cancer rate but is not ranked in the top 10 for lung, breast, or rectal cancer. Given Australia's biometeorology, its white population's genetics, and such behavioral patterns as ranching and surfing, one can understand why that country is the highest-ranked in skin cancer—but why should countries as diverse as Uruguay, France, and Singapore be ranked so highly for esophageal cancer? Furthermore, within each country, urban and rural rates can differ as greatly as international rates.

The extraordinary data richness at the micro scale in China was made clear in an atlas that portrayed the most dramatic spatial patterns of concentration (Editorial Committee, 1979). Not only was nasopharyngeal cancer concentrated among Cantonese on the Guangdong plateau, but there were extremely high foci of esophageal cancer west of the Taihang Shan in Shanxi and of liver cancer around the mouth of the Yangzi and southeastern coast. At the national level, these cancers, which are almost unknown in the United States, have been the major causes of cancer death until this last decade. With economic liberalization in the late 1970s to the 1980s, however, cigarette smoking reached prodigious levels, and the familiar consequence of lung cancer has built to a wave that is breaking now—greatly compounded by exposure to extraordinary levels of air pollution in the major cities.

The patterns of cancer incidence have changed over time. Lilienfeld and Lilienfeld (1980) have outlined several possible reasons. The changes may be *artifactual*; that is, they may result from errors due to changes in the recognition, classification, or reporting of the disease, or from errors in enumerating the population.

Real changes may result because the age structure of the population has changed, because people survive diseases that were once incurable and thereby live long enough for cancer to become manifest, or because genetic or environmental factors have changed. (As epidemiologists, the Lilienfelds include behavioral changes under environmental factors.)

Greenberg (1983) found a spatial convergence of cancer mortality in the United States between 1950 and 1975. There was a strong parallel trend for most types of cancer: Rates diverged by sex and race subgroups of the population, but converged geographically. He contended that this spatial convergence of cancer mortality (i.e., development of a homogeneous pattern) in the United States and other industrialized countries has been caused by change in the geography of risk factors associated with the diffusion of urban culture. These risk factors include air and water pollution, cigarette smoking, alcohol consumption, diet style, occupation, SES, stress, and medical practices. Ezzati et al. (2008) identified lung cancer as being responsible for the “reversal of fortune” as female life expectancy at the county level was decreased for more than a decade after 1983.

Mapping and Measuring Risk

The appropriate scale for comparative mapping of cancer has been especially problematic. Rare cancers, such as brain tumors, have a low range of rates compared to common cancers, such as colon cancer. The distribution of the African American population varies greatly across the United States, leaving regions of counties with almost no population denominator when cancer deaths are broken down by age and gender. One solution, used by the first National Cancer Institute county-level maps (Mason, McKay, Hoover, Blot, & Fraumeni, 1975) as well as a more recent cancer atlas by the U.S. EPA (1987), is to map white rates at the county level and black rates at the scale of state economic areas (which aggregate contiguous counties into clusters, creating an intermediate scale between the county and state levels). The *Atlas of United States Mortality* (from which our Plates 12 and 13 are taken) addressed the small-number problem by using health service areas as the scale of aggregation for all its disease maps (National Center for Health Statistics, 1997). As explained in other chapters, the rates are age-adjusted and mapped as mortality ratios compared to U.S. national rates, and the distribution of death rates by health service areas is presented in a graph. Hatching indicates where the data are sparse, mostly because of small populations at risk, to facilitate an honest interpretation.

Very often risk is measured in ways that take small numbers and emphasize their importance, so that people lose sight of the fact that very small numbers were involved in a study or event. One result is that the public gets scared, and then numbed and contemptuous, as chemicals and especially foods pop up as risks of cancer and then disappear (or earlier findings are reversed by larger studies). Yet, given that large-scale experiments can't be done with people and cancer, such small indicator studies are often the best possible way to connect with the animal toxicology results. Those reading any of this literature should understand how two measures work in particular: the *relative risk* and the *odds ratio*.

- *Risk* is the number of times (frequency) an event occurs relative to the total number in the study population—for example, 1 in 5.
- *Relative risk* is the risk of an event in those with a risk factor divided by the risk of an event in those without the risk factor.
- *Odds ratio* is the number of times an event occurs relative to the number of times it doesn't occur; that is, the ratio of event to no event in *exposed* divided by unexposed individuals.

Consider a population of 1,000, of whom there are 400 smokers and 50 cases of cancer. Forty cases of cancer occur to smokers, and 10 cases of cancer occur to nonsmokers.

The risk = 50/1,000, or 0.05.

$$\text{The relative risk} = \frac{40/400}{10/600} = \frac{0.1}{0.017} = 5.9$$

$$\text{The odds ratio} = \frac{40/360}{10/590} = \frac{0.111}{0.0169} = 6.6$$

Some Studies of Cancer

Geographers have studied spatial patterns to identify risk factors. As discussed above, the research questions and methodology may involve determining regions and investigating common or different factors by using *analogue area analysis* (trying to hold possible associations constant, except for the study variable) and applying the causative reasoning described above. For example, several regions can be identified in which stomach cancer has a high incidence. Iceland, Japan, and northern Minnesota/Wisconsin are three such regions. Copper seems to be unusually concentrated in the soils of one of these; is it also high in the others? People in Iceland eat smoked and salted fish and are of Scandinavian descent. Do people in Japan eat salted and smoked fish, and do people of Scandinavian descent in Minnesota and Wisconsin continue ethnic dietary practices? Similarly, some people in eastern Africa have a high incidence of esophageal cancer in a culture area in which home-distilled alcohol is made. People outside that culture area have lower cancer rates; perhaps the alcohol or method of its production is involved. The region around the Caspian Sea in Iran, however, has some of the world's highest rates of esophageal cancer, and people there do not consume home-distilled alcohol. Research suggests that the Iranian risk area is characterized by nutrient-deficient saline soil, halophytic vegetation, and poor crops. Alcohol could easily absorb elements from the pottery and cans in which it is distilled. Could trace element deficiency or toxicity be common to both regions? Geographers in Japan, Europe, and North America have analyzed the pattern of cancer occurrence in relation to environmental, socioeconomic, and demographic factors by means of multivariate statistical procedures. There are quite a few of such studies, but results have been inconclusive about the associations found.

Studies that are based on general populations, that do not select disease categories for research carefully and deliberately, or that target fortuitous places run into many data and methodological problems. They so often suffer from the ecological fallacy (see Chapter 1) that “ecological” (i.e., multivariate in statistical use) studies as a whole have earned a reputation for scientific weakness. Such multifactor, population-based studies are, however, a potentially powerful means for choosing places for case-control and other detailed studies. An important research question regarding disease distribution concerns the degree and scale of spatial clustering. These issues are explored in the discussion of neighborhood analysis in Chapter 9. Map patterns alone cannot eliminate chance in clustering, but observed patterns can be compared to randomized ones by using stochastic approaches, chi-square tests, and several spatial statistical methods.

In the discussion of time-space geography, exposure, and population mobility in Chapter 5, the analytic benefit of large, well-defined streams of migration that can be identified and followed is illustrated by the studies of breast cancer and stomach cancer among generations of Japanese in Hawaii. Environmental factors, genetics, lifestyles, and resistant customs such as favorite foods were separated by movement and generation. As noted there, however, such research is rarely possible. The study of micro-scale mobility and exposure to different environments can also be used to identify possible etiological factors and to test new hypotheses. Armstrong (1976, 1978) studied nasopharyngeal cancer, a disease of unknown etiology, among the Chinese in Malaysia. Using a case-control method (discussed above), he compared the exposure patterns of people with the disease and those of controls (people without the disease) to what he called “self-specific environments.” He measured how much time was spent daily in agricultural areas, in squatter housing or middle-class housing, in shop workplaces, in factories, in shopping on the street, in various means of transportation, in public places, and so forth; he also measured how much time was spent in smoky places, in crowded places, around chemicals, and under conditions of bad air pollution. He examined such indexes of traditional or modern/assimilated cultural patterns as having an altar in the house and eating meals in a formal manner, as well as details of what people ate and how they prepared it. He found that among genetically susceptible Chinese, important stimuli were associated with industrial or trade occupations, smoky workplaces, poorer housing, traditional lifestyle, a diet with less variety of foods, and childhood consumption of salted fish.

A different approach was taken by Glick (1979a, 1979b, 1980, 1982). *Spatial autocorrelation* is the association between a variable's values taken from two adjacent places. Instead of treating this lack of independence as a nuisance (as is usually done), Glick used the scale and intensity of that association to test etiological hypotheses. He used east-west and north-south transects across dozens of U.S. counties, and the spatial relationships of cities and industries among states, to study the spatial organization of cancer rates. Because different cancer-inducing processes act most strongly at different scales, knowledge of the relationship of scale and cancer variation can help identify the process. Glick started with theoretical models of carcinogenesis and inferred the spatial relationships. There has been regrettably little follow-up to this research.

There are two major classes of skin cancer. Basal or squamous cell carcinomas are fairly common and relatively innocuous because they can be successfully treated without hospitalization. Malignant melanoma, a rare form of skin cancer, is often fatal. Glick (1979a) had data from four special surveys of skin cancer. He used regression techniques to estimate the effects of age and of ultraviolet radiation levels and found that nonmelanoma skin cancer has a strong age effect, as one would expect if the carcinogenesis process requires repeated exposures and alterations of the cell (as illustrated in Figure 8.4). Glick reasoned that because nonmelanoma skin cancer involves more stages of exposure, it should have a steeper spatial gradient of mortality. Indeed, he found that the gradient in mortality rate for a transect (line) through Minnesota–Iowa–Missouri–Arkansas–Louisiana was very steep and constant for nonmelanoma skin cancer, going from about 1.0 to 2.6 per 100,000; mortality from melanoma ranged from 1.1 in Minnesota to 1.8 in Louisiana, with a less constant gradient.

Figure 8.5 presents a spatial correlogram for two types of cancer that illustrates the spatial pattern of point and area exposure to risk (see Vignette 12.3). Across the X-axis are the spatial lags in association. For example, at lag 1 the average autocorrelation of each county in the state with adjacent counties is expressed, whereas at lag 5 the autocorrelation is with five counties away. The autocorrelation is expressed as a normal (Z) score on the Y-axis. The farther away a score is from zero, either positively or negatively, the stronger the association among counties. The line that shows point exposure has a steep distance decay in autocorrelation from high positive (similar to the adjacent county) to high negative (different from the adjacent county). It represents a spatial pattern of isolated foci of incidence, such as might result from a contagious disease occurring in large cities but not yet diffused across the county line, or from behavior limited to ethnic enclaves. Stomach cancer in Pennsylvania shows such a pattern, whereas bladder cancer is characterized by a nearly horizontal correlogram for lags 1–3 (Glick, 1980). Such a pattern, as illustrated, portrays a

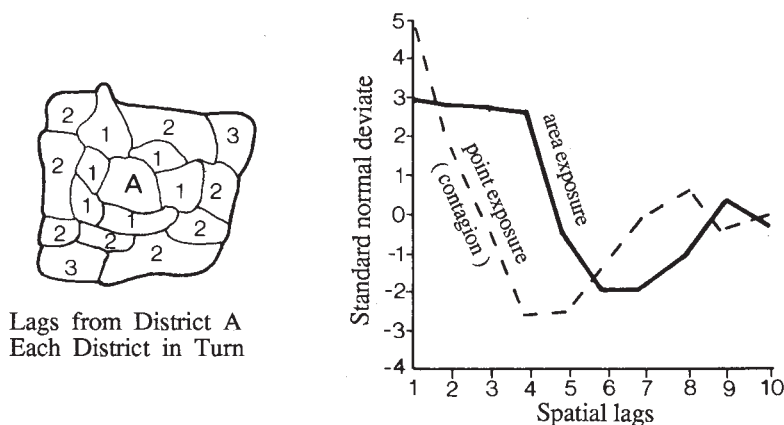


FIGURE 8.5. Spatial correlograms. Point and areal (regional) sources of exposure are reflected in different levels of correlation between spatial units when the units are lagged incrementally.

large region of similar rates and suggests an environmental carcinogen found over large areas (e.g., as water catchment basins), rather than a diffusion process.

Prostate cancer is one of the most prevalent forms of cancer in the United States, with especially high rates among African Americans. It is of unknown etiology, although strongly related to age. Hanchette (1988; Hanchette & Schwartz, 1992) used regression and trend surface analysis to draw out the association of ultraviolet light, prostate cancer, and a possible role of vitamin D and its associated enzymes. The strong north–south gradient of this disease is also correlated with colon cancer. Schwartz and Hanchette (2006) repeated this trend surface analysis for a 45-year period, cutting the map at 40° latitude because vitamin D synthesis does not occur in winter months at higher latitudes (Figure 8.6). They thus confirmed and extended their earlier results and found strong support for the hypothesis that vitamin D insufficiency increases risk for prostate cancer. Logically (again, see Mill's canons above), this is supported by the greatly higher rates for African American men (see Chapter 2 for a discussion of gene–environment interaction).

DISEASE ECOLOGY: CARDIOVASCULAR DISEASE

As noted earlier in this chapter, the group of diseases known as CVD constitutes the major cause of death in developed countries at the end of the demographic transition. It includes ischemic heart disease (heart tissue damage because of obstruction of blood flow to heart muscle itself), cerebrovascular disease (blockage or other damage to blood vessels supplying the brain), atherosclerosis (an arteriosclerosis, or hardening of the arteries, characterized by fatty deposition and fibrosis of the inner lining), hypertension (high arterial blood pressure), and rheumatic heart disease (reduced function because of heart inflammation and scarring due to previous rheumatic fever).

The resistance of the blood vessels to the flow of blood through them is blood pressure. This is affected by the volume and composition of the blood, by the dilation and constriction of blood vessels, and by narrowing of the arteries caused by the buildup of fatty deposits. Systolic blood pressure is the pressure on the arteries when the heart contracts and pumps out a surge of blood; diastolic blood pressure is the pressure on the arteries between contractions, the minimum pressure that the arteries feel continuously. High blood pressure for which the cause is unknown is referred to as *essential hypertension*. When the heart's blood supply is drastically affected, a myocardial infarction (heart attack) occurs. If the blood supply is only partially restricted, a chronic coronary thrombosis usually results. When the blood supply to the brain is blocked, a stroke results (apoplexy).

CVD is the leading cause of death in developed countries, although at the rate it is declining and cancer is increasing, this may change within a few years. It is the major cause of permanent disability, and it accounts for more hospitalizations than any other illness. CVD has been declining since the 1960s, however, especially among white males (whose rates were extremely high). The National Council on Cardiovascular Disease Prevention (Cooper et al., 2000) concluded that coronary

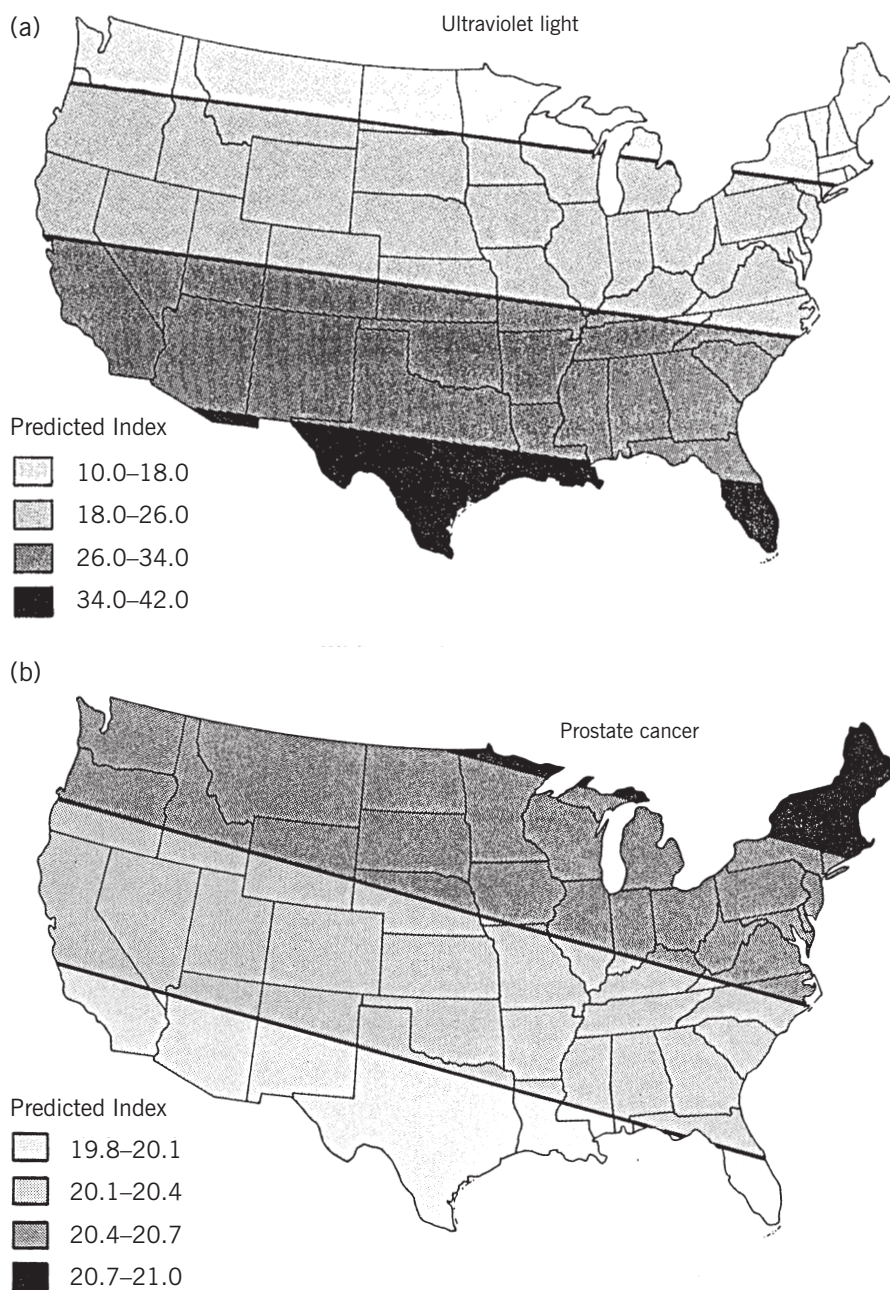


FIGURE 8.6. Maps of trend surface analysis of ultraviolet light as protective of prostate cancer. (a) Linear trend surface by county observations of ultraviolet light measured by epidemiological index. (b) Linear trend surface by county observations of prostate cancer mortality for white males, aggregated from 1970 to 1979. From Hanchette (1988). Used by permission.

heart disease mortality is still declining as a whole, although perhaps at a slower rate than in previous decades, but that stroke mortality rates have declined little in 20 years. The council also noted the striking differences in CVD rates by race/ethnicity, SES, and geography discussed earlier in this chapter. There are wide disparities among population groups and geographic areas in efforts at stopping smoking, increasing physical activity, controlling hypertension, and reducing obesity. The etiology of CVD rests on the familiar triangle of population–behavior–habitat interactions.

The Population Base

A wide range of genetic factors is involved in CVD. Several kinds of congenital defects may be involved, such as improperly closed heart valves. Genetic controls on metabolism are also important. People process, store, and remove fats in different ways. There may be a genetic predisposition to hypertension, acting through the renal enzymes that control fluid volume and excretion. Individuals react differently to chemicals (in drugs and foods) that promote vasodilation. The facility with which blood clots form has a strong genetic component as well. One of the most marked genetic differences is expressed in the lower incidence of ischemic heart disease among premenopausal women than among men—a protection that diminishes with the loss of estrogen. As explained earlier, however, the Women's Health Initiative found that taking estrogen after menopause actually increased the chance of stroke, rather than decreasing the chance of heart attacks. Sex hormones also influence how the body metabolizes and stores cholesterol and other fats. The level and composition of hormones differ not only between males and females, but also among individuals and from family to family.

Personality is also partly based on genetic combinations, although it has a behavioral expression and is influenced by environmental factors. It forms one of the bridges from genetic to behavioral considerations. There have been many intervention studies designed upon the distinction between Type A and Type B personalities. Type A personality involves aggressiveness, excessive drive, compulsive behavior, competitiveness, impatience, and insecurity. It is considered a coronary risk factor, and intervention programs try to get such people to slow down and relax. Numerous later studies have suggested that the dimension of hostility and anger is the most destructive element in the Type A personality syndrome.

Certainly the most targeted of genetic capacities is polymorphism for making and regulating the body's own cholesterol. Intervention with this inheritable trait through statin drugs that lower cholesterol levels has proven one of the most impressive forms of medical prophylaxis in the last decade.

The Behavioral Base

The behavioral base for CVD includes tobacco and alcohol use; eating salt, saturated fat, and triglycerides; poor exercise habits; various occupational factors; psychosocial insults; and stress and its management. Given the different kinds of chole-

terol (see below), the fact that the body makes its own, and the role of other factors (exercise, alcohol, and dietary combinations), today even the cholesterol in eggs no longer seems so important. Cigarette smoking is clearly a risk factor, however. The cadmium in cigarettes raises blood pressure, and chemicals they introduce to the blood circulation affect vasodilation. A little alcohol consumption is actually protective (apparently it favorably affects cholesterol balance), but heavy alcohol consumption interacts with nutrition, weight, and exercise to constitute a risk factor. Many dietary factors in CVD are still unknown. The high potassium levels found in some vegetables may be protective, for example, and calcium may be more important than previously recognized.

The role of cholesterol and other blood lipids (fats) is complicated by the fact that the liver makes some cholesterol itself, regardless of diet, and by the need to differentiate subcategories of lipids. High-density lipid (HDL) cholesterol acts to transport fats and clean out the circulatory system; it is the “good cholesterol.” Low-density lipid (LDL) cholesterol builds up in deposits on the artery walls; it is “bad.” The metabolic changes associated with exercise, with a little consumption of alcohol, and with some drugs increase the ratio of the protective HDL to the endangering LDL and so lower the risk of atherosclerosis. Exercise, then, has the dual impact of increasing the capacity of the circulatory system to pump blood so that it can better cope with sudden demands on it, and inducing a more favorable lipid balance through metabolic changes. Statin drugs act to greatly reduce the body’s own production of cholesterol. As this is being written, there is great excitement over a recent finding that even in people without elevated cholesterol, but with blood tests indicating the inflammation of arteries (which is itself a powerful risk factor), strong statin drugs lowered stroke, heart disease, arterial diseases, and overall mortality by 40–50% for everyone. Now long-term impacts need to be assessed, and public health concerns meant that needed changes in behavior are not discouraged.

Stress is a rubric that covers many evils. Stress and the body are connected through the immune and endocrine systems (see Chapter 2). Psychosocial insults come from situations as diverse as occupation and working conditions, neighborhood blight, domestic problems with spouse or children, and prolonged fear and anxiety. Many studies have found sudden myocardial infarction to be the illness most closely related to life event stress. Loss of job, upward mobility, marriage, divorce, death of a spouse, change of residence, and trouble with children are characteristically common among patients with heart attacks (see also Figure 2.1 and Audy’s levels of health). Similarly, the important occupational characteristics seem not to be specific job categories so much as lack of control, repetitiveness, inability to use one’s skills, and similar frustrations in one’s work. This seems related to the work of Evans et al. (1994) and Marmot and Theorell (1988) indicating that not simple deprivation, but relative status and hierarchy, are involved in CVD etiology. Marmot and Mustard (1994) consider the improvement in Japanese health, and especially the change at population levels from stroke as the leading cause of death uniquely in the world to CVD today, although both diseases are at lower levels than in the United States. Japanese propensity for strokes had previously been explained by ascription of a “shame culture” of tradition that turned stress inward. Presently,

work frustration seems to have become as important as among other developed countries.

All the stresses related to the poverty syndrome are especially relevant to CVD. Numerous studies within industrialized countries have shown positive health effects from a prosperous and growing economy, whereas times of recession and unemployment add tremendously to life stresses and to mortality. There is, however, a rather complicated relationship with socioeconomic change in the developing world. Cross-cultural studies have found that increased hypertension is often related to greater participation in the money economy. This participation generally entails major changes in diet, occupation, mobility, and other conditions. Economic change and loss of traditions add to stress. Economic growth and upward mobility are associated with increased risk of CVD. Not just the social organization and technology parts of the behavior vertex are involved here, but, importantly, beliefs and values. As Corin (1994) has put it, "The impact of changing conditions varies according to the significance attached to traditional ways of life" (p. 105). This can be compared to health risks and stresses in migrants.

Our knowledge of just how SES relates to biology, behavior, and social factors on the one hand and CVD morbidity and mortality trends on the other, especially with regard to minorities, remains limited. There is agreement about the need to improve our understanding of the concept of SES and the ways in which it reflects the conditions of everyday life for people of various economic strata (Lenfant, 1996). We know that there are strong, consistent inverse relationships: Lower SES means higher CVD risk factors (varying by age, sex, and ethnicity), such as smoking and obesity. The relations are weaker and less consistent with serum cholesterol, deliberate exercise, and lifestyle, all of which have poor national data. Psychosocial variables for SES groups of various ethnic backgrounds especially need research: What are the effects of job change, unemployment, economic instability, and social support, and how are these perceived?

Complications of race cut across all the SES and behavioral components. For example, it seems that not just race or SES, but also place of birth, matters (Fang, Madhavan, & Alderman, 1996). When CVD for blacks was compared with that for whites in New York City, but was also linked by census records to birth in the Northeast, the South, or the Caribbean, black mortality overall was indeed greater than white. But blacks born in the Northeast had almost the same death rates (per 100,000) as whites (white males, 285, vs. black males, 299; white females, 155, vs. black females, 165). Caribbean-born blacks had lower rates of death from heart disease than even Northeast-born whites in all age groups. Black males 25–44 years old who were born in the South had a 30% higher heart disease mortality rate than did black males born in the Northeast. Strokes for males ages 45–64 were a little less dramatically matched (white males, 25.3; black males, Northeast, 74.0; black males, South, 102.9), but the pattern is clear. Geographers cannot help remarking that, of course, other factors must be considered: spatial variation and regionalization of socioeconomic conditions, discrimination, and opportunity; differences in health care; and differences in cultural factors such as religion. The United States is big and diverse in culture, people, and environment; it is not at all a uniform

surface with an undifferentiated national population well studied by national sampling. More research needs to be done at varying scales. Such studies of race, SES, and place have led Gillum (1996) to editorialize that public health response and understanding have neglected the dynamic aspects of the problem, and especially the heterogeneity of rates among blacks. He has hypothesized associated factors at six stages: precolonial Africa; modern urban Africa; modern black populations in the West Indies; rural populations in the southern United States (similar to West Indies); poor in the inner-city United States; and affluent in the new suburbs.

The effects of stress on health have illuminated the importance of stress management. Intervention programs are being implemented to teach people how to relax, to use exercise as a stress release, to engage in meditation, and to soothe themselves by caressing pets and taking time to "smell the roses." The role of social support and close relationships in coping with stress has also come to the fore.

The Habitat Base

There are three important components of the habitat vertex of CVD. Available health services are especially important. The detection, treatment, and control of hypertension are among medical science's greatest interventions. The development of emergency medical systems and of cardiopulmonary resuscitation techniques has been important for heart attacks. The other two critical aspects of the habitat are weather and geochemistry.

Several physical mechanisms of acclimatization and adaptation to changes in pressure and temperature have been discussed in Chapter 6. CVD has a peak occurrence during periods of cold weather. Incidence seems to be responsive to the passage of fronts and other meteorological events. Because of physiological conditions of vasodilation and vasoconstriction, peripheral capillary networks for heat dissipation or conservation, and changes in the composition and thickness of blood, it is also likely that different climatic circumstances have something to do with the wide range of disease incidence.

An active area of research concerns the etiological significance of trace elements. These elements occur in such small amounts that only recently have scientists been able to measure them accurately, in parts per billion. The study of trace elements is further complicated by the constant interaction among them. Zinc, for example, is important to tissue healing. After a heart attack, most of the body's zinc is concentrated in the heart, and it seems reasonable that deficiency in zinc may play a role in fatal outcome. Zinc, however, is blocked by copper, so that the adequacy of zinc in the diet cannot be studied separately. Cadmium, a heavy metal and a systemic poison like lead, is known to raise blood pressure in laboratory animals. It may well be important in essential hypertension. Selenium, however, seems to block the action of cadmium and is a constituent of blood enzymes important for blood clotting and cholesterol metabolism. Selenium could not be directly measured until the 1970s by expensive techniques of nuclear absorption, so through the 1980s its associations were especially studied. Similarly, lead, nickel, chromium, manganese, lithium, potassium, sulfur, cobalt, and iron are variously involved in nerve transmission to

the heart, muscle healing, cholesterol metabolism, blood clotting, and hypertension. They all combine, enhance, block, and precipitate each other in various combinations. In addition, abundant elements such as calcium and magnesium are important electrolytes and have been implicated in arrhythmias and sudden death.

A geographer has to be concerned about the pathways that these elements can take to the individuals at risk. The *water factor* has been implicated in many studies. More than a score of international studies have found water hardness to be inversely related to CVD mortality. Hardness is usually based on the amount of calcium carbonate, but sometimes other minerals are included. The implication is that soft water dissolves more and so has a different composition of trace elements than does hard water. Only some elements, however, can be consumed in water in sufficient quantity to make any difference. Thousands of gallons would have to be consumed daily to obtain significant nutritional input of iron or potassium. Food is a much more concentrated source of most elements. Further complications can result from using analysis of surface water for populations that drink from artesian sources, or from water supply altered by treatment plants (e.g., through softening) and passed through municipal pipes. Figure 8.7 illustrates some of the geochemical pathways that trace elements take from the environment into human bodies. The original rock through which water moves is very important. Limestones are high in calcium and magnesium, which are accompanied usually by strontium, phosphorus, and zinc, and sometimes by cadmium and lead. Shales are high in aluminum silicates and, locally, in molybdenum, vanadium, and uranium. Sandstone is rich in

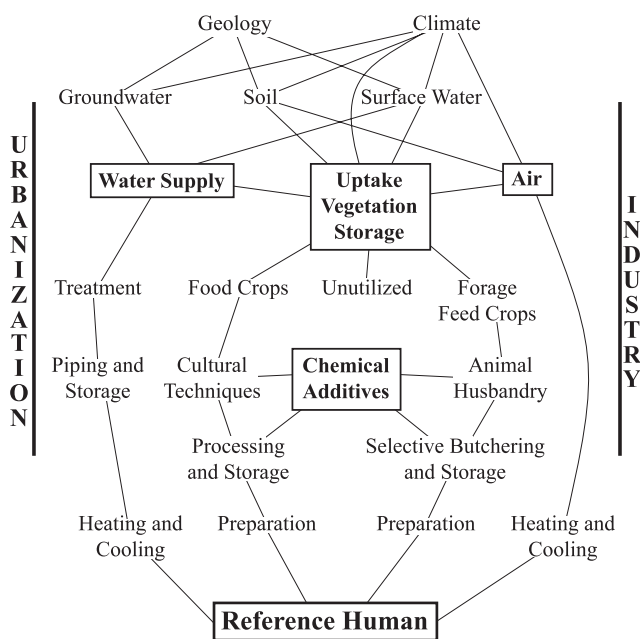


FIGURE 8.7. Geochemical pathways. Human consumption of trace elements is strongly influenced by the complex steps of processing and distributing of soil or water.

silica, but it is usually leached of trace elements. Serpentine rock is high in iron and magnesium, and often in chromium, nickel, and manganese. How these elements become bound and available depends on temperature, other minerals available, and groundwater and surface water supplies.

The uptake of these elements by vegetation may be controlled by the plants' metabolism, regardless of soil concentration, and the elements may be stored in plant parts that humans do not consume. Heating often causes the elements to combine and precipitate, while the passage of time during food processing and water piping permits them to be absorbed and to bind to other elements. Herbicides, insecticides, fungicides, and preservatives, which usually derive much of their potency from metals such as mercury or selenium, add to the concentrations of elements associated with the original soil or river water. After all these sources have contributed, the elements ingested by humans are still measured in parts per billion. It is no wonder that the effects of water hardness and various trace elements remain a controversial subject (Meade, 1980).

Patterns and Analysis

The high prevalence of CVD in industrialized countries is a basic component of the demographic and nutritional transitions (Chapter 5). This generality, however, obscures important variation among places. Most geographers have been concerned with mapping CVD and modeling its spatial variation, identifying patterns of clustering and changing associations at different scales; however, little has been done now in such analysis for 20 years. Similarly, in the 1970s there were about a dozen geographic studies of water hardness, which seemed correlated with CVD. The element most closely related to CVD in British water studies is calcium, which has been associated with sudden death due to arrhythmia. There is a strong north-south gradation of CVD in the United Kingdom (Scotland being high), and so there is interest in the inverse association with water hardness. The lower southeastern incidence coincides with the hard waters of Tertiary and Cretaceous formations; the higher northwestern incidence coincides with the soft waters of Precambrian rock formations. It has been noted that the Industrial Revolution began in the regions with soft water partly because of the water quality, and so early urbanization and industrialization might be confounding factors. The pattern persists, however, even without the industrial cities. Jones and Moon (1987) point out that studies of the water hardness factor offer good examples of the modifiable areal-unit problem (see Chapter 12, Figures 12.1 and 12.2). In Canada, the inverse relationship exists for all Canadian municipalities together, but it becomes positive when analyzed separately for British Columbia and the maritime and prairie provinces.

CVD in the United States also has a strong regional pattern (Plate 13). Schroeder (1966) was the first to study CVD and the water factor there, using data on finished water supplies in 1,300 municipalities across the country. At the national scale, significant inverse relationships were found between both male and female deaths and water hardness. Cerebrovascular disease is strongly clustered in the southeastern counties along the coastal plain of Georgia and South Carolina, a region sometimes

referred to as the *enigma area* (Meade, 1979). Again, the water factor has been implicated, as well as the deficiency of certain trace elements (e.g., selenium) in the soil and water. Temperature, population settlement and inheritance, agricultural patterns, poverty, late urbanization, and dietary factors may all be related as well.

The spatial correlation of a geochemical factor and CVD would be most convincing (again, see Mill's logical canons) if it held consistently at several scales and for all subgroups of the population. It would be further strengthened if geochemicals functioning as etiological agents could be plausibly connected with a biophysical mechanism. Researchers have noted that the more rigorous the study and the more focused the study area, the weaker and more contradictory the geological correspondence with CVD becomes (raising geographic questions about appropriate scale). As Comstock (1979) asserted, associations may be spurious, coincidental, or casual; however, the association of a geochemical factor with CVD is not spurious. Many independent studies in many countries have found statistically significant correlations. Comstock further lamented that most studies had relied on correlation coefficients (scatter around the line) rather than regression coefficients (slope of the line), which indicates strength of relationship (see Vignette 12.2). Using those studies that did publish regression coefficients, he estimated the relative risk at 1.25. This would certainly be low for, say, an occupational exposure; but water affects the entire population, and such risk would not be unimportant to public health.

Most of the regionalization of CVD within counties remains unexplained. The high death rates from CVD on the southeastern coastal plain of the United States, mentioned previously, cluster in a way that cannot be explained by the conventional risk factors. Counties of the southeastern coastal plain cluster for both stroke and heart disease, for whites and blacks, and for males and females (Plate 13). The lower Mississippi Valley has high rates, as the poor region of the Appalachian Mountains does for whites. (These two regions were combined to make one of Murray et al.'s [2006] "eight Americas" [see earlier discussion and Figure 8.1]. These high rates were not simple effects of race or income.) Analyzing the associated distribution and covariation of CVD with a series of overlays of geochemistry, agriculture, economics, ethnicity, industry, age of building construction, and various sources of water might suggest some new hypotheses.

Thinking that a micro-scale approach through study of population mobility might allow the identification of new environmental risk factors, Meade (1983) selected census tracts within the city of Savannah, Georgia, on the basis of socioeconomic variables that predicted stroke and blood pressure. Tracts with stroke rates that were well predicted served as controls for those where stroke was significantly higher than predicted. Interviews and 24-hour recall (see Vignette 4.1) of 10 households in each of 20 clusters within these census tracts (with rates that were high and predicted to be so, low and predicted to be so, and unpredictably high or low) ascertained the population's exposure to habitats within the city and its immediate environs. Figure 8.8 shows (a) the ellipses of activity areas constructed for the 20 clusters and overlaid on a map of land use; (b) an example of an activity area "cut out" and showing different land uses within; and bar graphs representing (c) the

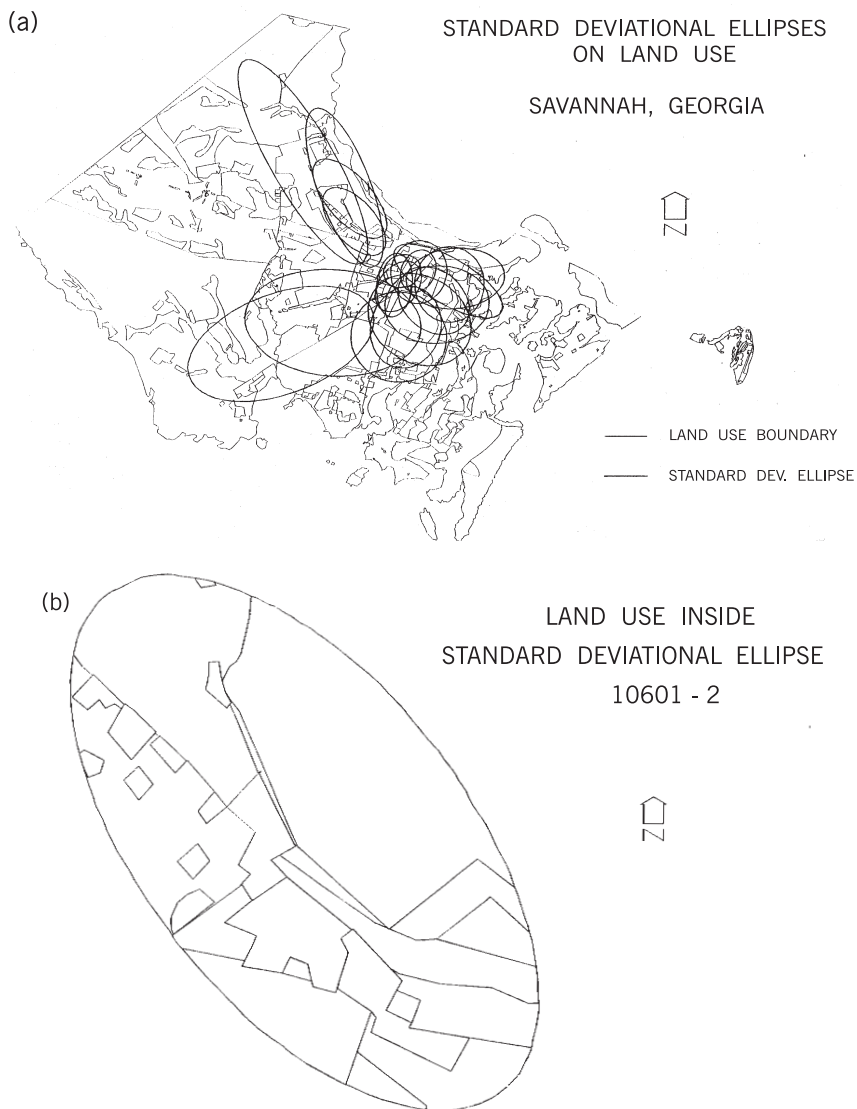
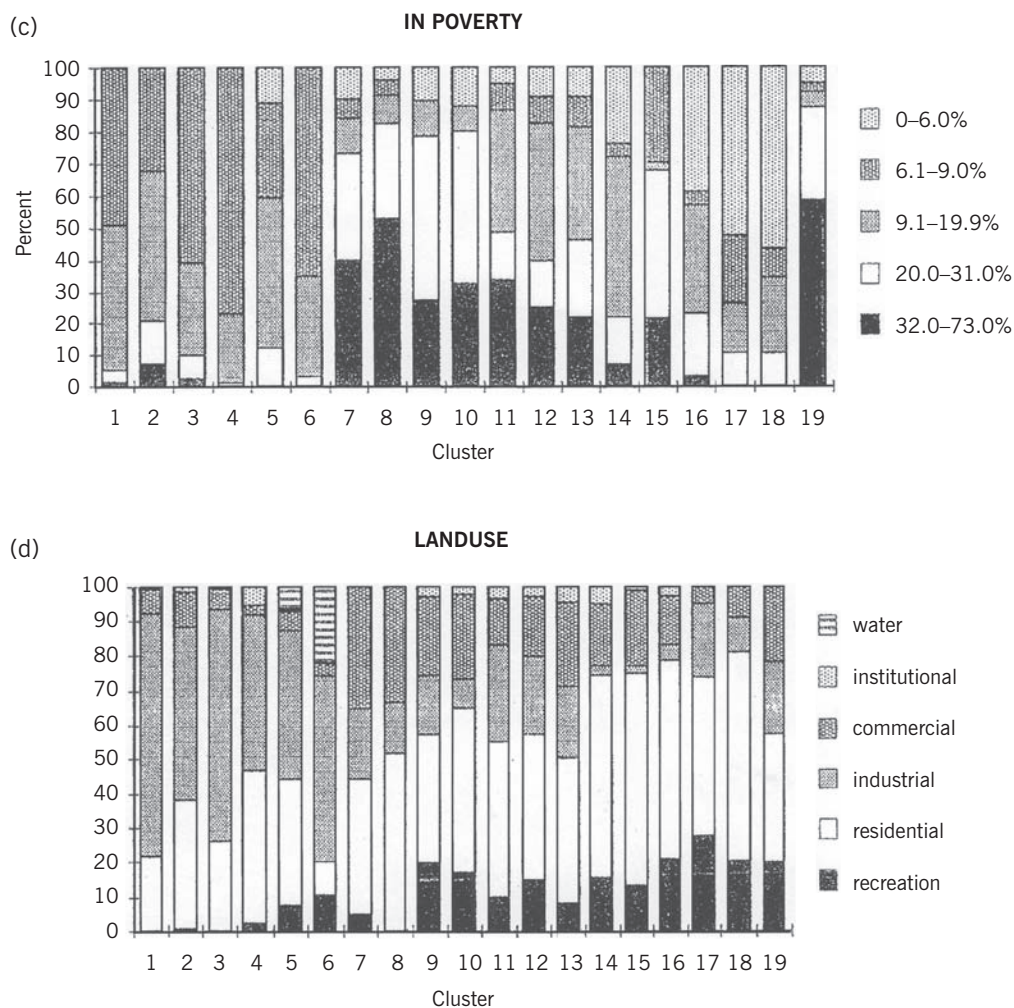


FIGURE 8.8. Ellipses of activity areas from interviewed clusters of households in Savannah, Georgia, (a) overlaid on land use and (b) cut out for cluster 2 in census tract 10601. The two bar graphs show the proportions of the activity areas composed of (c) various census attributes (e.g., percentage of households below poverty line) and (d) various land uses.

**FIGURE 8.8.** (*cont.*)

proportions of different census tracts with their demographic and statistical attributes, and (d) the proportions of the activity areas made up of the various land use environments (the habitats within which the people in the interviewed households moved and were exposed).

Japan has a geographic variation in CVD and in cancer that has attracted geographic and epidemiological study. The incidence of stroke is especially high in the northeastern parts of the main island of Honshu, the region of Tohoku, but lower around the Inland Sea. Takahashi (1981) found that salt consumption was higher in Tohoku, and that the water hardness ratio was low in Tohoku and high in the Inland Sea area. Kagami (1991) especially has sought to elucidate regional variation of disease in Japan as a geographic phenomenon at various scales. He examined

regional variation down to the prefecture level in cause of death throughout the demographic transition, from 1899 to 1985. Among the factors he considered (using statistical analysis, trend surface analysis, and mapping) were earned income, occupation, and labor; animal protein and salt in diet; agricultural cropping and land use; temperature; patient–doctor ratios; migration, and socioeconomic factors of industrial structure, population structure, and lifestyle. Kagami observed that through time, national diversity between urbanized and rural regions was reflected by the dominant diseases in each prefecture: tuberculosis (urbanized) and gastroenteritis and cerebrovascular disease (still rural). Regional variation of CVD was examined via trend surface analysis and was described as low in the southwestern plateau and high in the northeastern basins (differentiated by lower winter temperatures, mountainous terrain, and later industrialization). Within the northeastern region, Kagami analyzed Yamagata Prefecture by its regional divisions, using multiregression analysis and factor analysis. The highest mortality was characterized at division level by rice cultivation as the major industry, labor migration to the cities in winter, insufficient medical service and underdeveloped infrastructure, deficiency of animal protein and fat, and high intake of salt from winter consumption of pickled vegetables. Kagami concluded that previous studies had considered only main factors of diseases and had given little attention to regional characteristics as a complex of various scales. Since these studies with their rural associations were completed, heart disease has come to be the leading cause of death in urban Japan as well as other developed countries.

The main behavioral risk factors for CVD have been well established: diets high in saturated fats and heavy with salt; inadequate intake of fruit and vegetables; obesity; physical inactivity; tobacco use and heavy alcohol consumption. This knowledge led to efforts to change the problematic behaviors (aerobic exercise, broiling instead of deep-fat frying, lower sodium, smoking cessation, etc.), and along with medicine to treat the symptoms, the knowledge has done much to lower mortality rates and disability days, even extending the life expectancy of men. Yet altogether such risk factors account for only part of CVD, and they cannot explain the spatial clustering or patterns such as the coastal plain phenomenon. Some factors of the physical environment—water composition, sunshine—may be involved. Most health geographers today are most interested in patterns of the social environment, though: neighborhood (see Chapter 9) analyses of access to fresh produce, concentrations of harmful billboards or fast-food restaurants, and the like. There remains much to explain.

UNKNOWN ETIOLOGY AND OTHER QUESTIONS

A geographic perspective on a disease can result in discovering patterns of association, so developing etiological hypotheses for diseases of unknown cause is potentially a fruitful, but perilous, endeavor. Multiple sclerosis had a latitudinal gradient from high northern rates to low southern ones in Europe and North America. Only

Japan was anomalous, having very low rates even at high latitudes. The low incidence in Japan and among people of Asian descent in the United States suggested a lack of genetic susceptibility. Today the WHO death rates by country for 2002 show no such clear gradient (WHO, 2004b). France, Austria, Ireland, and Slovenia have the highest MS death rates (8 or 9 per 100,000); both Norway (2) and Italy (5, although further south) have lower rates. New Zealand has lower rates than Australia, although it is at a higher latitude; Costa Rica, Brazil, and Argentina all have the same rate (3). China's rate of MS is as low as Japan's, but at the national level so is the U.S., although there has been great north-south variation within it. Studies of population migration, which results in differential exposure to risks in separate places, provide some of the best analytic frameworks for addressing the etiology of MS (Williams, Jones, & McKeran, 1991).

The most obvious latitudinal environmental factor is solar radiation, but it may be spurious, except for diseases associated with vitamin D. Less obvious variations from north to south in both Europe and North America include agricultural crops, development of industry, level of urbanization, and economic standards. Migration streams to Israel from high-MS-rate areas of Europe and low-rate areas of Africa have been studied for MS incidence later in life. Childhood exposure to some unknown factor is important: Adults who migrate have the risk of their place of origin, whereas small children take on the risk of their destination. The new generation of native Israelis, surprisingly, has high MS rates despite low latitude, implying that some environmental conditions associated with industrialization and developed economies are important (Lowis, 1986). In the United States, however, California, though more industrial, has lower MS rates than Washington State. Migrants from the low-rate South and high-rate North to California and Washington have been studied. Children assume the high or low rates of their destination, as in the Israeli study. Adult migrants from the South to Washington State have higher MS rates than in their birthplaces but lower than for native Washingtonians. This implies that some protective factor may continue to shield the migrants in the presence of greater environmental hazard. In the absence of a U.S. registration system for rare diseases, however, it is difficult and expensive to get a large enough stream of migrants to be able to analyze such diseases.

Rheumatic diseases, which are responsible for widespread suffering and loss of work time, have received little attention from geographers. Rheumatic diseases include such degenerative conditions as osteoarthritis, rheumatoid arthritis, spinal disk degeneration, gout, ankylosing spondylitis, and diseases of the connective tissue. Etiological hypotheses in the general literature include the causal impact of wear and tear, calcium deficiency, metabolic deficiency in regulating blood uric acid levels, infectious agents, "bad" water, diet, occupation, and emotional stress. Genetic heritage has often been implicated, but the best-designed studies have found the least evidence for familial clustering. The rheumatic diseases are notoriously difficult to confirm or detect in any repeatable manner, especially at subclinical levels. Even the blood serum test for rheumatoid factor (anti-gamma globulin) is known to flip capriciously across the arbitrary border that separates positive and negative results.

Parkinson's Disease

Another geographically neglected disease is Parkinson's disease (paralysis agitans), which afflicts millions of people in the United States and around the world (WHO, 2004b). The age-standardized U.S. death rate from Parkinson's is 3.2, which is higher than Australia's (2.5) or China's (2.4), but lower than Jordan's (5.3) or Syria's (5.9). Rates for several countries in West Africa range from 2.7 to 3.0, but in Japan it is only 1.1. So the disease afflicts countries both early and late in the demographic transition, as well as both industrialized and less industrialized countries. Parkinson's disease affects famous and powerful persons and poor workers without distinction—more than 10 million people in the United States. In North Carolina in 2007, it killed a total of 537 people, accounting for just under 1% of all deaths. These deaths were 54.7% white male, 38.4% white female, 3.7% minority male, and 3.2% minority female. Only 2 of these deaths occurred to people under 50, and another 11 to people under 65. Most deaths befell people ages 75–84 (49%) and ages 85 and over (34.1%). That's what we know about its distribution, which is to say we know next to nothing. Mortality is a very poor measure of the prevalence of Parkinson's disease because most people (including those with Parkinson's) die from something else before age 75 or 85. It is a disease of unknown etiology that afflicts millions.

The disease first described by James Parkinson in the early 19th century is best known for its trademark symptom, an uncontrollable tremor—first of the hands, usually, but also of the legs, and almost any muscle eventually. The “motor” never stops on its own. Other symptoms vary among patients. There are balance problems, with an urge to throw oneself backward because one feels as if the body is falling forward. Muscles get stiff and weak. The gait gets jerky and unreliable. People fall. About half develop the “on–off syndrome,” in which their medicine works well and they can walk and go out and perform normally, and then suddenly the medicine cuts out and they are unable to move a finger. Most, but not all, people eventually “freeze” and need to be touched before they can move again. Speech slurs. People drool. Problems with swallowing develop. Through all this, a person can remain lucid, with intelligence undiminished—but eventually dementia ensues.

Parkinson's disease is due to degeneration and death of cells in one region of the brain (the substantia nigra). The dysfunction of these cells results in a loss of the neurotransmitter dopamine, which they produce, resulting further in blocked nerve impulses in the basal ganglia of the brain. The receptors on cells for picking up dopamine also seem to atrophy or decrease. Semiautomatic movements (such as movement of the tongue while speaking or swing of the arms while walking) become difficult, while involuntary movements (tremors) become common; the brain loses its ability to command the muscles. Balance is disturbed, stiffness occurs, paralysis develops, and the disease progresses slowly but inexorably as brain cells continue to die. The reason those particular brain cells die is unknown. This is why Parkinson's disease is one of the “poster cases” for the need for stem cell research. Embryonic stem cells are accepted by the brain, which has little immune function, and they can grow into new dopamine-producing brain cells. It is unknown how long this effect will last or whether these cells will in turn die (or what point of dis-

ease progression, what dosage, etc., is ideal). The medicine Sinemet (a combination of carbidopa and L-dopa; the latter is a precursor to dopamine that can get through the barrier into the brain) has been around for several decades. It is so effective that it is often the diagnosis: If a person responds to Sinemet, he or she has Parkinson's and not another kind of tremor. After about 5 years, however, Sinemet gradually loses its effectiveness, and other medicines are added. They have side effects, such as uncontrolled thrashing about (dyskinesia) or hallucinations and dementia. Most of the recent drug developments have been concerned with ways of extending the effectiveness of Sinemet and avoiding the complications induced by other medications.

There is no state or federal government register for Parkinson's disease. A few registers are variously maintained by targeted research hospitals or private organizations such as the National Parkinson's Disease Foundation. A person who dies of Parkinson's has probably visited a doctor for it many times. Many have been hospitalized several times for a "drug holiday," when all medication is taken away until breathing is threatened and then gradually returned and rebalanced. People with Parkinson's disease are not invisible or unknown; *they just are not counted in place*. Morbidity from a chronic disease is not reportable.

There is an uncommon, early-onset form of Parkinson's that strikes people in their 30s. A gene has been identified that is associated with this form. Most people, however, get the disease in their 60s, and no gene has been identified (although many researchers are looking for multiple ones that change their expression, as discussed earlier). Causal gene mutations account for fewer than 10% of Parkinson's cases (Vance, 2008). Identical twins get the disease a little more commonly than unrelated people, but usually a decade or more apart (suggesting a common susceptibility, but different life experience and exposure). Ordinary siblings don't get the disease much more than unrelated people, but if they do get the disease, they seem to get it when reaching almost exactly the same age. Overall "mega-studies" have concluded that there is some increased genetic *susceptibility*—but no one knows what the risk is that people are susceptible to!

Research on all aspects of the disease was hampered until a decade ago by the lack of an animal model to experiment upon. Other animals don't naturally develop the disease. Then some teenagers experimenting with making boutique drugs in their basement in California were brought into the hospital with symptoms resembling full-blown, advanced Parkinson's. Determined analysis isolated a contaminant produced in their manufacture of drugs in the basement. It was a chemical related to herbicides—a chemical that proved able to induce the symptoms in animals. Since then, many rats and monkeys have suffered the "disease" so that new drugs can be developed. The finding also sent researchers looking again for connections to agriculture and rural living, which were previously thought to be associated with risk. (Today, since most urban people didn't grow up in rural areas, the nature of this risk has become unclear—suburban herbicides, contamination of food, fumigation, insecticide spraying?) Factors thought to be associated with increased risk of Parkinson's disease include aging, white race, male gender, trauma (as from boxing), shy personality, and environmental exposures (to, besides

rural residence and herbicides/pesticides, manganese, iron, well water, wood pulp mills, the steel alloy industry, and the swine flu epidemic of 1918). There is a lot of confounding here from cohort experience. There was Parkinson's long before there were chemical pesticides; most of the 80-year-olds who have the disease now grew up in a rural America, but most of the 60-year-olds developing the disease today did not. Influenza viruses can cause encephalitis. For decades it used to be thought that the disease might be an aftereffect of surviving the terrible influenza epidemic of 1918, but as that generation has passed on, the disease has only increased. There are also rare and separate kinds of Parkinson's, like that induced by head trauma that kills the cells (as in the case of the great boxer Muhammad Ali), or that seen in welders exposed to manganese—but no patterns have emerged. There has been little geographic analysis of the spatial distribution and environmental circumstances of this serious disease. It is difficult to prevent a disease when we have no idea what causes it. Why haven't we figured that out?

What are the obstacles to advancing knowledge of causation and prevention? Parkinson's disease is a rare disease, so the small-number problem is intrinsic to its study, as the population by race, sex, and age is so unevenly distributed in national space. The disease takes decades to develop. During this time, people often move several times, sometimes between states and regions, so that they have been exposed to different environments for different lengths of time. Prevalence of more than a decade of morbidity is poorly reflected in mortality data. Because there is no national register, the places of diagnosis, places of birth, and perhaps interviews on migrations and occupations are not available for geographic study. These "ecological" studies across multiple scales confounded by population mobility have never been methodologically acceptable in epidemiology. Qualitative studies seeking commonalities in lives and exposures are rare and unconnected. Perhaps GIS can help us make some necessary connections. The question burns: *What* is it that people are susceptible to? Can we get it out of the environment and prevent some of the disease?

THE PRECAUTIONARY PRINCIPLE AND SOME POLITICAL ECOLOGY OF RESEARCH

As one reads through hundreds of articles and reports on cancer, heart disease, and diseases of unknown etiology, it is remarkable how little is known, studied, or said about prevention. The few exceptions include the recent promotion of diets rich in fruit and vegetables (perhaps organically grown), and finally, after more than half a century of opposition and inaction, serious promotion of smoking cessation. Even now, though, although the tobacco-growing state of North Carolina allows its university in Chapel Hill to ban smoking within a hundred yards of all its buildings (definitely including its hospitals), the legislature steadfastly refuses to raise taxes on cigarettes despite critical budget problems. And despite the transformative research usefulness of GIS in managing and analyzing spatial data sets (discussed in Chapters 9 and 12), comparatively little funding is available for analyzing

environmental exposures (geographic environment, not epidemiological behavior), developing surveillance of harmful chemicals, or monitoring population risk or public health interventions. Why has so little been done? What drives funding? Why do so much research funding and research promise today lie with genetics and drugs for treatment?

The *precautionary principle* is that when there is a clear indication of harm (to human life and health), that indication should be a trigger for action even if the scientific evidence of proof is not complete, and especially if delay may cause irreparable damage. The idea has been argued for decades, but it received a new life when the principle was endorsed in 1987 by European ministers of the environment who were concerned about deterioration of the North Sea. More recently, it has been important in the arguments over causation of and response to global warming. In the area of health, Sandra Steingraber (1998) advances the principle powerfully to address the proliferation of known carcinogens in our environment. She describes the experience of cancer, putting a human face on the victims and even on the cell lines of research. She describes the nature of the data (the small-number problem, the difficulties of cluster identification, the lack of a register, and the mobility of the population). But most of all, she describes the chemical soup in which we live, exposed to chemicals in food and drinking water, by showers, and in clothing, house cleaners, paints, agricultural spraying, and industrial food wrapping. She searches for the research on connections of the environmental context—agricultural chemicals in the water, industrial gases and particles belching into the air, uncountable products made of plastics, and countless other industrial, residential, and personal chemical uses (see Table 7.1) to which the population is exposed. By her best tally in 1997, 40 possible carcinogens appear in drinking water around the country, 60 are released by industry into ambient air, and 66 are routinely sprayed on food crops as pesticides. She cites in full the “Wingspread Statement” from a 1997 conference (of scientists, lawyers, farmers, physicians, planners, environmental activists, philosophers, and others) in which she participated. An extensive excerpt from that statement follows (Steingraber, 1998, p. 284):

The release and use of toxic substances, the exploitation of resources, and physical alterations of the environment have had substantial unintended consequences affecting human health and the environment. Some of these concerns are high rates of learning deficiencies, asthma, cancer, birth defects and species extinctions, along with the global climate change, stratospheric ozone depletion and global worldwide contamination with toxic substances and nuclear material.

We believe existing environmental regulations and other decisions, particularly those based on risk assessment, have failed to protect adequately human health and the environment—the larger system of which we humans are a part.

[We believe that there is] evidence of damage to humans and that worldwide new environmental principles are needed; recognizing [that] human activities may involve hazards, people must proceed more carefully than [they have done in] recent history: corporations, governmental entities, organizations, communities, scientists, [and] other individuals must adopt a precautionary approach to all human endeavors.

Therefore it is necessary to implement the precautionary principle. When an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically. In this context, the proponent of an activity rather than the public, should bear the burden of proof.

. . . The process of applying the precautionary principle must be open, informed and democratic and must include potentially affected parties. It must also involve an examination of the full range of alternatives, including no action.

Funding of Studies: Genetics and Environmental Risk

Many millions have died in the last few decades from preventable cancer. In recent years several shocking analyses and exposés have been published about the suppression of meaningful research on prevention. Two of the most powerful and respectable are described here. Devra Davis, who is now Director of the Center for Environmental Oncology at the University of Pittsburgh Cancer Institute and a professor of epidemiology, has looked back over a career of 40 years in epidemiology, research, and government regulation and has reviewed rare unpublished research reports going back to the 1930s. In her book *The Secret History of the War on Cancer* (2007), she reports on how much was known about the carcinogenic importance of benzene in petroleum products, benzidine (which causes bladder cancer) in dyes, chromium, asbestos, and of course tobacco. (The U.S. government even supported research on “safe cigarettes,” known [by scientist consensus] to be impossible.) With deep data, and interviews (many from 50 years ago), and detailed scientific explanations, Davis explains how the 1971 launch of the government’s War on Cancer “focused on defeating an enemy by detecting, treating, and curing the disease—not preventing it.” She goes on to describe how this campaign ignored and/or deliberately suppressed “proof that how and where we live and work affects the chances we may get cancers was basically ignored” (p. 15); how alliances (either naïve or pecuniary) of producers of products, makers of medicines, sellers of medical treatments, and academic researchers ignored how much was known about social and environmental causes of cancer from the sophisticated research done before World War II in their peer-reviewed funding decisions; and how industrial leadership acted “decade by decade” to downplay research on prevention, instead preventing environmental causes from gaining circulation in public information or benefiting the general public.

David Michaels (2008) has recently published a book titled *Doubt Is Their Product*. He is an epidemiologist and director of the Project on Scientific Knowledge and Public Policy at the George Washington University; during the Clinton administration he served as Assistant Secretary of Energy for Environment, Safety, and Health, with responsibility for protecting the health and safety of workers, neighboring communities, and the environment surrounding nuclear weapons facilities. Winner of the 2006 American Association for the Advancement of Science’s Scientific Freedom and Responsibility Award, he is no muckraker. Yet he further exposes the failure of

legislation, policy development, and agency regulation to prevent cancer and other serious diseases because of the warping and compromising effects of money on judgment and supporting research. He especially takes on the peer review process and the increasing involvement of industry in financing even academic research. Review of one's research by professional peers is critical to all academic research, whether social, physical, or medical. It is how science maintains standards, builds on previous knowledge, and gives vital feedback. Peer review affects research at two stages: the evaluation of research proposals resulting in determining which research to fund, and the publication of research results. Noting how corporations funding university projects in the 1990s blocked publication of research detrimental to their companies until leading journals refused to publish unless the contract left the (academic) research free of commercial interest, Michaels notes how current regulatory agencies (the EPA, the Occupational Safety and Health Administration, etc.) still have no mechanisms to identify conflicts of interest resulting from sponsor control. He makes 10 recommendations, 3 of which are that there should be full disclosure of any and all sponsors' involvement in scientific studies; that manufacturers should test chemicals for safety before exposing workers and the public to them; and that secrecy should be ended as it affects the health of the public, with chief executives being held responsible for submitting "accurate and complete" reports on what is known about the toxicity of their products.

Davis and Michaels agree that the responsibility for the safety of new products released to the environment (and thus the responsibility for testing these) should be put on the producers of the products. They also agree that those setting priorities for research funding have concentrated on treating the effects of pollution (new medicines, medical specialties and equipment and tests) and perhaps refining or reducing the pollution (better smokestack scrubbers), but have largely ignored prevention of risk, which would threaten or preclude the production and release of the dangerous chemicals themselves.

Genetics has become the dominant theme of research, funding, and popular explanations in the mass media as well. This is increasingly disturbing to many kinds of health scientists. What British authors Hall (2004) and Hedgecoe (2001, 2002) call the "geneticisation" of heart disease, schizophrenia, and other conditions seems an "unstoppable process." That is, genetic explanations are being expanded to include the role of behavior and environment in the expressions of genes, including those that result in complex conditions. However, such explanations marginalize individual, social, and even environmental influences, which are reduced to mere "triggers" of the underlying genetic process or predisposition. (Or is this merely the old outrage of social scientists that anything about human behavior could be genetically caused?)

This chapter has expounded upon the benefits derived from decoding the human genome. Much more is now understood about the working of genes, not only as individuals but in groups and in multiple expressions triggered by environmental experience. Promising new medicines are being developed for regulating Parkinson's disease, starving cancer cells, and treating asthma. Some lives are being

saved, no doubt about it. But what was it that threatened those lives in the first place? The academic researchers gain reputations; the pharmaceutical companies and the manufacturers of medical research tools profit; the hospitals fill their beds and coffers; and the manufacturers of the dangerous chemicals themselves, of whatever kind, continue with their growth, profits, and dividends. So many are benefited—except the millions of better-treated human beings, who perhaps did not need to be afflicted at all. Some social scientists fear a new eugenics movement, and others fear the loss of health insurance by millions, as new “susceptibility genes” are disclosed. The underfunded, underresearched, unanswered basic question of disease ecology still burns: What is it that people with these genes are susceptible to? And, by the way, why did their society expose them to it?

CONCLUSION

This chapter on the political ecology of noncommunicable diseases has reviewed some of the geographic research on diseases of primary importance in the economically developed countries. As the demographic, nutritional, and mobility transitions continue, and as the world’s population continues to urbanize and age, the etiology, prevention, and treatment of degenerative diseases are increasingly dominating world health concerns. Much of the review and analysis presented here, however, has been concerned with the noncommunicable diseases of the American people, and with their behavior in their environment. On the basis of what you have learned, how would you answer these questions?

- Why are mortalities of married men and women lower for each age group than those of single people, whether never married, widowed, or divorced?
- Why is “education” (what’s that?) inversely associated with so many noncommunicable diseases, such as CVD?
- Why does Hawaii have the lowest, and Mississippi the highest, age-adjusted mortality rates?
- Why do women have more morbidity but less mortality than men? (This question is known as the female paradox.)

This chapter has also examined some of the methodological and epistemological issues for causative reasoning, and some designs for modeling and analyzing multifactorial relationships. The spatial patterns, causal relationships, and time and scale parameters of these diseases are complex matters. If hypotheses are to have value, the geographic researcher needs to understand the disease processes as founded in the population’s biology and its interaction with habitat and behavior. The complexities of spatial pattern and disease etiology demand competence in statistics and demography as well as mapping/GIS. On this foundation rests the ability to project population health needs, as distinct from economic demand, and to plan for the delivery of appropriate health services.

REFERENCES

- Adler, N., Marmot, M., McEwan, B., & Stewart, J. (Eds.). (1999). Socioeconomic status and health in industrial nations: Social, psychological, and biological pathways. *Annals of the New York Academy of Sciences*, 896, 1–503.
- Amedeo, D., & Colledge, R. G. (1975). *An introduction to scientific reasoning in geography*. New York: Wiley.
- Ames, B. N. (1983). Dietary carcinogens and anticarcinogens. *Science*, 221, 1256–1264.
- Armstrong, R. W. (1976). The geography of specific environments of patients and non-patients in cancer studies, with a Malaysian example. *Economic Geography*, 52, 161–170.
- Armstrong, R. W. (1978). Self-specific environments associated with nasopharyngeal carcinoma in Selangor, Malaysia. *Social Science and Medicine*, 12D, 149–156.
- Brody, J. G., Moysich, K. B., Humblet, O., Attfield, K. R., Beehler, G. P., & Rudel, R. A. (2007). Environmental pollutants and breast cancer: Epidemiological studies. *Cancer*, 109(Suppl. 12), 2667–2712.
- Calabrese, E. J. (1985). *Sex differences in response to toxic substances*. New York: Wiley.
- Canfield, M. A., Honein, M. A., Yuskiv, N., Xing, J., Mai, C. T., Collins, J. S., et al. (2006). National estimates and race/ethnic-specific variation of selected birth defects in the United States, 1999–2001. *Birth Defects Research A: Clinical and Molecular Teratology*, 76(11), 747–756.
- Case, A., & Paxson, C. (2005). Sex differences in morbidity and mortality. *Demography*, 42(2), 189–214.
- Chesney, M. A., & Rosenman, R. H. (Eds.). (1985). *Anger and hostility in cardiovascular and behavioral disorders*. Washington, DC: Hemisphere.
- Center for Women's Health Research. (2007). *2007 North Carolina women's health report card*. Chapel Hill: University of North Carolina.
- Comstock, G. W. (1979). The association of water hardness and cardiovascular disease: An epidemiological review and critique. In U.S. National Committee for Geochemistry (Ed.), *Geochemistry of water in relation to cardiovascular disease* (pp. 46–68). Washington, DC: National Academy of Sciences.
- Cooper, R., Cutler, J., Desvigne-Nickens, P., Fortmann, S. P., Friedman, L., Havlik, R., et al. (2000). Trends and disparities in coronary heart disease, stroke, and other cardiovascular diseases in the United States: Findings of the National Conference on Cardiovascular Disease Prevention. *Circulation*, 102, 3137–3147.
- Corin, E. (1994). The social and cultural matrix of health and disease. In R. G. Evans, M. L. Barer, & T. R. Marmor (Eds.), *Why are some people healthy and others not?* New York: Aldine de Gruyter.
- Cossman, R. E., Cossman, J. S., Jackson, R., & Cosby, A. (2003). Mapping high or low mortality places across time in the United States: A research note on a health visualization and analysis project. *Health and Place*, 9, 361–369.
- Davis, D. (2007). *The secret history of the war on cancer*. New York: Basic Books.
- Dear, M. J., & Wolch, J. (1987). *Landscapes of despair: From deinstitutionalization to homelessness*. Princeton, NJ: Princeton University Press.
- Doyal, L. (1995). *What makes women sick?: Gender and the political economy of health*. New Brunswick, NJ: Rutgers University Press.
- Durant, S. T., & Nickoloff, J. A. (2005). Good timing in the cell cycle for precise DNA repair by BRCA1. *Cell Cycle*, 4(9), 1216–1222.
- Editorial Committee for the Atlas of Cancer Mortality in the People's Republic of China. (1979). *Atlas of cancer mortality in the People's Republic of China*. Shanghai: China Map Press.
- Evans, R. G., Barer, M. L., & Marmor, T. R. (Eds.). (1994). *Why are some people healthy and others not?* New York: Aldine de Gruyter.
- Ewald, P. W. (2000). *Plague time: The new germ theory of disease*. New York: Free Press.
- Ezzati, M., Friedman, A. B., Kulkarni, S. C., & Murray, C. J. L. (2008). The reversal of fortunes: Trends in county mortality and cross-county mortality disparities in the United States. *PLoS Medicine*, 5(4), e66.

- Fang, J., Madhavan, S., & Alderman, M. H. (1996). The association between birthplace and mortality from cardiovascular causes among black and white residents of New York City. *New England Journal of Medicine*, 335, 1545–1551.
- Geronimus, A. T., Bound, J., Waidmann, T. A., Hillemeier, M. M., & Burns, P. B. (1996). Excess mortality among blacks and whites in the United States. *New England Journal of Medicine*, 335, 1552–1558.
- Gillum, R. F. (1996). The epidemiology of cardiovascular disease in black Americans [Editorial]. *New England Journal of Medicine*, 335, 1597–1599.
- Glick, B. J. (1979a). Distance relationships in theoretical models of carcinogenesis. *Social Science and Medicine*, 13D, 253–256.
- Glick, B. J. (1979b). The spatial autocorrelation of cancer mortality. *Social Science and Medicine*, 13D, 123–130.
- Glick, B. J. (1980). The geographic analysis of cancer occurrence: Past progress and future directions. In M. S. Meade (Ed.), *Conceptual and methodological issues in medical geography* (pp. 170–193). Chapel Hill: Department of Geography, University of North Carolina.
- Glick, B. J. (1982). The spatial organization of cancer mortality. *Annals of the Association of American Geographers*, 72, 471–481.
- Greenberg, M. R. (1983). *Urbanization and cancer mortality: The United States experience, 1950–1975*. New York: Oxford University Press.
- Greenberg, M. R. (1989). Black male cancer and American urban health policy. *Journal of Urban Affairs*, 11, 113–130.
- Hall, E. (2004). Spaces and networks of genetic knowledge making: The “geneticisation” of heart disease. *Health and Place*, 10, 311–318.
- Hanchette, C. (1988). *Geographical patterns of prostate cancer mortality: An investigation of the relationship between prostate cancer and ultraviolet light*. Unpublished master’s thesis, University of North Carolina.
- Hanchette, C. L. (1998). *The geographic modeling of lead poisoning risk in North Carolina*. Unpublished doctoral dissertation, University of North Carolina.
- Hanchette, C. L. (2008). The political ecology of lead poisoning in eastern North Carolina. *Health and Place*, 14(2), 209–216.
- Hanchette, C. L., & Schwartz, G. G. (1992). Geographic patterns of prostate cancer mortality. *Cancer*, 70, 2861–2869.
- Harpham, T., Lusty, T., & Vaughan, P. (Eds.). (1988). *In the shadow of the city: Community health and the urban poor*. New York: Oxford University Press.
- Hedgecoe, A. (2001). Schizophrenia and the narrative of enlightened geneticisation. *Social Studies of Science*, 31(6), 875–911.
- Hedgecoe, A. (2002). Reinventing diabetes: Classification, division and the geneticisation of diseases. *New Genetics and Society*, 21(1), 7–27.
- Jones, K., & Moon, G. (1987). *Health, disease and society: An introduction to medical geography*. London: Routledge.
- Kagami, M. (1991). A geographical study on regional variations of disease mortality. *Science Report, Institute of Geoscience, University of Tsukuba*, 12(Sect. A), 65–89.
- Koblinsky, M., Timyan, J., & Gay, J. (Eds.). (1993). *The health of women in global perspective*. Boulder, CO: Westview Press.
- Kung, H.-C., Hoyert, D. L., Xu, J.-Q., & Murphy, S. L. (2008). Death: Final data for 2005. *National Vital Statistics Reports*, 56(10).
- Lenfant, C. (1996). Conference on socioeconomic status and cardiovascular health and disease. *Circulation*, 94, 2041–2044.
- Lewin, R. (1984). Trail of ironies to Parkinson’s disease. *Science*, 224, 1083–1085.
- Lilienfeld, A. M., & Lilienfeld, D. E. (1980). *Foundations of epidemiology* (2nd ed.). New York: Oxford University Press.
- Lippman, A. (1998). The politics of health: Geneticisation versus health promotion. In S. Sherwin

- (Ed.), *The politics of women's health: Exploring agency and autonomy* (pp. 64–82). Philadelphia: Temple University Press.
- Lowis, G. W. (1986). Sociocultural and demographic factors in the epidemiology of multiple sclerosis: An annotated selected bibliography. *International Journal of Environmental Studies*, 26, 295–320.
- Marmot, M. G., & Mustard, J. F. (1994). Coronary heart disease from a population perspective. In R. G. Evans, M. L. Barer, & T. R. Marmor (Eds.), *Why are some people healthy and others not?* (pp. 189–214). New York: Aldine de Gruyter.
- Marmot, M. G., & Theorell, T. (1988). Social class and cardiovascular disease: The contribution of work. *International Journal of Health Services*, 18, 659–674.
- Mason, T. J., McKay, F. W., Hoover, R., Blot, W. J., & Fraumeni, J. F., Jr. (1975). *Atlas of cancer mortality for U.S. counties: 1950–1969*. Washington, DC: U.S. Department of Health, Education and Welfare, Epidemiology Branch, National Cancer Institute.
- Mason, T. J., McKay, F. W., Hoover, R., Blot, W. J., & Fraumeni, J. F., Jr. (1976). *Atlas of cancer mortality among U.S. nonwhites: 1950–69*. Washington, DC: U.S. Department of Health, Education and Welfare, Epidemiology Branch, National Cancer Institute.
- Mayer, J. D. (1996). The political ecology of disease as a new focus for medical geography. *Progress in Human Geography*, 20, 441–456.
- McGlashan, N. D. (1972). Food contaminants and oesophageal cancer. In N. D. McGlashan (Ed.), *Medical geography: Techniques and field studies* (pp. 247–257). London: Methuen.
- Meade, M. S. (1979). Cardiovascular mortality in the southeastern United States: The coastal plain enigma. *Social Science and Medicine*, 13D, 257–265.
- Meade, M. S. (1980). An interactive framework for geochemistry and cardiovascular disease. In M. S. Meade (Ed.), *Conceptual and methodological issues in medical geography* (pp. 194–221). Chapel Hill: Department of Geography, University of North Carolina.
- Meade, M. S. (1983). Cardiovascular disease in Savannah, Georgia. In N. D. McGlashan & J. R. Blunden (Eds.), *Geographical aspects of health* (pp. 175–196). New York: Academic Press.
- Michaels, D. (2008). *Doubt is their product: How industry's assault on science threatens your health*. New York: Oxford University Press.
- Mill, J. S. (1856). *A system of logic*. London: Parker, Son & Bowin.
- Minowa, M., Shigematsu, I., Nagai, M., & Fukutomi, K. (1981). Geographical distribution of lung cancer mortality and environmental factors in Japan. *Social Science and Medicine*, 15D, 225–231.
- Molla, M. T., Madans, J. H., & Wagener, D. K. (2004). Differentials in adult mortality and activity limitation by years of education in the United States at the end of the 1990s. *Population and Development Review*, 30(4), 625–646.
- Murray, C. J. L., Kulkarni, S. C., Michaud, C., Tomijima, N., Bulzacchelli, M. T., Landiorio, T. J., et al. (2006). Eight Americas: Investigating mortality disparities across races, counties, and race–counties in the United States. *PLoS Medicine*, 3(9), e260.
- National Center for Health Statistics. (1997). *Atlas of United States mortality*. Washington, DC: U.S. Government Printing Office.
- National Center for Health Statistics. (1998). Table 8: Deaths and death rates for the 10 leading causes of death in specified age groups, by race and sex: United States, 1996. *National Vital Statistics Reports*, 47, 26–36.
- National Center for Health Statistics. (2007). *Final Deaths 2005*. Washington, DC: Government Printing Office.
- North Carolina Department of Health and Human Services. (2004). *Racial and ethnic differences in health in North Carolina: 2004 update* (A special report for the State Center for Health Statistics and Office of Minority Health and Health Disparities). Raleigh, NC: Author.
- North Carolina State Center for Health Statistics. (2003). *Health risks among North Carolina adults: 2001* (Report from the Behavior Risk Factor Surveillance system). Raleigh, NC: Author.
- North Carolina State Center for Health Statistics. (2007). *Behavior Risk Factor Surveillance System (BRFSS)*. Retrieved from www.SCHS.state.nc.us

- Ohno, Y., & Aoki, K. (1981). Cancer death by city and county in Japan, 1969–1971: A test of significance for geographical clusters of disease. *Social Science and Medicine*, 15D, 251–258.
- Schroeder, H. A. (1966). Municipal drinking water and cardiovascular death rates. *Journal of the American Medical Association*, 195, 81–85.
- Schwartz, G., & Hanchette, C. L. (2006). UV, latitude, and spatial trends in prostate cancer mortality: All sunlight is not the same (United States). *Cancer Causes and Control*, 17(8), 1091–1101.
- Singh, G. K. (2003). Area deprivation and widening inequalities in US mortality 1969–1998. *American Journal of Public Health*, 93, 1137–1148.
- Steingraber, S. (1998). *Living downstream*. New York: Vintage Books.
- Takahashi, E. (1981). Geographic distribution of cerebrovascular disease and environmental factors in Japan. *Social Science and Medicine*, 15D, 153–172.
- Travis, J. (1993). New piece in Alzheimer's puzzle. *Science*, 261, 828–829.
- U.S. Environmental Protection Agency (EPA). (1987). *The United States cancer mortality rates and trends 1950–1979: Vol. 4. Atlas*. Washington, DC: U.S. Government Printing Office.
- Vance, J. M. (2008). *Parkinson report*. Miami, FL: National Parkinson Foundation.
- Waldron, I. (2005). Gender differences in mortality: Causes and variation in different societies. In P. Conrad (Ed.), *The sociology of health and illness: Critical perspectives* (7th ed.). New York: St. Martin's Press.
- Wallace, R., Huang, Y. S., Gould, P., & Wallace, D. (1997). The hierarchical diffusion of AIDS and violent crime among U.S. metropolitan regions: Inner-city decay, stochastic resonance and reversal of the mortality transition. *Social Science and Medicine*, 44, 935–947.
- West, R. R. (1977). Geographic variation in mortality from ischaemic heart disease in England and Wales. *British Journal of Preventive and Social Medicine*, 31, 245–250.
- Williams, E. D., Jones, D. R., & McKeran, R. O. (1991). Mortality rates from multiple sclerosis: Geographical and temporal variations revisited. *Journal of Neurology, Neurosurgery and Psychiatry*, 54, 104–109.
- Wolch, J., & Dear, M. (1993). *Malign neglect: Homelessness in an American city*. San Francisco: Jossey-Bass.
- World Health Organization (WHO). (2004a). *Global burden of disease update*. Geneva: Author.
- World Health Organization (WHO). (2004b). *World health report annex*. Geneva: Author.
- World Health Organization (WHO). (2007). Health inequality, inequity, and social determinants of health. *Population and Development Review*, 33(4), 839–843.
- Yanos, P. T. (2007). Beyond “Landscapes of Despair”: The need for new research on the urban environment, sprawl, and the community integration of persons with severe mental illness. *Health and Place*, 13, 672–676.
- Ziegler, R. G., Hoover, R. N., Pike, M. C., Hildesheim, A., Nomura, A. M. Y., West, D. W., et al. (1993). Migration patterns and breast cancer risk in Asian-American women. *Journal of the National Cancer Institute*, 85, 1819–1827.

FURTHER READING

- Anderson, R. (1984). Temporal trends of cancer mortality in eastern New England compared to the nation, 1950–1975. *Social Science and Medicine*, 19, 749–757.
- Andrews, H. F. (1985). The ecology of risk and the geography of intervention: From research to practice for the health and well-being of urban children. *Annals of the Association of American Geographers*, 75, 370–382.
- Armstrong, R. W. (1971). Medical geography and its geologic substrate. *Geological Society of America Memoir*, 123, 211–219.
- Casper, M. L., Barnett, E., Halverson, J. A., Elmes, G. A., Braham, V. E., Majeed, Z. A., et al. (2000). *Women and heart disease: An atlas of racial and ethnic disparities in mortality* (2nd ed.). Morgantown: Office for Social Environment and Health Research, West Virginia University.

- Cleek, R. K. (1979). Cancer and the environment: The effect of scale. *Social Science and Medicine*, 13D, 241–247.
- Gardner, M. (1976). Soft water and heart disease. In J. Leniham & W. W. Fletcher (Eds.), *Health and the environment* (pp. 116–135). New York: Academic Press.
- Gardner, M. J., Winter, P. D., & Acheson, E. D. (1982). Variations in cancer mortality areas in England and Wales: Relation with environmental factors and search for cause. *British Medical Journal*, 284, 284–287.
- Howe, G. M., Burgess, L., & Gatenby, P. (1977). Cardiovascular disease. In G. M. Howe (Ed.), *A world geography of human diseases* (pp. 431–476). New York: Academic Press.
- Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H., & Subramanian, S. V. (2005). Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: The Public Health Disparities Geocoding Project. *American Journal of Public Health*, 95, 312–323.
- Lewis, N. (1998). Intellectual intersection: Gender and health in the Pacific. *Social Science and Medicine*, 46, 641–659.
- Lux, W. E., & Kurtzke, J. F. (1987). Is Parkinson's disease acquired?: Evidence from a geographic comparison with multiple sclerosis. *Neurology*, 37, 467–471.
- McLafferty, S., & Tempalski, B. (1995). Restructuring and women's reproductive health: Implications for low birthweight in New York City. *Geoforum*, 26, 309–323.
- Pyle, G. F. (1971). *Heart disease, cancer, and stroke in Chicago* (Research Paper No. 134). Chicago: University of Chicago, Department of Geography.
- Pyle, G. F. (1979). *Applied medical geography*. New York: Wiley.
- Shigematsu, I. (1981). *National atlas of major disease mortality in Japan*. Tokyo: Japan Health Promotion Foundation.
- Waterhouse, J., Correa, P., Muir, C., & Powell, J. (Eds.). (1976). *Cancer incidence in five continents*. Lyon, France: International Agency for Research on Cancer.
- Weinstein, M. S. (1980). *Health in the city: Environmental and behavioral influences*. New York: Pergamon Press.

CHAPTER 9

Neighborhoods and Health

THE CONCEPT OF NEIGHBORHOOD HEALTH

Although for the past century public health studies have typically involved the identification of individual risk factors for health and disease, recently there has been increasing appreciation of the role spatial dimensions play in determining health. A growing number of studies consider *neighborhood-level* determinants of health (Curtis, 1990; Diez-Roux, 1998, 2001; Kawachi & Berkman, 2003; Macintyre, Ellaway, & Cubbins, 2002; Schwartz, 1994; Susser, 1994). Risk and protective factors may inherently reside at the community level, and therefore measurement and intervention should take place at that same level. Spatial variations in health outcomes are caused by variations in the people who live in the areas and in the places themselves. The goal of such studies is to differentiate between *contextual* (area) effects and *compositional* (population) effects of a neighborhood on health.

Levels of influences on individual health are often explained through an analogy to a stream. Upstream influences may include a political decision at the national level, such as corn subsidies that may indirectly lead to increased obesity rates. Downstream influences may include individuals' choices to "supersize" their meals at the local fast-food restaurant. These upstream and downstream influences on health can also be conceptualized spatially. Figure 9.1 shows how individuals are part of neighborhoods, communities, and regions. Individual health can be influenced by neighborhood-level processes, such as whether or not there is a fast-food restaurant near a person's home, or further upstream at the community level. For instance, the town (community) that one of your authors (Emch) grew up in did not have any fast-food restaurants at all because the local government did not allow it.

A growing body of literature has demonstrated that overall living environments, including the neighborhoods in which people reside, have important effects on health. In studies using multilevel statistical methods to disentangle commu-

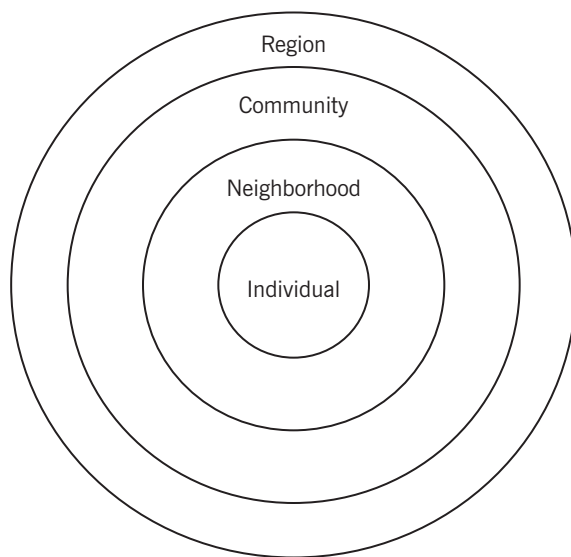


FIGURE 9.1. Neighborhoods within a nested spatial hierarchy.

nity-level and individual-level effects, neighborhood characteristics have been shown to independently influence a wide variety of health outcomes, including cardiovascular disease, mental health, overall mortality, adverse birth outcomes, self-rated health, asthma, and Lyme disease (Jones & Duncan, 1995). Neighborhood context has also been shown to affect behaviors and health outcomes, such as physical activity (Gordon-Larsen, Nelson, Page, & Popkin, 2006; Gordon-Larsen & Reynolds, 2006), food choices (Popkin, Duffey, & Gordon-Larsen, 2005), crime (Morenoff, Sampson, & Raudenbush, 2001; Sampson, Raudenbush, & Earls, 1997), hypertension (Morenoff et al., 2007), heart disease (Diez-Roux et al., 2001), smoking (Barnett, 2000; Barnett, Moon, & Kearns, 2004; Barnett, Pearce, & Moon, 2005; Moon & Barnett, 2003; Pearce, Hiscock, Moon, & Barnett, 2009; Twigg, Barnett, Pearce, & Moon, 2000), and self-perceived health (Dunn, Veenstra, & Ross, 2006). There is growing awareness of the role the human-made environment plays in health. For example, the layout of a community affects the occurrence of cardiovascular disease, mental health, and Lyme disease among residents.

The *neighborhood approach* to analyzing disease outcomes is similar to the medical geographic approach of disease ecology and its analytical approaches. The neighborhood approach is a conceptual framework that is spatial in nature and can be used to address the complexities of social and environmental correlates of disease over time and space. The medical geographic tradition of disease ecology is not only spatial, but holistic; it includes the integration of many different types of variables responsible for disease. The theory of disease ecology fits into both the spatial organization and human-environment traditions of geography. The fundamental question asked in studies using this approach is "What factors are associated with the time-space variation of disease?" It emphasizes the complex sets of interactions between people and their environments.

NEIGHBORHOOD DEFINITION AND UNITS

A neighborhood is not simply a spatial unit. You can visualize the neighborhood that you grew up in, but your former next-door neighbor may have a different area in mind for the boundaries of the neighborhood. In neighborhood health studies, an area specifically defined as a *neighborhood* is the unit of analysis. Such units may represent traditional neighborhood boundaries; for instance, Chicago is divided into 77 community areas that serve as the basis for urban planning (see Figure 9.2). Other neighborhood health studies may use local census units (such as blocks, block groups, or tracts) or alternative areal aggregations (including zip codes). When conducting such a study, the investigators need to think about the spatial scale, but unfortunately many do not. Orange County, North Carolina, home of the University of North Carolina at Chapel Hill (see Figure 9.3), is used below as an example to explain this concept.

There were 118,228 people living in Orange County at the time of the 2000 U.S. census. When one is studying the context of where they live as it relates to their health, the choice of the unit of analysis could influence the results of the study. For instance, if one were to study the impact of community context (e.g., crime, sidewalks, bike trails, playgrounds) on obesity rates, then dividing Orange County into different-sized census units would determine whether the study was based on a very local context (census blocks) or much coarser units (block groups or tracts). Figure 9.4 shows the various-sized census units that could be used in such a study.

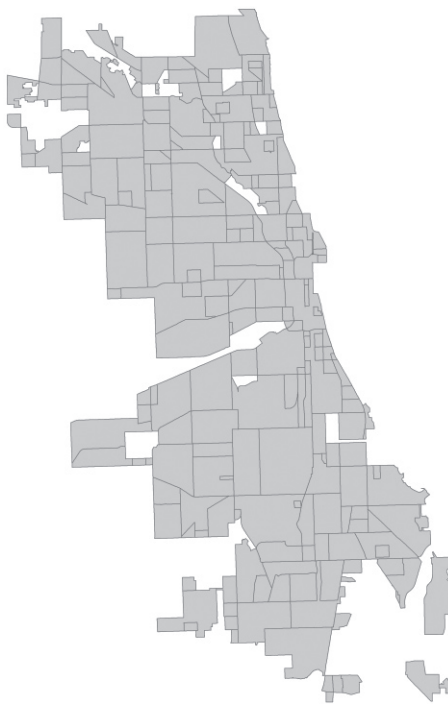


FIGURE 9.2. Chicago community areas.



FIGURE 9.3. Location of Orange County, North Carolina.

They divide the people into from 21 to 2,165 areas. Think about where you live and your choice of whether you will walk or drive to do daily activities, such as shopping, exercise, school, and work. If you consider the small area around your home and a variable such as the number of fast-food restaurants, and then consider the larger block groups and tracts, then that variable would be very different in the larger units. The issue is one of scale. If you have a fast-food restaurant in your block area, then you might go there more often than if it was a mile away. So if you are trying to measure the contextual effect of fast-food restaurants in close proximity to people's homes, and the effect is a very local-scale process, then the measurement of the effect must be at a local scale.

Many studies, however, do not use an appropriate spatial scale when measuring neighborhood contextual effects; they simply analyze data that are conveniently available. Many studies use zip code areas because they are widely available, but as Figure 9.5 illustrates, they are very large, irregular units that are used to deliver mail efficiently. Some neighborhood health studies have even used counties to study what the authors call "neighborhood effects." Although there is no unified definition of *neighborhoods*, they usually refer to more local-level areal patterns than would be found by analyzing the county-level or zip-code-level context. Geographers also are often concerned about what delimits, or forms, a neighborhood as a functional



FIGURE 9.4. Orange County, North Carolina: Census geography.

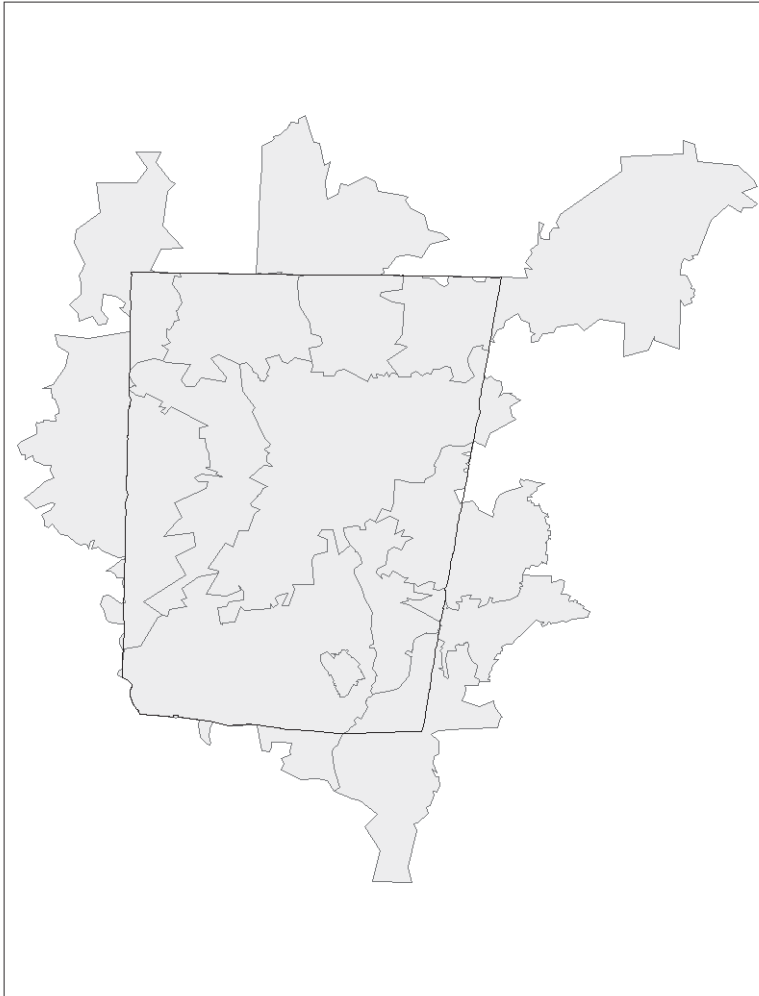


FIGURE 9.5. Orange County, North Carolina: Zip code boundaries.

region of interaction. Where, for example, do children go to school? From how far do people come to use a park? Where do people go for health care, shopping, or church? These social activity areas can define a functional neighborhood relevant to environmental exposures or epidemic diffusion (Bian, 2004), as discussed in other chapters.

Neighborhoods do not need to be defined by existing areal units, and sometimes it makes sense simply to consider the area within a given distance of a house. Figure 9.6 shows a hypothetical 1,000-meter buffer (Euclidean distance area) around house number 1, with 9 other houses (2 through 9) also in the neighborhood. Because houses 11 and 12 are outside the 1,000-meter buffer area, they are considered to be outside the neighborhood or area of influence. This type of neighborhood area is

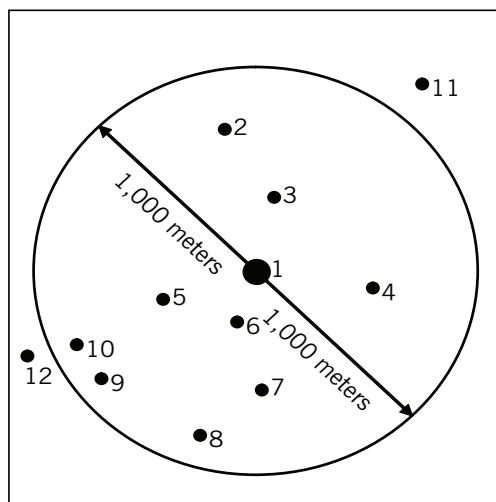


FIGURE 9.6. A neighborhood defined in terms of a 1,000-meter buffer around house 1.

easy to construct when there is a geographic information system (GIS) database for the study area.

The environmental context of the area being studied can also be used to define a neighborhood. Figure 9.7 shows an area in rural Bangladesh where a neighborhood study of cholera, a water-borne disease, was recently done (Emch, Ali, Root, & Yunus, 2009). In the zoomed-in view of the satellite image on the right, the light gray squares represent two hypothetical households that are connected via a water body. A neighborhood in this case could be defined as all of the households in close proximity to a common water body.

Figure 9.8 is a GIS database for this study area in rural Bangladesh. The database was created by digitizing ponds from satellite imagery and overlaying them with household points surveyed by using global positioning system (GPS) receivers. In the Emch and colleagues (2009) study, neighborhoods were constructed by modeling connectivity of households via water bodies, and cholera incidence was

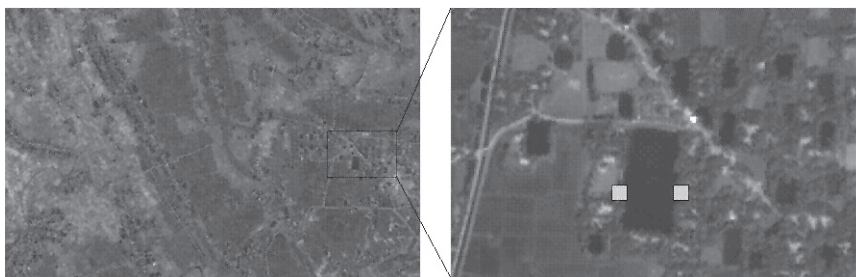


FIGURE 9.7. Households connected via ponds. From Emch, Ali, Root, and Yunus (2009).



FIGURE 9.8. Household clusters near digitized ponds. From Emch, Ali, Root, and Yunus (2009).

calculated for the different neighborhoods and compared to ecological variables. This section has shown various ways to define neighborhoods at different scales; all of these methods can be valid, given the right research problem.

ANALYZING NEIGHBORHOOD EFFECTS ON HEALTH

Once a suitable neighborhood unit has been determined, the analyst must differentiate between neighborhood-level and individual-level determinants of health. The goal is to disentangle the context of a neighborhood from the composition of the population. *Context* is about place, and *composition* is about people. The following is an example of how one might study a health outcome that involves both compositional and contextual factors. In North Carolina many birth defect rates exhibit a significant amount of spatial variation (Root, Meyer, & Emch, 2009). Environmental health studies suggest that these patterns are due to toxicants in a mother's environment (contextual factors). However, there is also evidence that birth defects are related to both behavioral risk factors (e.g., smoking) and socioeconomic risk factors (compositional factors). Spatial patterns of birth defects may be due to the fact that individuals with similar backgrounds and risk factors live in the same geographic area, producing larger area-level patterns of disease.

A neighborhood health study distinguishes between the environmental and socioeconomic factors that explain the geographic variation. The study in North Carolina found that gastroschisis (an abdominal wall defect) is more common in births to poor white women than in births to other women, and that there are significant clusters of gastroschisis births in the piedmont region of the state, which are hypothesized to be related to environmental factors (Root et al., 2009). Few studies of birth defects use methods appropriate for determining whether differences in prevalence across areas are due to characteristics of the areas themselves (contextual factors) or to differences between the types of people living in the areas (compositional factors). To address this limitation, multilevel regression models (see Vignette 12.2 for a description of regression analysis) can be used to pull out the independent effects of individual-level and area-level characteristics on birth defect outcomes. When modeling area-level variables, one must avoid committing the ecological fallacy—that is, avoid making inferences about individuals when one is investigating aggregations of individuals. (See Chapter 1; Chapter 12 provides a more comprehensive discussion of the ecological fallacy in a geographic context.)

EFFECTS OF THE BUILT ENVIRONMENT ON HEALTH

This entire book has thus far been about how geography influences health and health care. Where we live and where we go will affect our health. Much of Chapters 4 and 6 are about how the natural environment affects our health. Chapter 7 is about how human impacts on the environment come back at us to affect our health. The way we plan and build our neighborhoods, towns, and cities also influences our health. The concept of how our built environment affects us can be intuitively understood through the example of obesity. Although there may be genetic differences between individuals in whether they are more or less susceptible to becoming overweight or obese, much of the recent obesity epidemic can be attributed to changes in caloric input versus output. There are clearly more people taking in more calories and burning off fewer than in the past. The built environment can influence these patterns at a population level.

The walkability of a neighborhood can foster both physical activity and a sedentary lifestyle. The proximity of people's homes to other people's homes, measured by calculating the population density, can be important. If there are more people living in an area, then there are more opportunities to interact without driving. The proximity of people to commercial enterprises is also important. Mixed-use areas with residential, industrial, and commercial space foster walking because people can walk to their jobs and to shopping centers or to recreation areas. Connectivity also influences walkability. An isolated neighborhood with many dead ends that do not connect to other residential or commercial areas does not encourage walking. A neighborhood that is more "obesigenic"—where people have low physical activity and therefore higher obesity rates—will have low population density, fewer travel destinations within it, single-use zoning, and low connectivity.

The increasing obesity rates do not just involve caloric output, but also caloric input. Therefore, the local food environment is probably just as important as the walkability of a neighborhood. Access to supermarkets, exposure to fast food, and availability of healthy food options are all important aspects of the local food environment. The nutritional environment, including prices of different types of foods, is important to understand how neighborhood context can influence obesity rates. An aspect of the poverty syndrome discussed in Chapter 8 is the absence of outlets for fresh vegetables and fruits in many inner-city neighborhoods. Beyond the availability of food, advertising is also important. Fast-food restaurants advertise on billboards in some places more than others, and it's unlikely that this does not have an effect on people's usage patterns. The fast-food restaurants themselves have built-in advertising with easily recognizable signs. Often children can recognize these signs at a very early age, which is hardly surprising, given that the mascots of these restaurants are clowns or cartoon characters.

THE CHALLENGES OF NEIGHBORHOOD HEALTH STUDIES

Neighborhood health studies are difficult to implement, and there are many challenges to doing so. One major challenge, the so-called *endogeneity problem*, is that there may be two-way causality. In other words, we do not know whether the obesity is caused by the neighborhood. Skinny people may move to neighborhoods with walking trails (or build them once they live there), and fast-food restaurants may place themselves where people want to eat their food. Studies also need to take into account the dynamic nature of neighborhoods, and thus a *longitudinal* approach to the phenomena being studied is necessary. Longitudinal data sets are those that require measurement of phenomena at many different times. This is expensive and therefore is not often done. One study that follows a longitudinal design is the China Health and Nutrition Survey, which has collected information on 19,000 individuals in nine provinces in 2006 and 2009, and will do so again in 2011 (Popkin, 2008; Popkin & Du, 2003).

Much of the literature on neighborhoods and health, while it is ultimately interested in how health varies by spatial units, has in fact been *aspatial*. For instance, a study that uses census tracts because they are readily available, and uses a traditional multilevel regression modeling approach that does not take into account interaction between areas, is not spatial. The latter approach (using multilevel regression modeling that does not take into account spatial interaction between adjacent areas) can be particularly problematic. This approach probably does not represent reality in most cases because it does not consider the fact that we are more likely to visit the neighborhood next to our own more often than another neighborhood that is farther away. In the birth defects example above, a multilevel regression model would assume that people are only exposed to toxicants in their own neighborhood. In such a model, neither adjacent neighborhoods nor neighborhoods that are far away would be assumed to have an effect on the health outcome. This assumption

is likely to be true for the neighborhoods far away, but the adjacent neighborhoods are likely to have some effect on a person's risk of birth defects. The answer to this problem is something called a *spatial multilevel model*, which is an area of statistics that is in development.

An aspatial approach may not reveal reasons for spatial differences in the health outcome being studied. Alternatively, a study that starts with a theory about why a health outcome is related to space and proceeds to investigate the spatial variation of that outcome rigorously is more likely to be successful. Also, health outcomes may be multiscalar processes. People living in a particular census block in a certain town are influenced by the small neighborhood around them as well as by their neighbors themselves, from whom they get information, contagion, and help. But they are also influenced by their community decisions about planning and zoning, as well as by state-, regional-, and national-level organizations.

CONCLUSION

Studies of neighborhoods and health have become common during the past decade because health is not just about individuals. It is about the context of people's lives, including where they live, work, and perform other activities. The pendulum has swung away from studies assuming that health risk is only related to individual biology to a more holistic approach that looks at complex contextual reasons for health variations in the population. Although this latter type of research has been termed *neighborhood health studies*, it does not only encompass neighborhoods as traditionally defined. Any areal unit can be considered a neighborhood, including a census unit, a Euclidean distance area around a household, or even a town or county. The areal unit is all about scale, and some processes that may affect a population's health are at a coarser scale and others at a finer scale. Although there are still methodological limitations to our studies of how areal context influences a population's health, this trend in health research is an advance that better reflects the complex reasons why and where people get diseases. This approach to disease studies is very much in the spirit of the long tradition of disease ecology described in Chapter 2.

REFERENCES

- Barnett, J. R. (2000). Does place of residence matter?: Contextual effects and smoking in Christchurch. *New Zealand Medical Journal*, 113, 433–435.
- Barnett, J. R., Moon, G., & Kearns, R. (2004). Social inequality and ethnic differences in smoking in New Zealand. *Social Science and Medicine*, 59(1), 129–143.
- Barnett, R., Pearce, J., & Moon, G. (2005). Does social inequality matter?: Changing ethnic socio-economic disparities and Maori smoking in New Zealand, 1981–1996. *Social Science and Medicine*, 60(7), 1515–1526.
- Bian, L. (2004). A conceptual framework for an individual-based spatially explicit epidemiological model. *Environment and Planning B*, 31(3), 381–395.
- Curtis, S. E. (1990). Use of survey data and small area statistics to assess the link between individ-

- ual morbidity and neighborhood deprivation. *Journal of Epidemiology and Community Health*, 44, 62–68.
- Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *American Journal of Public Health*, 88, 287–293.
- Diez-Roux, A. V. (2001). Investigating neighborhood and area effects on health. *American Journal of Public Health*, 91(11), 1783–1789.
- Diez-Roux, A. V., Stein Merkin, S., Arnett, D., Chambless, L., Massing, M., Nieto, F. J., et al. (2001). Neighborhood of residence and incidence of coronary heart disease. *New England Journal of Medicine*, 345, 99–106.
- Dunn, J. R., Veenstra, G., & Ross, N. A. (2006). Relations between self-perceived relative social position and self-rated health: Results from a nationally representative survey of the Canadian population. *Social Science and Medicine*, 62(6), 1465–1473.
- Emch, M., Ali, M., Root, E. D., & Yunus, M. (2009). Spatial and environmental connectivity analysis in vaccine trials. *Social Science and Medicine*, 68, 631–637.
- Gordon-Larsen, P., Nelson, M. C., Page, P., & Popkin, B. M. (2006). Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics*, 117(2), 417–424.
- Gordon-Larsen, P., & Reynolds, K. D. (2006). Influence of the built environment on physical activity and obesity in childhood and adolescents. In M. I. Goran & M. Sothorn (Eds.), *Handbook of pediatric obesity: Etiology, pathophysiology, and prevention* (pp. 255–270). Boca Raton, FL: CRC/Taylor & Francis.
- Jones, K., & Duncan, C. (1995). Individuals and their ecologies: Analyzing the geography of chronic illness within a multilevel modeling framework. *Health and Place*, 1, 27–30.
- Kawachi, I., & Berkman, L. (Eds.). (2003). *Neighborhoods and health*. New York: Oxford University Press.
- Macintyre, S., Ellaway, A., & Cubbins, S. (2002). Place effects on health: How can we conceptualize, operationalize, and measure them? *Social Science and Medicine*, 55, 125–139.
- Moon, G., & Barnett, J. R. (2003). Spatial scale and the geography of tobacco smoking in New Zealand: A multilevel perspective. *New Zealand Geographer*, 59(2), 6–15.
- Morenoff, J. D., James, S., Hansen, B. B., Williams, D., Kaplan, G., & Hunte, H. (2007). Understanding social disparities in hypertension prevalence, awareness, treatment, and control: The role of neighborhood context. *Social Science and Medicine*, 65, 1853–1866.
- Morenoff, J. D., Sampson, R. J., & Raudenbush, S. W. (2001). Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology*, 39(3), 517–558.
- Pearce, J., Hiscock, R., Moon, G., & Barnett, J. R. (2009). The neighbourhood effects of geographical access to tobacco retailers on individual smoking behaviour. *Journal of Epidemiology and Community Health*, 63(1), 69–77.
- Popkin, B. M. (2008). Will China's nutrition transition overwhelm its health care system and slow economic growth? *Health Affairs*, 27(4), 1064–1076.
- Popkin, B. M., & Du, S. (2003). Dynamics of the nutrition transition toward the animal foods sector in China and its implications: A worried perspective. *Journal of Nutrition*, 133, 3898S–3906S.
- Popkin, B. M., Duffey, K., & Gordon-Larsen, P. (2005). Environmental influences on food choice, physical activity and energy balance. *Physiology and Behavior*, 86(5), 603–613.
- Root, E. D., Meyer, R., & Emch, M. (2009). Evidence of localized clustering of gastroschisis births in North Carolina, 1999–2004. *Social Science and Medicine*, 68, 1361–1367.
- Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277, 918–924.
- Schwartz, S. (1994). The fallacy of the ecological fallacy: The potential misuse of a concept and the consequences. *American Journal of Public Health*, 84(5), 819–824.
- Susser, M. (1994). The logic in ecological: I. The logic of analysis. *American Journal of Public Health*, 84(5), 825–829.
- Twigg, L., Moon, G., & Jones, K. (2000). Predicting small-area health-related behaviour: A comparison of smoking and drinking indicators. *Social Science and Medicine*, 50, 1109–1120.

FURTHER READING

- Balfour, J. L., & Kaplan, G. (2002). Neighborhood environment and loss of physical function in older adults: Evidence from the Alameda County study. *American Journal of Epidemiology*, 155(6), 507–515.
- Buka, S. L., Brennan, R. T., Rich-Edwards, J. W., Raudenbush, S. W., & Earls, F. (2003). Neighborhood support and the birth weight of urban infants. *American Journal of Epidemiology*, 157(1), 1–8.
- Chaix, B., Merlo, J., Subramanian, S. V., Lynch, J., & Chauvin, P. (2005). Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: The case of mental and behavioral disorders due to psychoactive substance use in Malmo, Sweden, 2001. *American Journal of Epidemiology*, 162, 171–182.
- Chuang, Y., Cubbin, C., Ahn, D., & Winkleby, M. A. (2005). Effects of neighbourhood socioeconomic status and convenience store concentration on individual level smoking. *Tobacco Control*, 14(5), 337.
- Corburn, J., Osleeb, J., & Porter, M. (2006). Urban asthma and the neighbourhood environment in New York City. *Health and Place*, 12(2), 167–179.
- Diez-Roux, A. V., Nieto, F. J., Caulfield, L., Tyroler, H. A., Watson, R. L., & Szklo, M. (1999). Neighbourhood differences in diet: The Atherosclerosis Risk in Communities (ARIC) Study. *Journal of Epidemiology and Community Health*, 53(1), 55–63.
- Diez-Roux, A. V., Schwartz, S., & Susser, E. (2002). Ecologic studies and ecologic variables in public health research. In R. Detels, J. McEwen, R. Beaglehole, & H. Tanaka (Eds.), *The Oxford textbook of public health* (4th ed., pp. 493–508). London: Oxford University Press.
- Diez-Roux, A. V., Stein Merkin, S., Hannan, P., Jacobs, D. R., & Kiefe, C. I. (2002). Area characteristics, individual-level socioeconomic indicators, and smoking in young adults: The CARDIA study. *American Journal of Epidemiology*, 157(4), 315–326.
- Ecob, R., & Macintyre, S. (2000). Small area variations in health-related behaviors: Do these depend on the behavior itself, its measurement, or on personal characteristics? *Health and Place*, 6, 261–274.
- Emch, M., Ali, M., Park, J. K., Yunus, M., Sack, D., & Clemens, J. D. (2006). Relationship between neighborhood-level killed oral cholera vaccine coverage and protective efficacy: Evidence for herd immunity. *International Journal of Epidemiology*, 35, 1044–1050.
- Emch, M., Ali, M., Yunus, M., Sack, D., Acosta, C., & Clemens, J. D. (2007). Efficacy calculation in randomized vaccine trials: Global or local measures? *Health and Place*, 13, 238–248.
- Lee, R. E., & Cubbin, C. (2002). Neighborhood context and youth cardiovascular health behaviors. *American Journal of Public Health*, 92(3), 428–436.
- Mayer, S. E., & Jencks, C. (1989). Growing up in poor neighborhoods: How much does it matter? *Science*, 243, 1441–1445.
- O'Campo, P. (2003). Invited commentary: Advancing theory and methods for multilevel models of residential neighborhoods and health. *American Journal of Epidemiology*, 157, 9–13.
- Pickett, K. E., & Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health*, 55(2), 111–122.
- Sampson, R., Morenoff, J., & Gannon-Rowley, T. (2002). Assessing neighborhood effects: Social processes and new directions in research. *Annual Review of Sociology*, 28, 443–478.
- Yen, I., & Syme, S. L. (1999). The social environment and health: A discussion of the epidemiologic literature. *Annual Review of Public Health*, 20, 287–308.

CHAPTER 10

Disease Diffusion in Space

There are only two ways in which anything—whether it be an idea, a type of mosquito, a food crop, a sophisticated technology for hospital care, or an infectious agent of disease—can come to be found in a particular location. Either it developed there independently, or it somehow moved there from another place. Movement is by far the more usual explanation of the existence of phenomena at a particular location. An understanding of the mechanisms that influence the spread of any phenomenon and its spatial pattern thus created is at the core of the geographic study of diffusion. The term *diffusion* implies a spread or movement outward from a point or beginning place. Geographers have for decades studied diffusion of all manner of things: fish hooks, types of barns, musical instruments, religious beliefs, agricultural seeds, and communications technology, among hundreds of others. Medical geographers have studied diffusion of new medical technologies, equipment, treatments, hospices, and the like, but not very much or to any notable end. Mostly they have focused on the diffusion of infectious diseases, especially nonvectored ones. In these studies many of the conceptual and methodological concerns of epidemiology and geography have been combined. Geographers, however, want to know not only how a disease spreads in a population, but how that spread occurs over space; not only how many cases might occur at a future point in time, but *where* those cases are likely to occur.

Western civilization has long generally recognized the existence and importance of *disease diffusion*. Disease quarantine, which could involve an individual, a household, or even an entire community or country, at least tacitly implies the existence of disease diffusion. Indeed, the term *quarantine* comes from the medieval port city of Raguna (in present-day Italy), when in 1377 it tried to seclude incoming vessels for 40 days in an effort to prevent importation of plague. In the United States, Noah Webster's (1799) *A Brief History of Epidemic and Pestilential Diseases* traced the spread of several late 1700s epidemics across New England. At about that same

time, William Currie (1792, 1811) was also describing the geography of a number of epidemics in the United States. The contagionists investigating the distribution of yellow fever in the early 1800s recognized that the disease was carried from place to place, spread by an infected individual. Although we might find their medical explanations of causation amusing, the quality of their geographic descriptions was, given the severe data limitations, excellent.

Most studies of geographic diffusion have emphasized an explanation of events in a space–time context. The extent and form of human intercourse are critical for both innovation and diffusion. Variables such as distance, the locations of settlements, and the distribution of jobs and facilities all influence the level and direction of interaction. Diffusion is patterned by the configuration of the networks that encourage movement and the barriers that discourage it. A major prerequisite of disease diffusion, for example, is the existence of a sufficiently large susceptible population. A sufficiently large immune population (or, more precisely, a sufficiently small susceptible population) can serve as an effective barrier to disease transmission. Knowledge of where a disease has been before and how it moved through population settlements, therefore, can help epidemiologists and medical personnel target vaccination campaigns on transmission channels or establish buffer zones of immune people as barriers.

New infectious diseases and newly virulent or antibiotic-resistant strains of old ones have appeared in recent decades. The evolution and emergence of new disease agents, largely as a result of changes in land use and transformation of population mobility, are discussed in Chapters 4 and 5. That development lies in the cultural ecology of population–behavior–habitat interactions. This chapter focuses on behavior, specifically population mobility and moving around elements (i.e., agents and vectors) of a disease system (Chapter 2). Whether they spread only locally (such as Legionnaires' disease) or diffuse as great world pandemics (such as HIV/AIDS), the threat of emerging diseases challenges the development of diffusion prediction—which relates not just to the mathematical or virtual space of differential equations, but to real towns needing real health care personnel for sick people at specific locations. After a brief look at epidemiological terminology, types of diffusion and the importance of urban hierarchy are explained. The main forms of diffusion modeling are explained and illustrated. Diverse influenzas are modeled separately. Finally, the pandemic of AIDS has its own section with a spectrum of geographic approaches and methodologies: disease ecology, several types of diffusion modeling, and discourses on context and consequences for specific places.

TERMINOLOGY

Epidemiological Terminology

Epidemiologists are concerned with disease infections moving through a population (one that seems not to exist distributed in space but as a whole occupying a point). Several of the terms from Chapter 1 apply to this process (see Figure 1.3). A person can be infected but never develop symptoms (*subclinical*), or can develop

symptoms (*clinical*). The person can be even clinically sick, but not pass the infection to another person (insufficient *viremia*, a *dead end*). How likely is a disease to develop in an infected person (*pathogenicity of the agent*, *resistance of the host*)? How virulent is the infection that occurs (*case fatality rate*)? How likely is the infection to be passed on (*secondary attack rate*)? The *basic reproductive rate* is the average number of secondary infections produced when one infection occurs in a virgin population (i.e., one in which everyone is susceptible). How long is the time period between infection and when a person becomes infectious to others (*incubation period*, which is actually a little different from period and time to symptoms; *latency*, because people with some infections start to “shed” the disease agent in respiratory droplets, etc., before they actually develop symptoms such as fever)? People who can be infected are *susceptible*. Those who have already had the disease, in reality or through vaccination, and can no longer be infected are *immune*. Those who are actively able to pass the agent on are *infectious*. For some diseases it is possible for individuals to have passed through the course of the disease and no longer have symptoms themselves, yet still shed disease agents and so be *carriers*. “Typhoid Mary” is the infamous example: Mary Mallon, an Irish immigrant woman working in the New York City area in the late 19th and early 20th centuries as a cook, repeatedly infected people in sequential households where she worked, causing 56 cases of typhoid and several deaths. She was herself healthy and denied her role, believing that prejudice against the “dirty Irish” had led to her wanton persecution. She returned to work as a cook even after being released from 3 years of quarantine upon her signing an affidavit not to do so. After another death followed, Mary Mallon was forcibly quarantined for another 38 years until she died at 69. The case of Typhoid Mary spread the concept of *carrier* widely among the public, and of course led to new law on quarantine powers. This was invoked a few years ago, for example, when a man with multiple-antibiotic-resistant tuberculosis traveled by plane against recommendations.

Spread of contagion requires contact in space. *Distance* and *density* are important parameters. The likelihood of an infection’s spreading from one host to another depends on such things as whether the potential host is susceptible to the infection (age immunity, vaccination, nutritional status, etc.); whether the distance between them (survival of the agent in the environment) is bridgeable; and how likely one or more contacts are (density of population with the opportunities for other secondary infections). Likelihood of contact can be related to social networking as well as “herding” (into schools, prisons, nursing homes, etc.). The cultural process of herding, putting many animals into close proximity, helped create the intensity of transmission that led infectious agents to cross species and infect humans back at the beginning of domestication, densely settled agricultural villages, and the first cities. *Herd immunity* is the proportion of immunity in a specific population needed to prevent the infectious contact from finding a susceptible host, thereby ending the epidemic. *Epidemic* originally meant a disease’s spreading in a population, but in recent years has been used (so as to refer also to nontransmissible diseases) to mean a disease’s occurring in excess in a population. *Pandemic* means that a disease is spreading among populations across national borders. The World Health Organization (WHO) today technically uses the term *pandemic* to mean sustained spreading

within communities in two or more countries (stage 5 in the WHO's rating system) or in two or more WHO regions of the world (stage 6, globally pandemic).

Knowledge of these parameters enables researchers to model and predict future developments, as well as possible intervention strategies when an epidemic starts. If *this* many people are infected at a certain point in time, the argument goes, and *this* is the incubation period, the secondary attack rate, and the proportion of the population susceptible, then at the next point in time *that* many people will be infected. Such calculations can be complicated. The proportion of susceptible persons may vary with such characteristics as age, vaccination history (sometimes substituted for by income and education levels), or prevalent health status, for example. Nevertheless, epidemics in a population are often mathematically reduced to a set of differential equations. Epidemiologists with such knowledge can predict the need for hospital beds, the economic impact of work absenteeism, and such matters for the whole population. Researchers can model the comparative advantage of targeting various population subgroups for a vaccination drive, for example. Far too often in such efforts, however, the question of *where* is left out.

Much has been learned about the movement of a new infection through a population. An epidemic typically follows an S-shaped curve, or *logistic curve* (see Vignette 10.1). At first only a small proportion of the population is infected (*early innovators*, in geographic parlance). The numbers of infected people increase exponentially at first, but then growth slows down as the proportion of susceptible individuals decreases and such individuals become increasingly difficult for the agent to contact. The epidemic finally wanes, leaving a small proportion of the population uncontacted. If the population is not of a critical size, the disease then becomes extinct. If a population is large enough to produce continually enough new susceptibles (usually children), the disease may become endemic. Eventually it may circulate like such "childhood diseases" as mumps and measles.

Whole populations do not, of course, occupy a point in space. Diseases diffuse over space because population in all its diversity is distributed over space. All such parameters as vaccination coverage, age, income, education, ethnicity, density, and settlement size vary over space. Day care centers, schools, and medical facilities are located in space. Contact is made in space and must overcome distance to occur. The remainder of this chapter is about geographic modeling of disease diffusion—surely the purest application of the spatial-analytical tradition of geography discussed in Chapter 1.

Types of Diffusion

The geographic process of diffusion goes on constantly and affects many aspects of our lives. Information about someone's death or illness is usually spread from mouth to mouth, by friend calling friend or neighbor telling neighbor. If a new product becomes available (e.g., a new contraceptive or medical treatment), at first the supply will be limited, and people will need to travel to a large urban area or specialized medical center because the innovation will not be available locally. Perhaps a foreign student has finished medical study in a U.S. university and is return-

ing home to practice and teach the new techniques he or she has learned. These examples are medically related, but they could as well be about political rumor, a new sports franchise, or the ongoing restoration of the American bald eagle to regions in which it was recently extinct. They are examples, respectively, of *expansion*, *hierarchical*, and *relocation diffusion*.

Relocation diffusion involves the introduction of an innovation to a location that is not part of the network or system of interactions from which it comes. It often involves leaps over great distances and intervening populations. People with knowledge may migrate to a new place, for example, as when Italian immigrants taught New York Irish Americans to make pizza. We have described several examples of relocation diffusion in earlier chapters. The movement of plague bacilli from burrowing rodents in central Asia to rodents in European cities to burrowing rodents in American, Argentinian, and South African grasslands constituted a whole series of relocation diffusions. When hunters who had depleted the local raccoon population captured raccoons from a Florida population in which rabies circulated and introduced them into the woods of West Virginia, they set off an epizootic of raccoon rabies that has diffused across a region currently extending from Massachusetts to North Carolina, where raccoon rabies is now enzootic. Relocation diffusions happen. Modeling and prediction of diffusion efforts such as those discussed in this chapter usually begin with their arrival.

The terms *expansion*, *contact*, and *contagious diffusion* are often lumped together or used interchangeably. Adoption of innovations, which takes place through contact between people, is strongly affected by distance. Geographers speak of the *friction of distance* and measure *distance decay* (see Chapter 11, Figures 11.1 and 11.2, for illustrations of distance decay). These terms simply mean that the farther apart people are, the less likely they are to have contact or interact. The strong effect of distance causes the process of contact and expansion diffusion to have a distinctive pattern, which is sometimes likened to an ink spot spreading on paper or to ripples moving out from a stone thrown into water. The major difference among these types of spatial diffusion is that ideas can travel over telephones, electric lines, and so forth with almost no friction of distance, whereas a contagious microbe must be transmitted from one individual to another within some physical proximity. Ideas are frequently disseminated to millions of people at a time via television or the Internet. The movement of real people is required for contagion. (The transmission of germs through fomites, such as infected blankets, is extremely rare.) Although most geographic literature is concerned with expansion diffusion, it is not the only kind of contact diffusion.

The Urban Hierarchy of Central Places

The final type of diffusion, *hierarchical*, is characterized by disease agents' moving by contact among people through urban areas according to the areas' relative size. Some of the most important theoretical constructions in geography have for decades addressed the regular spatial arrangement of settlements (villages, towns, cities) of different sizes, and have described how population size relates to economic func-

tions and the distance decay (and hence market regionalization) of such activities as retailing. The settlements of different size and function are referred to as *central places*. These exist in a framework of hierarchical relationships among which goods, ideas, money, and especially people move up and down.

Some goods and services, like getting gas, buying bread, or seeing a dentist, are needed by everyone frequently. These are known as *low-order* goods and services and need to be located as close to people as possible, or people will go to an intervening opportunity. Low-order goods are widely distributed. But every location with a gasoline pump can't support a skilled mechanic or a supermarket. People have to travel to a central place, a town. Every central place with a shopping center and a garage with a skilled mechanic can't support a piano store. A small proportion of the population buys pianos, and usually only once or twice in a lifetime. The *threshold of population* for specialized goods and services, which are *high-order* goods, is high. For example, the number of people it takes to keep a laser eye surgery facility is large, and so they must be drawn from a wider market area. Small central places, through the people living there, provide low-order goods and services frequently for local people. The people of many small towns travel to a larger nearby city to get more variety, greater choice, and specialized goods and services. These cities in turn look for certain more specialized services, such as financing or information media, to higher-order central places. In health services, for instance, general practitioners and pediatricians should be widely dispersed, and every county hospital should be able to tend broken legs, heart attacks, and appendices gone bad. But to get a skilled brain surgeon or radiation and chemotherapy for cancer, one has to go to a higher order center. A burn ward dealing with a very specialized service of high training and cost (one that is fortunately needed by few people very frequently) is a very high order service that usually serves a large area. Anyone severely burned from South Carolina to southern Virginia, from a large city like Charlotte to a coastal military base, will be stabilized and transported to the burn facility at North Carolina Memorial Hospital in Chapel Hill. The area necessary to provide the population needed to support the highest-order goods and services is very large, sometimes regional.

Gerald Rushton at the University of Iowa got the idea of modeling the "market" area over which people traveled to their health services like a watershed for rainfall—a *market-shed*. People do not necessarily go to the dentist or general practitioner in a small town in their own county, but instead go to the higher-order central place, which offers more variety and skill (e.g., an orthodontist and a pediatrician). When a disease agent diffuses, it moves through such interaction of people up the urban hierarchy to the largest city, and then is said to *cascade down the hierarchy* to other cities. In Figure 10.1, where was the agent introduced?

Networks and Barriers

Diffusion occurs over a web, or *network*, of places, people, and the communication links through which information, people, and goods flow. The diffusion of a given phenomenon may be studied at several different time periods in order to

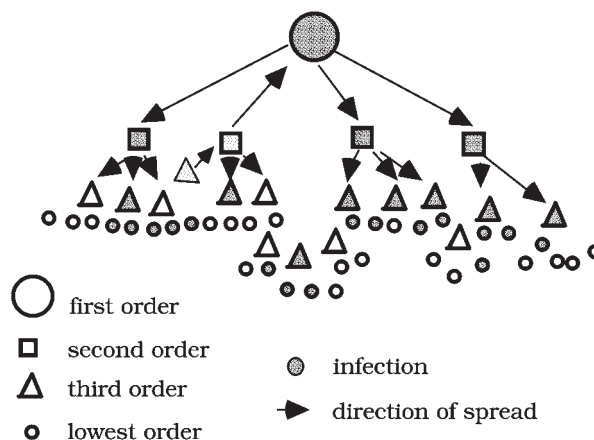


FIGURE 10.1. Infectious disease cascading down the urban hierarchy.

gain knowledge about changes in network structure. Alternatively, research on the simultaneous diffusion of the same or similar phenomena in different networks can illuminate the impact of types of network structures. These strategies are pursued by several of the studies discussed later (see “Modeling Disease Diffusion”).

In opposition to networks, which pattern and support diffusion, *barriers* slow and shape the process. Barriers have three basic effects. *Absorbing barriers* simply stop an innovation. Populations vaccinated against smallpox were an example of an absorbing barrier. Unable to contact susceptible persons, isolated cases had no chance to spread, and smallpox disappeared. Sparse populations in arid lands can behave like absorbing barriers to disease diffusion when the low population density does not facilitate contact transmission. *Reflecting barriers* channel and intensify the local impact of a diffusion process while blocking its spread to another locale. For example, the presence of lakes or unbridged rivers may cause rabid raccoons to turn back and limit an epizootic to a territory, or the presence of a hostile ethnic group or gang may cause people to shop at a certain place or travel in one direction instead of another. Finally, the barrier may be *permeable*, allowing some diffusion but slowing the process. Most international borders are permeable barriers.

Barriers are often physical—oceans, deserts, mountains, rivers, and so forth. Broad barriers, such as oceans or deserts, were effective barriers to contact diffusion during the period of slow transportation. The disease might run its life course during the journey and never reach the destination. Aboard a ship, the susceptible population might be too small to maintain the contagion through the course of a trip. Rapid transport today has almost eliminated the effectiveness of such barriers.

Cultural barriers are especially important in the diffusion of styles of health care. The Chinese carried their medical system with them as they migrated to Southeast Asia in the early 20th century (relocation diffusion). Comparatively few people indigenous to that area have chosen to adopt the Chinese system, however, perhaps indicating the cultural separation of the population groups. It has spread

more, for example, among the Thai, who are Theravada Buddhists, than among Malays, who are Muslims. Programs such as child vaccination, family planning efforts, and AIDS prevention programs diffuse in a similar way, with varying types of constraints.

DISEASE DIFFUSION

The characterization of infectious disease as either endemic or epidemic is often inadequate because a disease can be both. A disease heretofore nonexistent in an area may suddenly appear, often virulently and in the absence of cultural buffers. More commonly, an endemic disease develops in epidemic proportions, perhaps seasonally. How or even whether it develops depends on several well-known factors. The first of these are the qualities of the microorganism that influence its relationship with a human host. These qualities include the number of organisms needed to initiate symptoms and produce clinical disease; for most agents, subclinical infections also produce immunity. Second are the number of organisms shed by a carrier of the infection, the period of shedding, the method of shedding into the environment, and the survival time of infectious organisms in the particular environment outside a human host. Third is the probability of contact with a susceptible host during the period of infectiousness—a factor influenced by the agent's virulence, as well as by the numbers and distribution of susceptible people. A fourth factor is the population's resistance, whether from prior infection, vaccination, or genetic predisposition.

The diffusion of infectious disease moves both through population and over space. Most studies and efforts to model and predict the course of diffusion start with a time series. An epidemic starts when the incidence of cases exceeds expectation, whether that expectation is no disease at all or an endemic clutter of occasional cases. Many diseases follow a cyclical pattern in which a virus or other agent contacts a cohort of susceptible children, born since the last epidemic, and spreads rapidly within the group. After a while, as the proportion of immune members of the group increases, it becomes more difficult for a disease agent to contact a susceptible person, and the rate of spread slows down (the logistic curve again). The periodicity of measles epidemics, for example, especially in less populated areas, was regulated by the fairly consistent number of children born each year and the substantial minimum susceptible population needed to maintain the infection (Cliff & Haggett, 1983; Cliff, Haggett, & Smallman-Raynor, 1993). Artificial immunization has greatly reduced the number of measles cases today, but any relaxation in the completeness of vaccination could reestablish a susceptible population. Such was the case in the early 1980s when numbers of college students began coming down with measles. The relatively new vaccine they had received as small children was found to confer less than permanent immunity. The college epidemics were stopped by requiring all students to get booster shots immediately.

The regularity that results according to these known factors has led to major developments in deterministic modeling. These follow the general approach begun

by Bailey (1957, 1975): susceptibles $\rightarrow$ case incidence $\rightarrow$ infectives $\rightarrow$ removals $\rightarrow$ immunes. Because these models generate epidemics that become extinct, the influential Hamer–Soper model added birth rates at the beginning to generate a continuously susceptible population. The size of population and its subpopulations can then be fed into the differential equations to generate predictions that are mathematically satisfying. The development, mathematics, and applications of these models are thoroughly explained by Thomas (1992). The geographer's frustration at the spatial "blindness" of these models and modelers is well expressed by Gould (1993), especially in his chapters "Time But No Space: The Failure of a Paradigm" and "The Response: How Many Bureaucrats Can Dance on the Head of a Pin?"

The remainder of this chapter discusses ways the diffusion of contagious disease over space and through time can be visualized and modeled. For the most thorough and graphic explanation and presentation of the subject, readers should consult Cliff and Haggett (1988) or, more mathematically, Thomas (1992). Geographic disease diffusion research is often based on models developed to study other types of innovations and applied outside the area of health concerns. These general models have been extended and expanded to fit the special circumstances of disease diffusion. The studies by Brownlea (1972) on hepatitis; Haggett and others on measles (Cliff et al., 1993; Cliff & Haggett, 1982, 1983, 1988; Cliff, Haggett, & Ord, 1986; Cliff, Haggett, Ord, & Versey, 1981; Haggett, 1976); Stock (1976) on cholera; Pyle on cholera (1969) and influenza (1984, 1986); and Gould (1993) on AIDS are used in the following section to look at the diffusion of disease as a geographic subject. The historic application of these methods of analysis, and their limitations due to missing ecology, are addressed with poliomyelitis (Trevelyan, Smallman-Raynor, & Cliff, 2005a, 2005b). The new method being pioneered by Ling Bian of using individual-based models developed out of Hagerstrand's lifelines and stochastic processes is also briefly discussed (Bian, 2003, 2007; Bian & Liebner, 2005). A subsequent section presents the diffusion modeling with some ecology for various influenzas. A final section brings geographic studies discourse to the pandemic of AIDS.

MODELING DISEASE DIFFUSION

Most of the epidemiological research in disease diffusion has involved micro-scale field investigations, which rely on relatively small numbers. The range of scales used by geographers in diffusion studies offers a different view. A long-range advantage to broad-scale models of disease diffusion is their possible importance for forecasting, even for a specific place. If we can identify those variables (centers, barriers, population distributions, spatial processes) that consistently influence the diffusion of a particular disease, or if the geographic diffusion pattern of a place is repetitive and thus predictable, health officials could focus vaccination programs and health education on blocking or minimizing outbreaks. In this section we look at models of hierarchy, distance decay, stochastic diffusion, gravity models of population potential and spatial transformation, and spatial autocorrelation. We also examine how these approaches have been employed to better understand measles and poliomy-

elitis—and, in subsequent sections, influenzas, and AIDS. We end the discussion of AIDS with a consideration of context and consequences, and the limitations of this spatial analysis.

Diffusion over Distance and through Hierarchy

The two major diffusion processes—contact diffusion as expansion diffusion, which involves traversing distance over time, and hierarchical diffusion, which involves the channeling of contact through the population interactions that create the urban hierarchy—have been discussed previously. Figure 10.2 illustrates how they are expressed. Try to identify four types of diffusion in this figure before reading further.

Contact diffusion is best expressed as taking place when the disease agent arrives later at places farther away from the source. Traversing distance dominates the movement of the disease along the coast and up the lower river in Figure 10.2. The movement along the upper river becomes more complex at the appearance of hierarchical diffusion. The infection moves up the urban hierarchy from the second-order coastal city to the first-order city quickly. All the large cities of the second functional order get the infection immediately after the coastal primate (first-order) city does. As the disease cascades down the hierarchy, the size of the urban place becomes more important than the distance to be traversed. Of course, contact diffusion begins to occur locally from each of the infected urban places, too. Radial

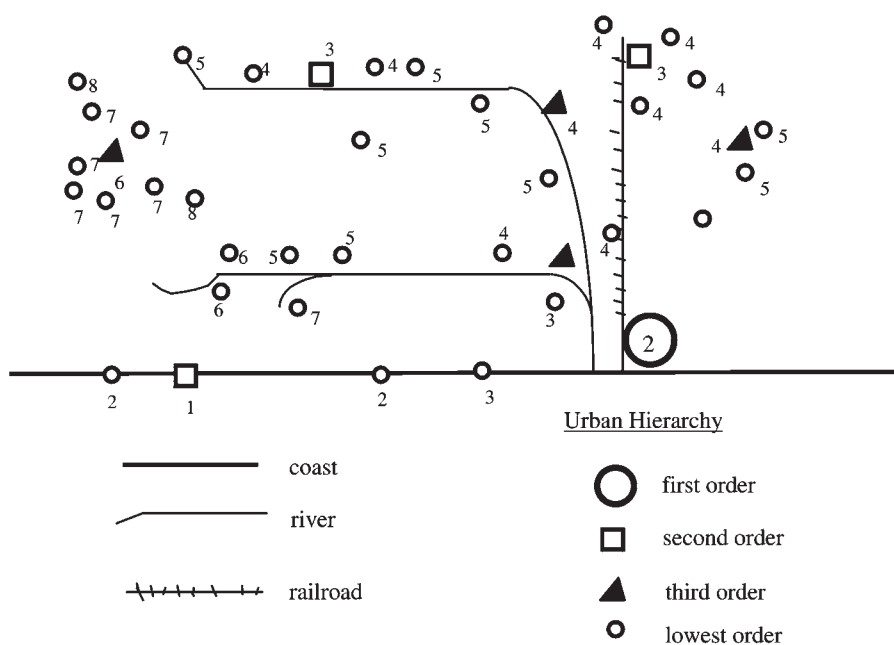


FIGURE 10.2. Distance decay and urban hierarchy. Numbers indicate time periods of a disease's arrival.

diffusion from a settlement to the nearest settlements is obvious, especially among the cluster of villages in the west. They get the disease so much later than the other places, despite the third-order urban place and despite the proximity to the initial settlement infected, that a strong barrier effect is clearly present. It is easy to visualize rugged interfluvial mountains and perhaps a hill tribe ethnic group. Relocation diffusion, of course, occurs when a ship arrives at the coastal port and introduces the first infection to the coastal settlement system. This composite diagram of diffusion types was inspired by the patterns of cholera diffusion in West Africa that were actually mapped, diagrammed, and analyzed by Stock (1976), whose work is discussed soon. The relationships of distance and hierarchy can be represented on a graph.

Figure 10.3 portrays distance from the coastal primate city on one axis and the period of the disease's arrival on the other. The symbols used to represent places indicate their status in the urban hierarchy. It is clear that hierarchical diffusion has become most important as size of place dominates distance in time of disease arrival. The minor (mathematical) vector, however, suggests the presence of an unintegrated system within the country—perhaps similar to the lower river valley of Figure 10.2—in which distance is still the dominant factor for diffusion. This perspective on disease diffusion was first developed through the studies of cholera diffusion by Pyle (1969), whose map of cholera spreading in 1866 appears in Figure 10.3 also. These two conceptually important studies, by Stock and Pyle, are described following a short history.

The cultural ecology of cholera is described in Chapter 2 and spatially analyzed for Matlab, Bangladesh, in Chapter 12 (see Figures 12.3 through 12.7). It is an acute intestinal bacterial disease that is internationally reportable to the WHO. Its symptoms of severe diarrhea and vomiting result in mortality from severe dehydration and the consequent shock and cardiovascular collapse. Historically, mortality (for untreated cases) can be more than 50%. The bacterial agent, *Vibrio cholerae*, has two accepted *biotypes* and may be developing a third. The cholera vibrio, biotype *cholerae*, has an ancient endemic focus in the Ganges Valley of India. Indeed, the dreadful black Hindu goddess of death and disease, Kali, has her own temples in Calcutta (Kolkata). For millennia cholera has broken out epidemically in India when pilgrims returned to their homes from periodic Hindu festivals in the sacred Ganges. By the early 19th century international trade networks and population mobility had developed to spread cholera to southeast and southwest Asia, where the Hajj (pilgrimage) to Mecca became a center of diffusion. Six 19th-century pandemics of cholera spread from the Ganges to the great ports of India, and then through Suez and on to Europe and North America. Each pandemic was cleared elsewhere and collapsed back to only the Ganges (like a primary natural nidus).

Cholera epidemics in the early 20th century seemed effectively limited to Asia. Then a World War II-era epidemic introduced cholera to the eastern islands of the Indonesian archipelago, where it stayed and evolved into the *El Tor* biotype. *El Tor* cholera has a lower case mortality (i.e., it is evolving toward commensalism). Infections may involve only diarrhea, or even no symptoms at all. Infective but unknowing carriers of the disease resulted in the WHO's dropping the last pretense of con-

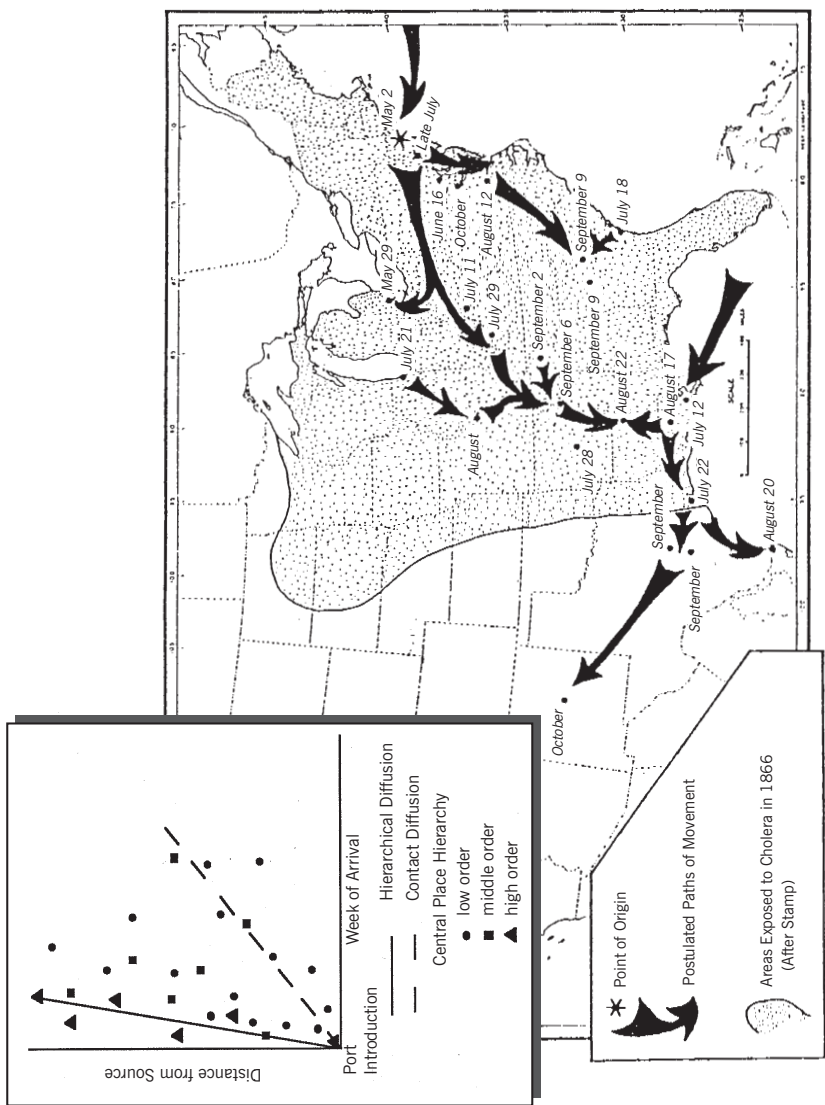


FIGURE 10.3. Diffusion of the 1866 cholera epidemic in the United States. From Pyle (1969, p. 66). Copyright 1969 by Ohio State University Press. Reprinted by permission.

trol through vaccination, which is no longer required for travel (and was never more than partially effective in the short term). The new cholera spread after 1961 to the southern states of Europe and the (former) Soviet Union, entered the United States and Japan (where sanitation contained it), and apparently for the first time spread to West Africa (where it has stayed). In the 1990s the agent was brought in ballast water by ships from Southeast Asia to Peru's coastal algae, and it then diffused through South America. The diffusion of El Tor to a susceptible world population set the stage for studies of disease diffusion.

Pyle was the first to realize that the diffusion of a disease such as cholera could be used to analyze the spatial structure of a country. That is, the pattern of the diffusion could illustrate the type of transportation system, the development of national economic integration, and the evolution of a functioning urban hierarchy of commercial and mobility interactions. Even before the development of germ theory and vital statistics, the arrival and presence of cholera in a place were noted in the local newspapers. Pyle studied the diffusion of three major cholera epidemics in the United States by developing the methodology depicted in Figure 10.3. In the first epidemic, in 1832, cholera was introduced (relocated) at several points: the riverine ports of Canada, as well as New York City and New Orleans. Diffusion occurred along the Atlantic coast and the riverine systems. Clearly the farther along the coast or a river a place was, the later the arrival of cholera, in a system dominated by water transport. The second epidemic, in 1849, occurred when the hierarchical structure of the urban system was emerging. Railroads were becoming important, although much of the interior was not yet connected to the East by rail. The epidemic was introduced from Europe in both New York City and New Orleans. From these cities it diffused up the rivers and along the coast in two separated regions; within each region, however, hierarchy (i.e., size of urban place) had emerged as the major predictor of how soon the epidemic would arrive. By the time of the third epidemic, in 1866, the rail system was well established but did not cross the continent. Cholera was again introduced in New York City, but the epidemic now moved through the country as one system of urban hierarchy determining time of infection. Even the exceptions (residuals) to this served to emphasize the structure: Detroit was infected early for a city of its size, as explained by its proximity to nearby Canadian cities; San Francisco was infected more than a year late for a city of its size, but of course the transcontinental railroad was not yet built and pioneers still walked or sailed there; and Washington, D.C., which then was almost deserted during summer, did not receive the epidemic until Congress convened in the fall.

The arrival of El Tor cholera in West Africa (1970–1975) in a “virgin” or totally susceptible population provided Stock the chance to develop one of the most comprehensive and conceptually clear studies of diffusion. He examined the applicability of existing diffusion models. First he needed extensive fieldwork to identify, locate, and map the local timing of the outbreak of cholera throughout the region. He identified behavior at the village level that buffered or increased risk. This ecology was modeled as a local system. It involved such things as caring for the sick, preparing the dead for burial, washing hands, eating with hands out of common plates, use of latrines, and presence of local medical service. Then, addressing regional

scales, he identified and developed four models (which are generically represented in Figure 10.2) to describe the overall diffusion pattern.

There were three models of contact diffusion: *coastal*, *riverine*, and *radial*. Fishermen carried the disease along the coast, and it spread from coastal villages to nearby towns. Diffusion along the Niger River was also dominated by the friction of distance. Larger towns received the infection no earlier than smaller towns closer to the origin. The radial pattern of contagious diffusion was identified around Lake Chad. Here the disease intensity was particularly virulent, and as people fled the lake area, there were no developed transportation routes or natural channels to structure the diffusion paths. People from a village fled to the nearest ones around, from which people fled outward to the next ones. Only in Nigeria did a clear example of hierarchical diffusion emerge. Coastal diffusion introduced cholera to the primate city of Lagos, from which river, bus, and rail routes emanated. A settlement's level in the urban hierarchy, rather than its distance from Lagos, determined when the epidemic arrived. Overland transportation routes channeled diffusion. The least affected areas of the country had low population density, little urbanization, and undeveloped transportation.

Stock's (1976) monograph emphasizes the complexity of the geographic structure of diffusion. All the basic models had some applicability. Immune populations, deserts, and sparsely populated areas were permeable barriers, but rivers, coast, and developed overland routes helped channel movement. Stock suggested that the linear pattern along the Niger River loosely resembled the 1832 cholera diffusion along the internal waterways of the United States discussed earlier, whereas the hierarchical diffusion in Nigeria more closely resembled the 1866 U.S. epidemic's structure. The difference between Nigeria and Mali in economic development in the early 1970s thus seems similar in spatial expression to the past structural development of the United States.

Local Diffusion by Chance: Stochastic Simulation

There have been few attempts to apply to disease diffusion the simulation methods formulated by Hagerstrand, which have otherwise been very useful in modeling innovation diffusion (see Vignette 10.1). This may be because the influences on the multiple waves of an epidemic are very complex. The aging but still the best example of the application of stochastic diffusion simulation to disease is a study of infectious hepatitis in Wollongong, Australia (Brownlea, 1972).

The infectious hepatitis virus is transmitted through fecal contamination. The virus is robust and tolerant of a wide range of environmental conditions. Water, fish, seafood such as oysters, and pet hair may be vehicles for disease transmission. Brownlea identified cyclical fluctuations in the number of cases reported. During epidemic years, incidence rates of 2 per 1,000 or more affected mainly children and showed an equal sex ratio, strong spring seasonality, and no spatial concentration. During interepidemic years, rates were rather uniformly 0.3 per 1,000, affected all ages, and showed little seasonality and no spatial concentration. Because hepatitis is often poorly reported, Brownlea aggregated cases over 15 years and used a Poisson

probability test to determine just where active spread was present. He identified the center of this as a wave that moved spatially through time, and labeled it a *clinical front*.

Brownlea developed several models for hepatitis behavior at different scales within this industrial–suburban, rapidly growing region. The basic model was based on Hagerstrandian-type stochastic simulation of a random walk. Assuming a closed population, an equal chance of diffusion in all directions, and essential community immunity as the epidemic passed, he simulated a random diffusion in which the clinical front would advance as a ring from Wollongong's initial node. Time periods for measurement were determined by incubation period. The bulges and bends in the actual, observed advance of the clinical front—that is, the deviation from the random—he attributed to ecological parameters he had identified, which operated as barriers or channels (Figure 10.4). Modifying the simulation to fit the actual disease behavior by adjusting the cell probabilities showed that concentration of young families, light sandy soils, concentration of older people, and a polluted lake used for fishing and swimming were important parts of the physical and sociodemographic surfaces that retarded or channeled epidemic diffusion. Recalibration of cell probabilities led to a close match between the simulated and real diffusion patterns.

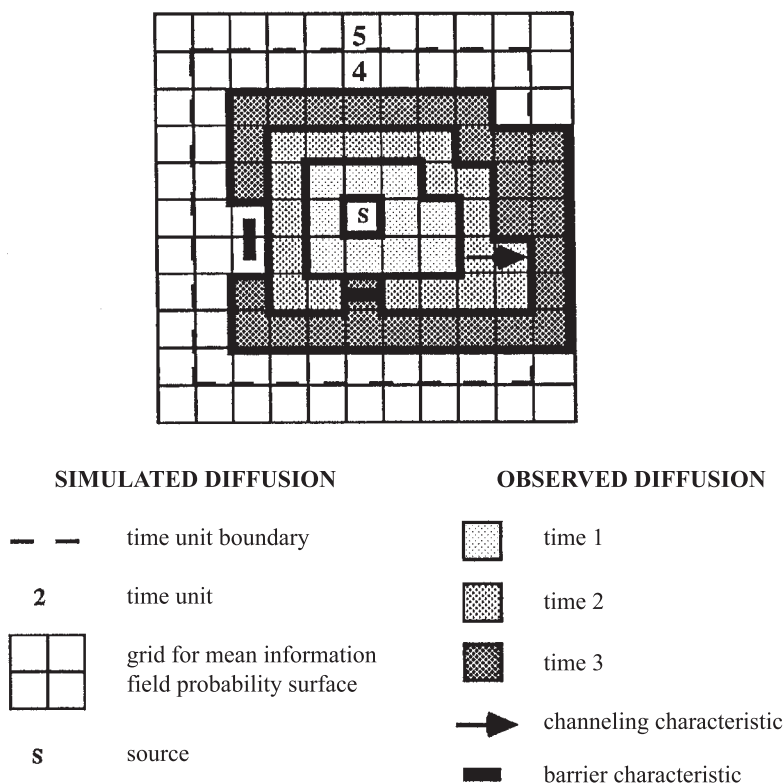


FIGURE 10.4. Simulation of the diffusion of infectious hepatitis.

These findings were applied to the movement of the clinical front among the settlement nodes in the study area. Brownlea found that the disease diffused from each population center, exhausting the susceptible population at the core and rippling out to the surrounding nodes. If a peripheral node developed a sufficient concentration of immigrants and young children, it became a new center of infection and node of diffusion. The first, exhausted node was occasionally reinfected from the surrounding, newly active nodes, which caused the in-filling epidemiological pattern of the interepidemic years. The model laid a basis for explaining the disease's rhythmicity and spatial patterning, and it provides a powerful example of using Hagerstrand's stochastic simulation for analysis of disease diffusion.

Approaches from the Family of Spatial Interaction Models

One of the most robust nomothetic models in geography is the *gravity model*, the prototype for geographic modeling of spatial interaction. It has been widely used to account for, and predict, an amazingly wide variety of flows: job migration, telephone traffic, airline passenger movements, mail delivery, commodity shipping, and the spread of information, among many others. Because it is about interaction, contact, flow, connection, and the spatial patterns that result, it is also about the diffusion of disease.

The basic concept is elegantly simple. The greater the number of people who live in two places, the greater the likelihood of contact between them; the further apart the two places are, the less the likelihood of contact between them. We would expect more people to move, write, or telephone each other between Chicago and Philadelphia than between Chicago and further Houston. We would expect more of such contact between New York City and Washington, D.C., than between New York City and smaller Buffalo, New York. In an important way, even the hierarchical diffusion discussed earlier is "gravity model thinking," as the likelihood of interactions between the populations of the largest cities overwhelms the friction of the distance between them. If, furthermore, instead of the attraction between only pairs of places, we consider the influence of all places in a study region upon each other's contacts, we create a surface of potential interaction among people called *population potential*. (Vignette 11.1 develops the calculation of population potential.)

The gravity model was derived from efforts in the late 19th century to analyze society "scientifically" and create a "social physics." The inverse relationship between the attraction of population size (i.e., mass) and the discouragement of distance seemed obviously analogous to Newton's gravity model. Some of the terminology used today, such as *potential*, is derived from Newtonian physics. Obviously, however, the surface of the earth and the spatial arrangement of its settlements do not exist in a void, or on a perfect plane. Such things as direction, road routes, and mountain or sea barriers really matter. Different "frictions" of distance encountered by different modes of movement (whether by foot, car, plane, etc.) could be empirically adjusted in the model by varying the exponent on the distance (instead of dividing by distance squared, à la Newton) until the formula worked.

After this initial model “calibration,” more elaborate ways were developed, using constants to “constrain” the model so that the total flow predicted from an origin to a destination equaled the actual total flow observed. The computation of gravity model predictions of migration, for example, has become very complicated. The large literature on the gravity model and other models of spatial interaction can be conceptually approached for a general explanation (Abler, Adams, & Gould, 1971) or more advanced and mathematically sophisticated explanations (Cliff & Haggett, 1988, 1989; Thomas, 1992).

The family of spatial interaction models and their derivatives can provide a way to put real space into diffusion within a population. For example, the former Soviet Union used to collect remarkable incidence data (daily reports from more than 100 cities) and had developed detailed mathematical models of the diffusion of influenza within its population. These formulas ignored distance between the cities in that vast country, however. The models were adapted by British epidemiologists and others to the kinds of data available in the West, but they remained aspatial. When Pyle (1986) used the gravity model and U.S. data to put distance (between cities) into their predictions, he showed that distance as well as size (mass) was important by accounting for 70% to more than 80% of the variance among U.S. cities. He further added to his prediction by using population potential of the elderly population in his regression analysis.

The following two examples—of Gould (1993) using the gravity model as a “spatial filter” to transform the diffusion space of Ohio, and of Hunter and Young (1971) using a surface of population to explain the basic movement of an influenza epidemic through England and Wales—should clarify the gravity model and provide some sense of the power of these techniques in medical geography.

Diffusion through Transformed Space

Cartographers have long transformed space for analytical purposes. For example, distances on roads may be represented by travel time instead of actual miles, so that areas get foreshortened or extended. Cartograms, such as demographic base maps, represent the content of a region as space. Indeed, it is impossible to represent the surface of our sphere on a flat sheet of paper without transforming space (Chapter 3). Gould (1993) used the gravity model as a “spatial filter” to transform the space of Ohio so that accessibility, in this case to the contagion of AIDS, could be illustrated and used predictively.

Gould mapped the diffusion of AIDS across Ohio’s 84 counties and three large cities of Cleveland, Columbus, and Cincinnati from 1982 on. As he claimed and as Figure 10.5 shows, there is educational value in the dramatic impact of maps showing AIDS diffusing—and arriving. Instead of dismissing AIDS as spreading among some “other kind of people” or being someone else’s problem, the reaction of Ohioans viewing the maps in sequence (and of course locating their own place first) was often “It’s coming; . . . it’s here!”

The three largest cities have a lot of interaction among their populations, of course, as well as contact with large cities outside the state. The gravity model was

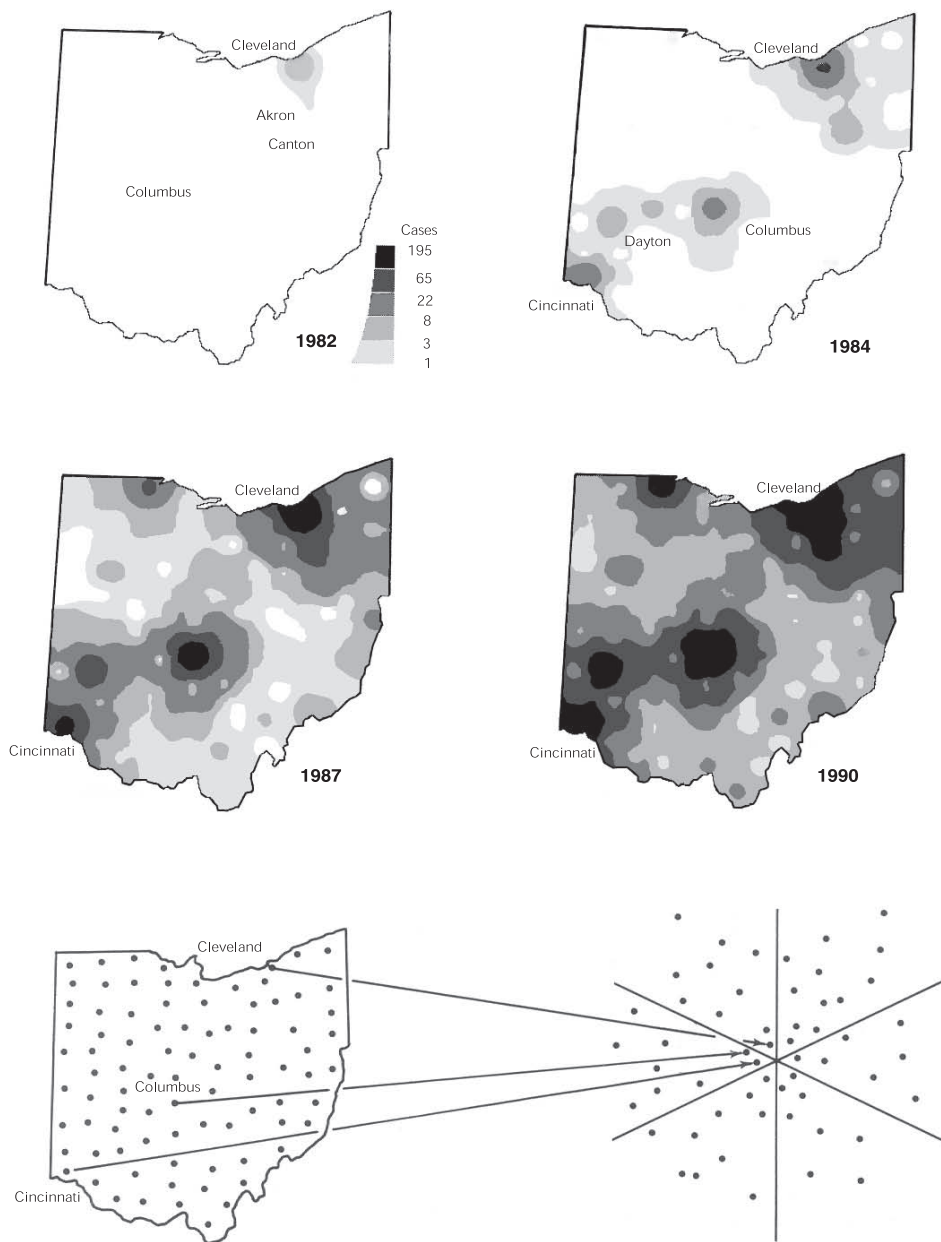


FIGURE 10.5. Diffusion of AIDS in Ohio and in space transformed by the gravity model. The gravity model can be used as a "spatial filter" to transform the space over which infectious agents spread. From Gould (1993). Copyright 1993 by Blackwell Science. Reprinted by permission.

used to calculate all 3,486 potential interactions among the counties, which were then mapped into multidimensional "AIDS space." The rural counties isolated from other contacts, or places connected by proximity and population size, then showed the spatial logic of the mapped pattern of diffusion.

Figure 10.6 depicts another population potential surface for England and Wales. After examining the uses of absenteeism at school and work, and various other estimates of infamously unreported influenza incidence, Hunter and Young (1971) used the incidence of acute pneumonia (primary and influenzal, and well reported) to map the Asian flu epidemic of 1961 week by week on areal and demographic base maps. They located the *centroid* (the mean center, itself a kind of gravity model calculation) of cases for each week. They calculated the population potential by county and mapped it (see Vignette 11.1 for techniques for both of these). As the figure shows, the center of the epidemic advanced week by week along a ridge of population potential until it reached the "potential mountain" of London. Population potential and week of onset were highly correlated, but not total incidence.

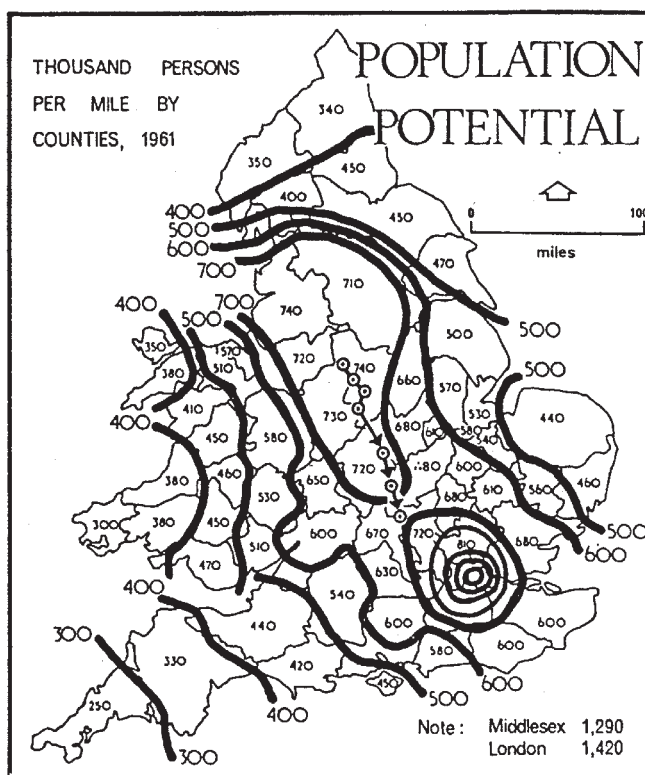


FIGURE 10.6. Population potential and diffusion of influenza in England and Wales in 1957, showing the movement of the epidemic center along the ridge of highest population potential. From Hunter and Young (1971). Copyright 1971 by the Association of American Geographers. Reprinted by permission.

Testing Models of Diffusion with Spatial Autocorrelation

The methodology of analyzing changes in relationship (or similarity) over space by correlating the values of geographic units, such as counties, at sequentially increasing distance for sequentially increasing time can provide a powerful and sensitive analysis of disease diffusion. This method of analysis, *spatial autocorrelation*, is described and developed in Chapter 12 (see especially Vignette 12.3). In this section we look at how, in the hands of masters such as Andrew Cliff and Peter Haggett, it can illuminate the disease process and its human geography. The illustrations and models that follow are all from their work (Cliff & Haggett, 1982, 1983, 1988; Cliff et al., 1981, 1986, 1993; Haggett, 1976).

Measles

Measles is endemic throughout most of the world today. The virus is highly infectious, and nearly all susceptible individuals contract measles after close contact with an infectious person. Epidemics flare when a sufficiently large population of susceptible children are born in a population that is otherwise more than three-fourths immune as a consequence of previous infection (i.e., at the saturated end of the logistic curve). In small island populations the disease agent disappears (ceases to exist) when the virus that has killed or created immunity can no longer find a susceptible person to infect. There is then no measles in that place until population mobility introduces it again from outside. Numerous mathematical models and island studies have determined that a minimum population size of 250,000–500,000 frequently interacting people is needed to generate the new susceptibles and contact rate required to maintain the disease. The difference in these population estimates is undoubtedly related to the relative isolation or intensity of interaction of the people. The need for such a large size of interacting population is strong evidence, as discussed in Chapter 2, that measles could not have existed as a human disease until the first cities were built in the trading nexus of today's Middle East.

Haggett (1976) first studied measles diffusion in the relatively isolated county of Cornwall in southwestern England. He posed seven possible models from what was understood at the time about the process of spread. Which was important: spreading in separate regional systems; spreading from cities to surrounding rural areas; contact diffusion spreading in waves across distance from the largest city; local contagion from one geographic unit to another; contact by people in their "journey to work"; or simply the population size of the places or the population density of the unit areas? He had a two-wave measles epidemic to study—a 222-week series from late 1966 to the end of 1970, across 28 political jurisdictions. The genius of this study was to redefine *nearness* to count *joins* in terms of his postulated models, not just spatial contiguity. Each unit could be correlated with its neighbors, but so could commuting origins and destinations, cities of the same class size, units of the same density, and so on. Haggett discovered that different spatial processes of diffusion were more important during different stages in the epidemic, as well as during endemic or epidemic periods. During endemic periods, the population size model

showed the persistent cases of disease in an otherwise weakly structured low level of contagion. Early in the epidemic, population size became less important, and waves of contagion effects (shortest distance) and journey to work took over. During the epidemic peak, local contagion was most important. As the epidemic waned, it did not shrink in geographic extent but decayed in place, with some tendency for lower-level hierarchical spread to continue. These observations could not be generalized, of course, beyond this one epidemic in this singular place and time. A more ambitious study was needed, on a greater scale.

Cliff and Haggett (1983, 1986, 1988, 1989) found in Iceland a large island population system that could be “closed” for analysis as a region and that had the great advantage of meticulous records kept over more than a century. They followed the introduction, spread, and disappearance of multiple epidemics of both measles and influenza. Over time the population grew large enough to maintain endemic disease, and the infrastructure of transportation, medical care, schooling, and general circulation of information and people was also transformed. They were able to study how intervals, peaking, duration/frequency, age structure, and other dynamics were affected. Spatial autocorrelation methods were at the core of the most powerful of these analyses.

As shown in Figure 10.7, they used the method developed in the Cornwall study. They defined *joins* in various ways to test hypotheses of process. The geographic patterns of diffusion of contagious measles and influenza could be expected to reflect population movements for work, schooling, shopping, holidays, and such social activities as visiting relatives. An *index case*, or infected person, could theoretically travel to any one of the 50 medical districts (used for epidemiological data). The center of each district could be considered a *node*, and nodes were linked to each other by links of minimum distance (represented by the *minimum spanning tree*, or MST). When joins were defined not only in terms of minimum distance, but also in terms of airports, road connections, and nearest larger place with or without contiguity (i.e., urban hierarchy), the progression, duration, and concentration in time of successive epidemics of measles and influenza could be examined over decades of change. On each of the histograms in Figure 10.7, the vertical axis plots the number of occasions in 17 influenza and 8 measles epidemics on which the black-white statistic (Vignette 12.3) showed that network to be statistically significant at the 95% level in a one-tailed test for positive spatial autocorrelation. Time in months is on the horizontal axis. For “nearest larger place,” the bar graph shows nearest larger contiguous place, and the line graph shows hierarchical size regardless of contiguity (i.e., pure population). The MST and the road network are said to be “spatially constrained”; that is, they represent contagious diffusion in which contiguity and distance are important. The air network and nearest larger place represent hierarchy, nodes, and population connections as more important than distance.

The graphs show the importance of contagious spread for both diseases. In the case of influenza (the b graphs), for example, the MST was important from 1 month before to 2–4 months afterward, and for measles (the c graphs) at all stages. The road network was most important during the peak of an epidemic. The influenza epidemics waned in place, not through constriction of territory. So did measles, but

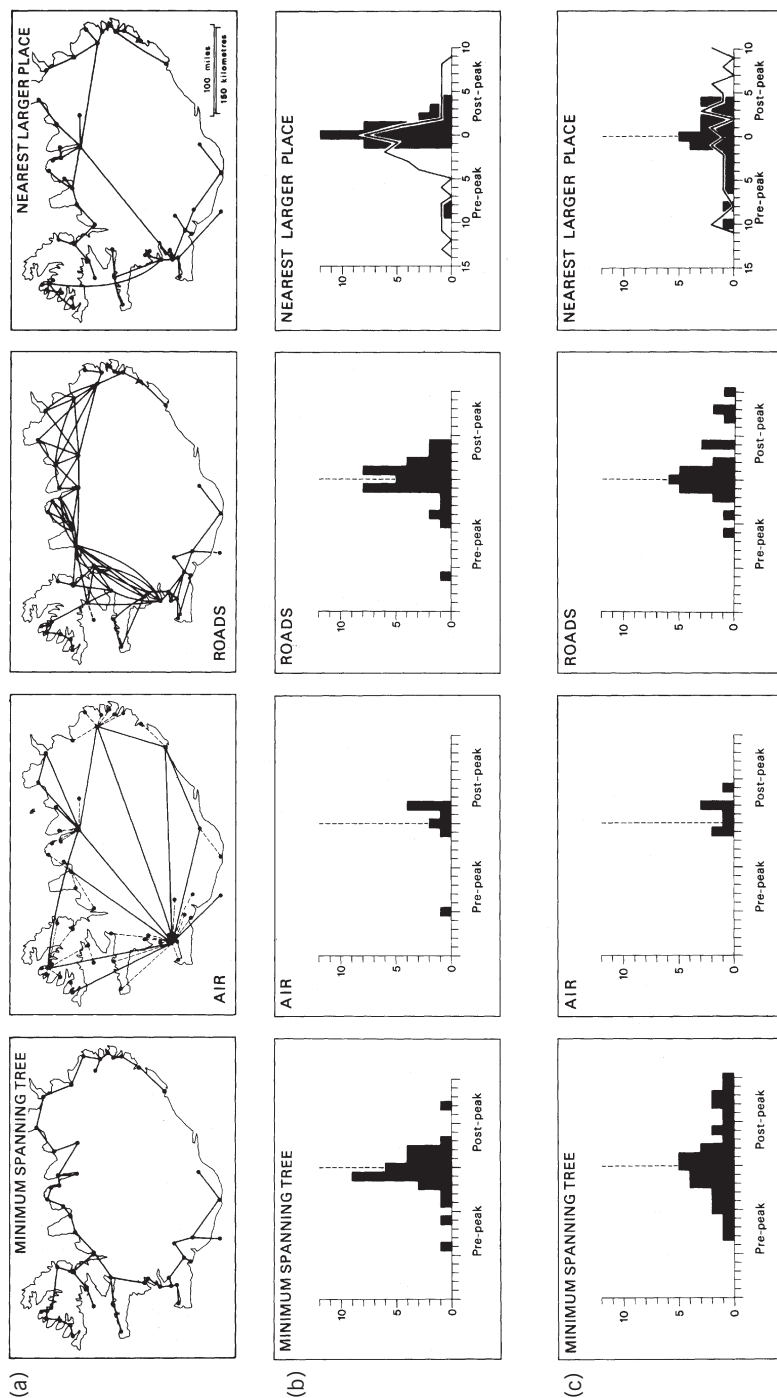


FIGURE 10.7. Join count statistics for estimating paths followed by influenza and measles epidemics in Iceland, 1945–1970. (a) Nodes and joins for diffusion models; (b) influenza; (c) measles. Spatial autocorrelation was used to study the links and pathways of diffusion of influenza and measles in Iceland. Number of significant deviates ($\alpha = .05$ level, one-tailed test). From Cliff and Haggett (1988, p. 206). Copyright 1988 by Blackwell Science. Reprinted by permission.

it showed some role for urban hierarchical diffusion in filling in the latter stage. Contiguous hierarchy was important for influenza only at the peak and immediately after; pure population size hierarchical diffusion (the line) was important in the initial spread, during the peak, and until the end. In contrast, unconstrained hierarchical diffusion was never important for measles. For both diseases, the air network was surprisingly insignificant. The researchers speculated that the period of infectivity is too short for both of these diseases for long-term travel to be as important as local contagion. The difference in the importance of population size is likely related to the fact that measles gives permanent immunity to the exposed population, whereas influenza continuously mutates and finds new susceptibles.

As the transport system, school locations, residence types, and commercial connections in Iceland changed over decades, so did the frequency, peakedness, duration, intensity, and other characteristics of the epidemics. From studies in spatial analysis such as this, a lot can be learned both about the characteristics of an infectious disease and about the connectivity and structure of the human geography. Such knowledge is especially useful for developing strategies for vaccination (such as those used in the eradication of smallpox) and epidemic containment (Cliff et al., 1986).

Poliomyelitis

Geographers have recently looked at the epidemics of poliomyelitis that emerged in the early 20th century—in particular, the great epidemic of poliomyelitis in 1916 in New York City, and its spread (Trevelyan et al., 2005a, 2005b). Polio loomed as a horrific public health crisis until the late 1950s, when effective vaccines led to its retreat in the United States. Poliomyelitis is now *almost* extinct in the world, hanging on in northern Nigeria and war-torn places like the Pakistani frontier and Somalia. After being reduced to 38 cases in Nigeria in May 2008, it flared again to 258 cases in May 2009. It spread back into 10 African countries, and from India to flare again in Indonesia (which mobilized another campaign in Java to vaccinate all children). The WHO has again studied the situation and resolved that it is possible to eradicate this terrible disease. Analysis of how it spreads spatially will be useful in this effort.

Using data from a seminal 1918 public health bulletin to examine the spread of polio in 1916 within New York City to its environs, Trevelyan et al. found a cluster of maximum activity with a 128-kilometer radius, and integrity as a cluster to 500 kilometers, with the level and rate of propagation of the wave (the logistic curve; Vignette 10.1, Figure 2) declining with distance. They determined that it “probably” followed a mixed contagious-hierarchical diffusion in the U.S. Northeast. Using data from the *Mortality and Morbidity Weekly Report* (Centers for Disease Control and Prevention [CDC]), and analyzing spatial autocorrelation and wave periodicity over 50 years, they found that polio activity increased in velocity of epidemic transmission over the decades (i.e., as U.S. transportation became increasingly networked, as in Iceland; see Figure 10.7) and became more spatially coherent. Over this half century the center of viral spatial activity shifted from the Northeast to the West Coast to occurrence across all states. After mass vaccine for polio was developed

and implemented from the mid-1950s on, there was an increase in epidemic volatility and increased dependence upon endemic reservoirs in the most populous states, until in the end not only the incidence but the spatial structure collapsed.

These analyses of poliomyelitis epidemics as they emerged and disappeared in the 20th century, using urban hierarchy, spatial autocorrelation, and infrastructure networking, seem to fit the process and pattern of the influenza and measles models above. There is something lacking in this basic spatial analysis, though, we think. Polio has a reversed disease ecology when compared with other 20th-century infections of childhood. It was a great puzzle. Polio, or infantile paralysis, was more common among the middle class and even the rich (e.g., Franklin D. Roosevelt) than among the poor. It was not a disease of filth and bad water; it was not vectored. In World War II it affected American soldiers in Egypt and/or the Middle East, but it did not seem to affect locals. It increased steadily as the U.S. infrastructure grew better, the population healthier, and the economy wealthier. The new postwar suburbs were its most fertile ground, and it was a terror there. Finally it was found that the polio virus was ubiquitous—a human infection that almost every child had previously contacted in infancy, everywhere, over the centuries. Babies received antibodies from their mothers' milk, which protected them until, constantly exposed, they developed their own antibodies. With improvements in sanitation and water supply, children were protected from the virus; some never contacted it. When children without any maternal antibodies did contact the polio virus, some developed infantile paralysis, or indeed died from the infection. Vaccination was needed as a cultural buffer to replace the mothers' antibodies, which safe water and sanitation (so protective against other enteric diseases) had removed, leaving children exposed. From reading the analyses of diffusion of polio epidemics above, and from thinking about how polio epidemics spread around the world (following the demographic transition) as India, Kenya, and Brazil improved their sanitation and water supply, it seems that what is diffusing is not the virus but protection from the virus—in other words, development, infrastructure, modernization. Leaving the disease ecology out of the spatial analysis would make the patterns of advance recorded above seem mere reflections in a mirror.

Emerging Methods

New computer models and geographic information system (GIS) simulations are opening up new methodological approaches. Old disease diffusion data, such as those on polio described above, can be approached in entirely new ways. One of the diseases being studied is the great 19th-century summer epidemic killer, yellow fever (see Chapter 4). Using data from public health reports cleverly applied to geocoded insurance lithographs of residences, Curtis, Mills, and Blackburn (2007) studied the great yellow fever epidemic of 1878 in New Orleans, officially reported as 27,000 cases with more than 4,000 deaths. They analyzed a new statistic, a *spatial basic reproduction number* (secondary infection), by putting 50-meter buffers keyed to *Aedes aegypti* activity range around each death as a "neighborhood" (see Chapter

9). Thirty-three neighborhood index locations were finally identified, with age and nativity pattern for each, which followed the epidemic. The study highlighted the importance of multideath residences in numbers of deaths and as nodes in patterns of neighborhood diffusion sequences. In innovative hands, GIS applications are beginning to tease new insights out of even our oldest data.

Networks in Time–Space

A new method of analyzing diffusion uses the power of spatial data managing in a GIS to address the mobility in time–space issues raised in Chapter 5. In particular, Hagerstrand's lifelines in space are "back" after decades of neglect and can be analyzed with ArcGIS. Ling Bian (2003, 2007; Bian & Liebner, 2005) has challenged traditional epidemiological models based on uniform populations with uniform distributions interacting globally with equal contact rates. She has constructed a conceptual framework with network models of diffusion based on individuals interacting at multiple scales (see Plate 14 for an example). Time matters for both continuous processes and discrete events. Time periods in these models can be set to latency and incubation periods. Individual life paths intersect in two "layers," of homes and workplaces. These nodes are linked in a network of connections between groups, so that within-group and between-group interactions represent local and long-distance transmissions. This is stochastic modeling of discrete individuals moving in space, dispersing through networks (not waves). Data are gathered on occupations and workplaces; transportation routes/means and commuting; night and day residential patterns; shopping patterns; and so on to place individual people into multiple levels of spatial interactions, each level a medium of contagious diffusion. This form of modeling is only at its beginning, but holds enormous promise.

INFLUENZAS

The "swine flu epidemic" that did not happen in 1976 provides a major example of the need for geographic modeling of diffusion pathways. After the type of influenza isolated in an outbreak at Fort Dix, New Jersey, in January was thought to be the same as that which killed 20–40 million people in a world pandemic in 1918–1920, massive efforts were made to vaccinate the entire U.S. population. A particular characteristic of the deadly pandemic was that it killed not only elderly, very young, and other vulnerable people, but also healthy young adults. The entire population was at risk. There was no way to target vaccinations because little was known (or spatially considered, of course) about how such an epidemic would spread. In the end, there was no outbreak. Not only was a large amount of money spent, but because there are inevitably some people who have reactions to any vaccination, the serious side effects became more important than the danger of the swine flu epidemic that did not develop. The vaccination campaign created a popular backlash that

impugned both science and government. How much better it would have been if the vaccine could have been produced and storehoused while a rational strategy for containment of any outbreak and spatially targeted vaccination coverage was implemented.

Basic Ecology

The ecology of influenza is intrinsically interesting. Influenza is notorious as an annual winter epidemic. The usual epidemiological explanation for winter infections has been that people gather together inside and so communicate contagion better. Even in places that are crowded all the time (e.g., prisons), however, influenza is a winter epidemic. It has been suggested that in lower humidity, or perhaps in lower ultraviolet radiation, the virus may survive longer in the air between people and so be able to infect more of them; still, nothing could be proved. Cliff and Hagggett have addressed how influenza that is newly epidemic in the Northern Hemisphere in winter also becomes epidemic in the Southern Hemisphere in winter, but seems to occur at low rates in the tropics all year long (Cliff et al., 1986). Now that it has been found that the guinea pig gets the flu, and so can be used as the animal model that was previously lacking, the mystery has been solved (Lowen, Mubareka, Steel, & Palese, 2007; Lowen, Steel, Mubareka, & Palese, 2008). Measurements have shown the virus to survive longest in cold temperatures and to be most contagious in low humidity because the virus adheres to heavier water molecules and falls to the ground. The chance for transmission in hot, humid air is very short.

Influenza is one of those few contagious diseases that could have occurred even when human population on earth lacked great cities. There are two reasons for this: Humans share the influenza virus (or viruses) with several animals, especially swine and ducks, and the virus has an extraordinary ability to mutate in gradual (drift) and radical (shift) transformations that baffle the recognition and memory of the human immune system. Three major classes of influenza are known. Influenza C is least known, but does not seem to cause epidemic disease. Influenza B causes morbidity in youthful populations and has low case mortality. Human epidemics are caused by influenza A strains of virus. Influenza A strains have a coat, or envelope, of glycoprotein in the form of spikes of hemagglutinin (H) and neuraminidase (N), and are referenced by ratios and patterns of these spikes. The H is responsible for binding the virus to host cells. The N facilitates release of the virus from infected cells and the spread to others. As "the flu" spreads endemically in a population, continuing all year but spreading especially in winter, the virus changes the angle, bend, and other properties of its spikes. This gradual "drift" finds enough new susceptibles for the contagion to continue its presence. Then, with irregular periodicity, the virus rearranges its spikes into different HN ratios and patterns that are not recognized by the immune system ("shift"), and the whole population becomes susceptible to a new infection which can spread in a great pandemic wave. There were three pandemics in the 20th century. The so-called "Spanish flu" (which did not come from Spain) killed 20–40 million people in 18 months from 1918 to 1920, as

noted above. It was caused by an H5 virus, which usually does not occur in people but in birds. The Asian flu of 1957–1958, H2N2, is considered mild, as it killed “only” a million people while causing massive absenteeism and misery. The Hong Kong flu of 1968–1969, H3N2, was another mild pandemic thought to have killed “only” a million people.

This rearrangement of H and N surface proteins seems to occur through genetic exchange as multiple varieties of the influenza virus infect pigs and perhaps ducks simultaneously. They emerge where the “three P’s”—people, pigs, and poultry—reside together. Cultural processes of herding pigs and ducks and agricultural domestication and intensity of rearing have played a role in creating pandemics, but the recombination process probably also occurred occasionally in wild birds and mammals in prehistoric time. There has been a tendency for new varieties to emerge from the cultural hearth that domesticated pigs and ducks, southern China, which continues an especially numerous and intense association among the three P’s. This has given Hong Kong special importance as an advance warning surveillance site for the development of appropriate vaccines for the future winter. It should be noted, however, that the U.S. Midwest also has an exceptionally intense historic focus of swine and people, and that the great North American flyways (of migrating waterfowl) may be implicated in the early emergence of epidemics. Pyle (1986) noted the xenophobia implicit in naming influenza epidemics for other places. Indeed, the “Spanish flu” (or deadly “swine flu,” in today’s vernacular) of World War I was quite possibly brought by American farm boys from the Midwest to Europe.

Simulating U.S. 20th-Century Epidemics

Pyle (1986) tried to discover whether these drifted (endemically epidemic) and shifted (rapidly pandemic) forms of influenza have different and predictable spatial patterns. Such footprints would be useful in detection, but especially important for developing containment and intervention strategies. Using harmonic analysis (see Vignette 6.2) to map the timing of epidemic arrival and peak, Pyle studied centuries of influenza epidemics in Europe and decades in the United States. He reconstructed diffusion pathways to identify patterns of great “frontal” assaults sweeping across continents; patterns of year-to-year continuous advance; and patterns of multiple-node, gradual fill-in. Having seen these footprints of shift and drift, he then combined many of the modeling techniques discussed in this chapter to simulate how the great epidemics of swine flu, Asian flu, and Hong Kong flu spread in the United States, compared with the endemic spread of intervening years. Pyle simulated how influenza would spread randomly if it began in the sites harmonic analysis had designated as “repeated events,” and how it would spread if hierarchical diffusion dominated. He added map surfaces based on population potential for the elderly instead of city locations. He then used regression to test the prediction of these simulations for the spatial spread of drifting and shifting influenza. Clever though these models were, the results are somewhat muddled by the quality of data: Even while the newly shifted strains spread, the old normal

form also continued to circulate, and there was no realistic way to separate their infections.

Getting Sick with the Birds and the Pigs

We humans are now studying the origins of influenza and trying to prepare for what may well be the next great, and not mild, pandemic. Influenza is considered overdue for a shift. In 1997 an influenza virus that sickened chickens in Hong Kong managed to infect people and led to the immediate slaughter of all the millions of chickens in Hong Kong. It was only the beginning. The agent was and is H5N1, an avian influenza virus (AIV). There have now been more than 400 human cases of H5N1, with a mortality of 60%. Cases have all resulted from contact with a sick bird, except for a handful of cases involving very close contact with an infected person (e.g., from a chicken to a grandmother cleaning the chicken for cooking to a granddaughter). At this writing, there has been no “casual” transmission of H5N1. Although this AIV is thought to require only minor genetic changes to adapt to human cell receptors, it has not yet developed the ability to pass between human hosts through nasal aerosols. Production of a vaccine for the world’s population is an especially difficult undertaking. Not only does it take months after isolating a strain to mass-produce enough of it to make vaccines, but the medium used to grow the virus has in the past been chicken eggs—and H5N1 is a chicken disease! New high-technology, 21st-century production systems need to be developed. This AIV has spread via bird migration routes and winter contacts, and probably some human (poultry) trade routes as well. It has now occurred—resulting in the mass slaughter of many hundreds of millions of chickens and ducks—from China, Vietnam, Thailand, and Indonesia to Nigeria, Saudi Arabia, Egypt, and Turkey, as well as to Germany, Poland, Russia, and England (see Plate 15a). The “bird flu” has moved out of the headlines, but there remain catastrophic possibilities *when* it crosses species. Much effort has gone into trying to develop a sentinel system, good reporting, and an emergency vaccination containment strategy. Geographers are using GIS to overlay environment, people, fowl, and many other variables to develop predictive models of where it will emerge (see Chapter 12).

In March and April 2009 a new influenza of swine origin that *was* contagious among people emerged in Mexico. In contrast to the avian flu, this swine flu pandemic is already deemed unstoppable; scores of countries and all WHO regions have been infected, with more than 134,000 confirmed cases at this writing (Plate 15b depicts the situation as of July 2009). All 50 U.S. states and 5 territories/districts are epidemic, with 43,771 cases confirmed. H1N1, as WHO prefers to call it instead of “swine flu” (a name that led to Muslim countries’ slaughtering uninfected pigs—e.g., those raised by Coptic Christians in Egypt—and more broadly to inaccurate comparisons to the deadly “swine flu” epidemic of 1918), became a pandemic on June 11, 2009, but this was not declared until July. WHO officials were afraid of causing hysteria; they feared people would not separate the issue of geographic extent from that of severity of the contagion. They hastened to explain that despite its alarming beginning, with high reported mortality among the first cases known

in Mexico, the H1N1 epidemic has in fact proven milder than usual “seasonal” flu. Case mortality rates are less than 1%, usually 0.4% or 0.6%. The WHO stopped reporting in late July the number of cases and weekly deaths by country, and instead urged countries to shift from finding and confirming cases to monitoring trends (e.g., is the median age [13–17] getting older as contagion moves out of schools)? The H1N1 flu epidemic has the characteristics of a herald wave (a precursor outbreak announcing, or warning, of the epidemic to follow). Even as it has spread rapidly to tens of thousands in Chile, New Zealand, and other countries with winter cold in the Southern Hemisphere, it has continued, unusually, to spread through summer heat in North America. Unless population can be protected by a new vaccine quickly, it seems likely that H1N1 will explode in a great epidemic next winter, even as the “regular flu” continues to drift. Volunteers are already (as this is written) being recruited to test the first vaccine.

Combining a few human and avian influenza genetic traits with more swine influenza traits, H1N1 *may* have developed in the hog-raising (“North Carolina model”) facility with large, fetid “lagoons” of hog waste, and may have been airborne to the nearby Mexican village of initial outbreak as well as to its workers. From such a place of recombination, many variants of strains could emerge. There is some variation in mortality around the world, which inevitably raises concerns over selection of more virulent strains. Are clusters of higher mortality due to characteristics of the hosts? That is, are they less well nourished, already sick, more crowded? Risk factors weakly associated with deaths include, in the United States, asthma, obesity, respiratory disease, cardiovascular disease, and cancer (again, see www.cdc.gov). Establishing risk factors or even case mortality rates has proven difficult because of concepts discussed earlier. What proportion of infected people are clinical, and what proportion are hospitalized? What proportion of all clinical cases are reported and tested by doctors for laboratory confirmation of the virus? The disease seems mild enough that most people stay home untreated, and get well in a few days. It is proving difficult to “get a denominator” of infected population for this disease so a true case mortality can be determined. It is not known whether the first outbreak in Mexico was indeed more virulent, or just a tip of a surprising iceberg.

Modeling diffusion from such data would seem hopeless. However, a Canadian team with experience in trying to model the possible spread of severe acute respiratory syndrome (SARS) in Canada a few years previously tackled the problem in an innovative way by using the previous year’s international air travel. They analyzed flight itineraries for all passengers departing from commercial airports in Mexico between March and April 2008 (Khan et al., 2009). Their analysis showed that 2.35 million passengers flew from Mexico to 1,018 cities in 164 countries. The United States and Canada received 80.7%; the remainder of the Western Hemisphere, 8.8%; Western Europe, 8.7%; and East Asia and the remainder of the world the rest. The detailed patterns showed a strong degree of correlation with how H1N1 spread in 2009. Countries receiving more than 1,400 passengers from Mexico were at elevated risk for importation. This diffusion model was more than 92% sensitive and 92% specific in predicting importation of the new virus.

GEOGRAPHIC APPROACHES TO THE PANDEMIC OF AIDS

Among the plethora of emergent and reemergent infections—from epidemics of multiple-drug-resistant tuberculosis and malaria to outbreaks of hantavirus, bubonic plague, and Ebola—the diffusion of no other agent in recent times has had nearly the profound impact on people and their societies as has that of human immunodeficiency virus (HIV) and the acquired immunodeficiency syndrome (AIDS) it causes. AIDS was indubitably the plague of the late 20th century. Its spatial diffusion has similarities to other epidemics and can often be modeled and successfully predicted, as we shall see. The basic ecology of the virus (crossing species from an animal to adapt to human transmission) is not unique, although the challenge it poses to treatment and development of vaccines is exceptionally difficult. The disease AIDS, however, has truly been “constructed” in perception, stigmatization, blame, fear, altruism, and destitution; new social organization and political invective; and family, economic, and social collapse as no other in our time has been. The literature, both scientific and humanistic, on AIDS is already vast, while just its geographic contributions continue to accrete by the day. This final section of one chapter cannot do justice to the problem and issues involved. We aspire here only to introduce briefly various aspects of the medical geography of AIDS, to acquaint readers with the current dimension of the epidemic, and perhaps to inspire further geographic research and insight. Although grounded in diffusion analysis, any analysis of this disease brings the broad spectrum of geographic approaches to bear: spatial, social, ecological, and humanistic.

Ecology

The disease agent is different from any other described in this book. HIV is a *retrovirus* in the *lentivirus* or slow-virus family. Slow viruses taking years to produce symptoms have been known to infect many animals, but have not previously been associated with disease in people. In animals they are variously known to produce wasting, pneumonia, severe neurological complications, and some types of cancer. HIV consists of a string of ribonucleic acid (RNA) and an enzyme called reverse transcriptase. RNA usually is an intermediary that carries information from a cell's genetic code, DNA, to protein molecules in the cell and causes them to organize in the form of a copy of the DNA. This virus uses its enzyme to “write itself backward,” to copy its own RNA as two chains of DNA, and so is classified as a *retrovirus*. The new DNA then works its way into the cell's DNA and waits, apparently until the cell is triggered to start reproducing rapidly by an infection it must fight. The cell then starts to manufacture enormous quantities of the virus's RNA. The HIV is *in* the white blood cells: macrophages, monocytes, and especially “helper” T4 cells, which recognize an infection and mobilize the production of the proper antibodies. This is because the envelope of the HIV has two glycoproteins that lock onto molecules (CD4) on the surface of white blood cells and then penetrate the membrane and merge. Thus HIV uniquely accomplishes several things (which took years for researchers to understand): It hides inside the very cells responsible for finding and

destroying “nonselves”; it then destroys those cells in the process of replicating itself; and it does this at the very time they are most needed to fight an infection, and so are multiplying and concentrating and optimizing the target for infection by HIV. Eventually, of course, the white blood cells are so totally destroyed that the everyday infections and cell mutations of all kinds, usually routinely suppressed, run amok.

Many consequences of this disease process are now understood. It is months before antigens to HIV can be found in the blood, and during this time not only will tests be falsely negative, but donated blood can be a source of infection. It often takes years before an infected person develops symptoms and the vulnerabilities of AIDS, but during this time he or she is infectious to others. Some people, probably because of the frequent challenges to their immune system from sources of multiple infections (e.g., consuming contaminated water or engaging in anal sex), move into HIV replication and develop AIDS years earlier than others. For example, a child who simultaneously is infected with schistosomiasis, tuberculosis, malaria, and dysentery will not live long. HIV mutates at an extraordinary rate, several times faster than even the influenza virus, and this rapid mutation has so far helped confound efforts to develop a vaccine. The “cocktail” of anti-AIDS medicines that has been developed to interfere with the various steps of this process (such as attachment to the cells, fusion, transcriptase, replication, and passage between cells) has been successful in extending life for years and protecting white cells. It is, however, expensive beyond hope for people in developing countries and for some people in the United States. The retrovirus cannot exist in air, water, or sunlight, and thus cannot be transmitted by casual contact of hands or breath. It has been isolated from fresh blood in the stomach of mosquitoes, but not from the muscle cells or mouth parts (i.e., it does not replicate or live in a mosquito’s tissue, and thus cannot be vectored in this way, except perhaps mechanically after an interrupted meal is quickly resumed from someone else). The disease agent is contagious between people only through the passage of fluid from one body into another. Because people with hemophilia require transfusions of plasma concentrated from the blood of scores of people to get enough of the clotting factor they need, they are exceptionally at risk. In many places an entire generation has died. Many persons engaging in intravenous drug use (IDU) are similarly endangered by the efficiency of infection via injected blood. About a third of babies born to infected mothers used to be themselves infected, but that rate is going down in developed countries where new treatments are now available for the pregnant women. In sub-Saharan African countries, an infected child is usually dead within 1 year and an adult within 2 of developing AIDS.

There are several reasons to be interested in where a disease evolved. Perhaps the most important is that in the place where exposure to a disease agent has been longest, resistance and even immunity to it are also likely to have evolved the most. There must be such a place; any new virus must result from a mutation of an older one existing somewhere in space. Once understood, any mechanisms of resistance and defense might then be replicated by human technology (as in the production of a DNA vaccine). It is also important to determine whether or not a new infection is a nidal disease, and to identify the reservoir and any vectors or other components of the nidus if there is one (Chapter 4). Then research needs to determine what

the nature of human intrusion or connection with it was. As discussed in earlier chapters, the sheer dimension of land use change and population mobility in the world today makes new epidemics in the future likely. To establish surveillance and warning systems, not to mention intervention or containment, we need deeper understanding of the processes of emergence.

HIV evolved in sub-Saharan Africa. Like such disease agents as measles and influenza before it, the retrovirus apparently crossed over to people and mutated into a viable human infection. It was established over a decade ago that there are many variants of simian immunodeficiency virus (SIV), which occur widely in various primates (Gould, 1993, summarized the literature to that time in his Ch. 2; see also Grmek, 1990; Huet et al., 1990; Marlink et al., 1994). The first SIV identified—common and without symptoms in African vervet (green) monkeys, although deadly to Asian monkeys—is widespread across the continent, but can be genetically separated from human forms. Chimpanzees are known to hunt and eat vervets, so it is possible that they got the virus in this way. There are three known forms of HIV. HIV-2 occurs among the human population of West Africa (with some cases elsewhere in recent years). It is comparatively low in virulence and transmissibility, has few disease outcomes, and causes only a little reduction in lymphocytes over a long period. It is extremely closely related to an SIV common in the sooty mangabey monkey population of West Africa. As Gould (1993) put it, greater genetic “distance” and variation exist among SIVs isolated from sooty mangabeys than between them and the HIV-2s isolated from people. HIV-3 has only recently been found in central Africa, with yet too few cases identified to form a pattern. HIV-1 was first found in central Africa and has only been recently found in populations in western Africa. It was mutating and diversifying there as early as 1960 (Worobey et al., 2008). It is the virus of the world pandemic.

A very closely related virus, probably the source of HIV-1, has been isolated from a subspecies of wild chimpanzees. This may explain why chimpanzees inoculated with HIV-1 in laboratories show symptoms but do not die. It gives researchers an important, although precious, animal model. Numerous genetic subtypes, or *clades*, of HIV-1 have now been identified. Some are more contagious and some more virulent than others. Clade E, the most infectious, is the most widespread in Africa, whereas clade B spread in Europe and North America. It is also not at all certain that a vaccine developed for one clade would be effective against the other clades. It has proved especially difficult, in any event, to develop a universally effective vaccine—in large part because of the extremely high mutation rate of the virus, which results from errors it makes in its process of reverse transcription. The greatest diversity characterizing a source area has been located in west central Africa. The genetic difference between strains there has convinced scientists that the diversification of HIV-1 began long before recognition of the AIDS pandemic, emerging between 1890 and 1910 (Worobey et al., 2008). It is likely that the virus was brought to the first large cities, especially Kinshasa as it grew near the end of the 19th century, and there found the density of population and of men living away from their families and villages to establish itself as a new human disease. The virus almost certainly crossed to people as they killed and butchered

(always a bloody activity) chimpanzees for food. As logging roads penetrate the forests today, there has been a boom in very poor men hunting and marketing “bush meat.” It might have infected isolated people living in small groups and villages many times over centuries, but any effects remained contained locally. HIV needed cities, transportation routes, urban hierarchies, and wage labor migration to emerge in the world (see Chapter 5).

Within tropical Africa there are many ways in which the virus could have been spread. These ways include the penetration of skin by unsterilized instruments (knives and needles used for purposes as diverse as circumcision, tattooing, and esthetic scarification); vaccination for smallpox eradication; more recently, childhood vaccination for tuberculosis, cholera, and whooping cough; and other mass interventions, including antibiotic treatments. In developed countries, unsterilized needles are mainly used to inject heroin, as medical needles are disposed of after a single injection. In poor places without supplies or even means to sterilize used needles, especially mobile clinics, needles are routinely reused by nurses and doctors. Moreover, the blood supply remains contaminated in many developing countries. For all the world, however, the overwhelming majority of HIV transmissions have occurred through sexual intercourse. As a sexually transmitted disease (STD), it is most contagious when people have lesions from other such diseases.

The Dimension of the Problem

The rate of new infections with HIV has apparently increased in the United States in the last few years because young adults have started regarding AIDS as a chronic disease and become careless about safe sex. In the United States and other developed nations, almost universal treatment with a “cocktail” of drugs aimed at different viral processes has enabled HIV-positive people to live for years when previously they would have died. It is not only an expensive treatment, however; it is a sickening, weakening, miserable treatment for those able to tolerate it. There are strong signs today that the virus is becoming resistant even to the mixture. At the same time, however, the Global Fund for Malaria, Tuberculosis and AIDS, and especially the work of the Clinton Foundation in negotiating prices and copyright-free production, has allowed millions in developing countries to be treated at last. As the lives of fathers and mothers in Africa are prolonged, parents will be able to raise crops and earn incomes long enough to raise their children. Children today, probably born before their mothers’ infection, fill villages of orphans and populate families with grandmothers who care for 20 or more grandchildren in utter poverty. There are annual international conferences of all countries, and many agencies and nongovernmental organizations (NGOs) that report studies and statistics. With the help of the WHO, the Joint United Nations Programme on AIDS (UNAIDS) presents the best data available. These data can be accessed from the UNAIDS website (www.unaids.org); they include not only a report with the latest information (UNAIDS, 1008, at this writing), but PowerPoint presentations with maps and graphs available for papers and educational uses. Although any data presented in this textbook will be soon dated, at least they can be easily updated!

When the first edition of this text was being written in 1987, 40,770 cases of AIDS had been reported to the WHO in 1986 from 91 countries. Although 81% of these were from the Americas (primarily the United States), this was because of more complete reporting. The WHO estimated at the time that there were more than 100,000 cases of AIDS, the majority of which were in Africa. In the United States, 1% of the cases were children; 2% were blood- (i.e., transfusion-)related; fewer than 7% of cases were among women; 14.5% were among persons engaging in IDUs; 70% were in gay men; and the remainder were in men who were both gay and engaging in IDU. HIV had recently been discovered, but little serology had yet been done. It was estimated that there were 5–10 million people infected worldwide, including up to 15% of adults in some localities.

When the second edition of this text was being written in 1999, 2.2 million AIDS deaths had been reported, and about 1.5 million new AIDS cases had been added to the 3.3 million “living with HIV,” or HIV-positive—a cumulative total in 1998 of 7 million. Serological tests for HIV had been developed and become somewhat common worldwide. The WHO estimated that more than 25 million adults (42% of whom were females) and 1.5 million children were living with HIV/AIDS (compared, for example, with almost 500 million infected with malaria). Table 10.1 continues the epidemic statistics to the time of the present writing. Since many countries have no vital statistics, and cause-of-death reporting occurs mainly in selected cities and WHO special registration areas, the numbers of dead as distinct from still living with HIV have been estimated from prenatal and other clinic experience with identified populations. Extension of registration areas and sentinel cities (e.g., from 155 to 1,100 in India) and better data for districts and regions, enabling construction of age-specific rates, have improved estimates dramatically. UNAIDS (2008) revised the estimated rate at which people lived and died in specific districts and countries, leading to a fall of almost 7 million worldwide in those living with HIV.

As this third edition is being written (2009), a little over 2 million people have died from AIDS in the past year, and a few hundred thousand more than that number have become newly infected (Table 10.1). Between 32 and 34 million people are living with HIV; that is, 0.8% of the world's population is serologically positive for HIV infection. Of those with HIV/AIDS, 47% of adults are female. Of the 2.5 million children living (all too briefly) with HIV, 90% are in sub-Saharan Africa. The use of AZT in pregnant women has almost ended the birth of babies with the disease, except where it is too expensive and/or not allowed by governments. A few countries, most notably Thailand and Uganda, have even “reversed” the epidemic, or decreased the percentage of people infected. Both nations fully educated their people about the cause and means of spread. Thailand used its family planning program, which has brought the nation's fertility rate to below replacement level, to promote safe sex and condoms. Uganda termed its initially successful program “ABC” (Abstain until marriage; Be faithful to spouse; if failing A or B, use a Condom). The global pandemic is said to be “stable”—no longer increasing exponentially.

AIDS has attacked most ferociously in sub-Saharan Africa. As Table 10.1 shows, countries in southern Africa have reproductive adult populations (ages 15–49) 18–26% infected by HIV. United Nations demographic projections have shown that

TABLE 10.1. The AIDS Epidemic

	Millions ages 15–49			
	Deaths	Newly infected	Living with HIV	Percentage
World, 1998	2.2	1.5	3.3	0.8
World, 2001	2.9	0.5	40.0	1.2
World, 2008	2.1	2.5	33.2	0.8
Sub-Saharan Africa	1.6	1.7	22.5	5.0
Botswana				23.9
Lesotho				23.2
Swaziland				26.1
South Africa				18.1
Nigeria				3.1
South and Southeast Asia	0.27	0.34	4.0	0.3–0.5
India				0.3
Thailand				1.4
Indonesia				0.2
Latin America	0.07	0.12	1.8	0.5
Haiti				2.2
Brazil				0.6
Eastern Europe/Central Asia	0.06	0.15	1.6	0.1–0.9
Russia				1.1
Ukraine				1.6
North America	0.02	0.02	1.3	0.6
United States				0.7
Other ^a	0.07	0.07	2.0	<0.4

Note. Data from UNAIDS (2008).

^aOther regions are the Middle East/North Africa, Western Europe, Eastern Asia, and Oceania, all with percentages of <0.4%.

these countries may have their populations decreased by AIDS mortality instead of proceeding through the demographic transition (Chapter 5); under this mortality onslaught, the decline of overall mortality and fertility seen in other countries has been slowed but apparently will not halt. HIV was not introduced into South and Southeast Asia until the late 1980s/early 1990s. The extension of surveillance and reliable data collection to more than 100 cities in India has helped to revise estimates about how fast HIV will spread now that it's there. "Only" 0.3% of the population is positive for HIV, but that is a percentage of more than a billion people: by 2025 an additional mortality of 31 million is expected.

In the United States the CDC (2008) has also revised, improved, and extended (by 10 states) surveillance and reporting to improve the accuracy of estimates. The new system has indicated that in 2006 the United States had an estimated 56,300 new infections (95% confidence level), including 35,312 diagnosed and reported

new cases. There were 14,016 deaths from AIDS, in a population of about 300 million. Data on risk factors were gathered (CDC, 2008). The U.S. population with HIV/AIDS in 2006 was 73.9% white, 12.4% black, and 14.8% Hispanic. Of the new infections, 53% resulted from males having sex with males (MSM); 12% resulted from IDU with contaminated needles; 4% resulted from MSM + IDU; and 31% resulted from high-risk heterosexual sex (i.e., with partners engaging in IDU or MSM). Of those getting infected, 73% were male (of these cases, 72% resulted from MSM). The racial-ethnic breakdown was as follows: 35% white, 45% black, 18% Hispanic, 1% Asian/Pacific, and <1% Native American. Among the 27% of new cases in females, 80% of these resulted from heterosexual sex. The incidence of HIV infection for black females was 14.7 times, and that for Hispanic females was 3.8 times, that for white females. The purpose of this intensive data collection was to help guide local, state, and federal intervention measures and tailor them to populations at greatest risk. The main conclusion was that all intervention levels were failing to reach, educate, test, and serve the young black gay population.

MSM has always been the leading form of HIV transmission in the United States, as in Europe; it began as an epidemic by spreading in the gay population. In other parts of the world (in Asia and Latin America) HIV was introduced through IDU in the ports, and sometimes also through sex tourism, but quickly spread to the heterosexual population through a commercial sex industry. Some men then brought the infection home to their wives and unborn children. In Africa it has always spread almost entirely through heterosexual sex. These differences in the most important means of transmission are reflected in the percentages of female deaths and of women living with HIV. The ratio of women to men has remained stable as new infections, deaths, and numbers living with HIV have changed. For the world, 47% of those living with HIV are female; in the United States, 20%; in the Caribbean, 42%; in sub-Saharan Africa, 61% (hence the much greater infection of children there).

Geographic Modeling

When AIDS first came into public view, in 1981 among homosexual white men in the United States, nothing was known about its causes, origins, means of spreading, contagiousness, treatment, or prevalence anywhere. Even when the first retrovirus was identified 2 years later and case numbers began to rise into the hundreds, knowledge of these things had not advanced much. The question most frequently asked about this new disease by students in class was whether they could get it by kissing. The first reaction of geographers attempting to advance helpful knowledge, of course, was to map it. Mapping proved especially difficult because for years, even when AIDS was reported, concerns for patient confidentiality prevented the release of data. Although two dots placed on a county base map at a scale that would fit a whole state on a single piece of paper would be totally anonymous, individual data had conventionally come attached to variables (age, sex, race, income, ethnicity, education, etc.) that, if published, could be used to identify someone. (Because an

address can identify someone, it is not published in that form.) Behind several years of geographers' frustration, however, lay a more basic problem. Health professionals of all kinds, and most of the public, perceive an epidemic as spreading through a population over time, but they do not envision the process as happening over space or having geographic consequences. This was one of the points Gould was trying to make (see Figure 10.5).

The first geographic studies attempted to find spatial patterns in the occurrence of AIDS, and so to lay the basis for inferring something about the processes of its etiology and spread. What were the patterns of prevalence and recent incidence in various population groups; how extensive and dynamic were these; and how were they related to existing spatial models? In the United States, Africa, and the world in general, how was its location related to highway infrastructure and settlement system, to mobility, to international travel for different purposes? How did it relate to demographic structures and fertility levels, prevalence and knowledge of contraception, use of condoms, availability of medical care, and all the other structural and intervention variables? This set of concerns has continued into the present involvement in establishing national rural sampling frames that can be used for resource allocation and targeting intervention. (Most of the references for this section in the second edition have been moved to "Further Reading.")

The maps showed the spread of AIDS across Africa; to New York and San Francisco, then the metropolitan United States, and then the Midwest, the South, and rural areas nationally; to the great cities of Brazil and Mexico, and then to cities across South America; and at the end of the 1980s to Asia. There is clearly a pattern of hierarchical diffusion with local expansion, together with suggestions at least of linear diffusion along the interstate highways that have led to epidemiological studies of drug selling, prostitution, and truck driving. Postulates about these diffusion patterns and population composition led to spatial modeling. Gould's (1993) effort in particular to "predict the next maps" unraveled the spatial processes involved. His use of the gravity model was discussed earlier (see Figure 10.5), but that was just a small bit. He established surfaces of AIDS in the United States and in cities such as New York that could be useful for prediction and preparation in real places. Gould argued in the international (WHO) arena for the importance of a spatial perspective when the dominant paradigm was (and still is) just to subdivide the differential equation of epidemic contagion into smaller and smaller groups, until there are no data at all. That is, contact rates are determined not only for homosexuals, bisexuals, heterosexuals, prostitutes, those with IDU, and those with hemophilia (and others receiving blood transfusions), but also for white, black, Hispanic, Asian, young, old, rich, poor, female, male—tens of thousands of transmission rates reduced to absurdity (Gould, 1993, p. 161), while the epidemic moves from here to there.

The basic differences in approach indicated in these studies are not simply separations between geographers and epidemiologists. In their space-time models, geographers have been concerned with reconstructing the pathways of HIV diffusion to provide short-term predictions and advanced warning for education, health

care provision, community organization, and other responses (Gould, 1993; Lam, Fau, & Liu, 1996; Loytonen, 1991; Williams & Rees, 1994). Whether such spatial diffusion studies are working with the logistic curve to create probability surfaces or using spatial autocorrelation to examine differences among regional structures, the strength of these studies is the production of realistic local estimates that can be used to mobilize and educate communities as a whole. Rather than promoting blame, separation, marginalization, or begrudging of resources on one hand, and the endangerment of false security and ignorance among the majority on the other, studies of *place* can be used to bring people together.

Thomas (1996) and others have argued, however, that long-term policies and targeted interventions require systems models that elicit the underlying epidemic structure (i.e., demographic) and can be used to simulate the impact and effectiveness of alternative public health interventions. Thomas's multiregion models, for example, have demonstrated that population growth rates and risk activity rates are important for simulation and targeting of vaccination and other interventions. Those simulations, however, have not been effective in predicting the observed pandemic pathway and are said to need revision to specify better geographic variations in model values (i.e., a return to spatial modeling).

Little effort in all this modeling has been directed at health care delivery and remedial services. Kearns (1996) noted the limited attention to facilities such as AIDS hospice location, or indeed to the space of gay men, their experience, or the evolution of their community-based organizations. There has been some simulation, noted earlier, for targeting public health interventions. Community-based intervention is important for greater effectiveness. *Community* is used here in the British sense of an interacting subgroup of the population, rather than the U.S. meaning of a place in which people interact. One of the most successful intervention programs internationally has involved using commercial sex workers as peer educators to educate their fellow workers about HIV/AIDS and the importance of using condoms. Because not all bars in any country can be targeted for recruitment into such programs, Welsh (1994) studied the territorial extent of diffusion of information among commercial sex workers and their customers to different bars and to other communities to which they traveled. He used computer mapping and (distance) road buffers to determine the locational and other characteristics of the bars most central to the network within a provincial city in the Dominican Republic. Then he successfully used the gravity model to establish the existence of hierarchical diffusion of the information regionally.

Geographic study of the AIDS pandemic has largely failed to address either the macro-level parameters of the patterns (global economic restructuring, urbanization, and political domination) or the micro-level constraints on everyday human agency (e.g., the behavioral choices of women). Even the best of the studies discussed above have described or fitted the pandemic retrospectively. Useful predictions or interventions for other populations have been lacking, mainly because of insufficient context and resulting insight. In the following section, a few examples are briefly considered for some diffusion context.

Discourses on Contexts and Consequences

AIDS has not only caused tens of millions to die young. It has evoked more fear, prolonged suffering, loathing, distortion, desperate “acting up” and other protest, stigma, shunning, and humiliation than any contagion in half a century—probably even since drugs first treated syphilis and leprosy. It is a molecular-level, biological fact that HIV crossed species, became contagious among humans, and spread from an origin in the African rain forest. This is not a “Eurocentric construction of the region as the source of the virus” or any other “construction”; it is a biological happening. HIV is a virus. It does not discriminate, but infects all the human white blood cells it can, wherever it can find them. In the same manner, smallpox, measles, pneumonic plague, influenza, and most of our contagious diseases crossed species, became contagious, and spread from an origin. This is how new human diseases emerge. HIV may be the first to do so in recent memory, but it is almost certainly not the last, as SARS, bird flu, and cryptosporidiosis attest.

The disease system is another matter entirely. It is aspects of the behavioral vertex of the triangle of human ecology (see Figure 2.2) that build disease systems. Behavior builds transportation systems, economic systems, educational systems, medical systems, and technological research-and-development systems. It is cultures that assign meaning and significance, as well as societies that determine powerlessness and ascribe rights and worthiness.

Stigma, Blame, Denial, and Their Consequences

More than 40% of HIV infection in the world today occurs among poor females. In several U.S. cities, AIDS is the leading cause of death among all women ages 15–45 (Farmer, Connors, & Simmons, 1996). If such information seems incongruent with the popular image of AIDS as a disease of men (mostly gay men and those with IDU), geographic studies have done little to inform Americans or to analyze local cultural and global socioeconomic structures (Yeboah, 2007).

The perceptions created by the composition of early reports of the new disease in the United States did a lot of harm. The American mass media and even some medical literature referred to AIDS as “the gay plague,” and some thought it punishment from God on bad people (although children and patients with hemophilia were troublesome because they were not “to blame”). For the most part, therefore, “normal” people did not have to worry; education in high schools about “safe sex” was objectionable. Funding for research, vaccine development, or education explicitly targeted to people at greatest risk was not much of a priority. The disease came out of “deepest, darkest Africa” and probably involved immoral sexual practices or weird tribal rites. (For instance, didn’t an anthropologist say it came from tattooing monkey blood?)

Within the United States it was thus easy to deny risk by further marginalizing people with infection. Perhaps, then, it is not surprising that African governments were not very cooperative with U.S. efforts at researching HIV/AIDS. African leaders, constructing their own stories, protested that Africans were not homosexuals,

did not use drugs, and were being unfairly blamed. The Americans, probably the CIA, must have created the disease to fight the Soviets. Alternatively, maybe the virus was deliberately being tested on Africans to keep them poor and powerless. It did not matter that not even the Americans knew at that time what a retrovirus was, or that even in warfare no chemical or biological agent could be used without knowledge of an antidote to protect one's own. Denial was such that several national leaders had to suffer the loss of their sons and whole devastated villages of orphans had to be created before international research, medical care, health education, or other interventions were accepted. Meanwhile the disease spread. Because health researchers think about epidemics as processes occurring in time rather than in space, this mere snapshot of the early stage of the pandemic was widely perceived as an infection that occurred in populations at risk. Perhaps there was genetic susceptibility involved. In Asia, the epidemic potential was therefore dismissed because it was not occurring (there) in any Asian population. AIDS was obviously a *farang* disease, Thais said; perhaps it was genetic, or resistance to it was. As the epidemic started in Thailand at the end of the 1980s, denial and concern about protecting the tourist industry let the storm gather.

In the early years, the CDC declared that being ethnically "Haitian" was by itself a risk factor. Perfectly healthy people lost their jobs as a result. HIV apparently did come from Haiti to the United States as early as 1969, 12 years before its first description (Gilbert et al., 2007). It may have been reintroduced to Haiti from the United States through sex tourism. Its broader pattern in the Caribbean is a reflection of the slave trade; of neocolonial plantation and trade structures; and of more recent economic reallocations, displacement from land, and power realignment with the end of the Soviet threat. Farmer (1992) put it this way, after following the struggle for health and livelihood of Haitian villagers and describing the political and economic structures and anthropological meanings of their lives:

One way to avoid losing sight of the humanity of those with AIDS [i.e., not just to accumulate information on genes and mores and myths] is to focus on experience and insights of those who are afflicted. . . . Link the large-scale events and structures of the world AIDS pandemic to the lived experience and commentary of people. . . . For AIDS in Haiti is about proximity rather than distance. AIDS in Haiti is a tale of ties to the United States, rather than to Africa; it is a story of unemployment rates greater than 70 percent. AIDS in Haiti has far more to do with the pursuit of trade and tourism in a dirt-poor country than with [exotic sexual or religious rituals]. (pp. 262, 264)

Similarly, in East Africa AIDS is not simply a matter of diffusion along highways with truck stops and prostitutes, and of the majority of men living away from their families and laboring in mines, plantations, or urban factories. It is the history of the mobile labor force that was created by colonial economic policy, of apartheid, of currently unstable governments, and of war and environmental refugee movements—and, more broadly, of urbanization, rapid population growth, land use change, and the developmental impacts discussed in Chapter 5. Writing about public health study of the global pandemic in general, Farmer and colleagues (1996,

p. 242) discussed studies that noted the association of HIV with cigarette smoking but ignored the also present association with dirt floors in Haitian slums, or studies that examined biological forces in such detail as “mucosal and squamous epithelial discontinuity and integrity in the susceptibility to AIDS of Nairobi sex workers” but ignored the social causes found in the same studies and results as “prostitutes from the lower socioeconomic group were more than twice as likely as their counterparts in the higher socioeconomic group to be infected with HIV.” Socioeconomic factors that drive the spread of HIV usually were (and still are) effectively ignored.

On the other hand, denial of the reality of the pandemic because of denial of the negative stereotypes has denied (and continues to deny) treatment, intervention, and prevention to millions. Not only did resentment of negative stereotypes and European constructions about cultures of Africa lead to delay in identifying genetic varieties, sources, and animal responses to the virus that are thought to be crucial to developing a vaccine, but to initial defiance of education action, testing of blood in hospitals, and even use of AZT to protect newborns from HIV-positive mothers during birth. Not until Nelson Mandela announced the death of his own son from AIDS could some other African leaders admit their own loss and allow NGO programs to help. The president of South Africa denied that AIDS was caused by a virus, instead blaming malnutrition, poverty, and colonialism. Meanwhile, newborns continued to be infected by their mothers, and HIV spread to millions of vulnerable and ignorant (by government fiat) South Africans. In the United States, the denial by Christian churches and leaders of the black community that black men could be gay led to denial of peer educator programs, blood testing, or educational and empowerment programs for female teenagers, even as the epidemic grew disproportionately among them. It still does.

There is now a vast body of work on the perception, experience, impact, legends, and history of AIDS in every social science discipline and in literature and art as well. This discussion cannot do justice to this work; it merely presents a few examples.

Thailand

Thailand still has the highest prevalence (1.4%) of HIV in Asia, but it is less prevalent than 10 years ago, and the incidence is annually reduced. The northern region of Thailand is no longer being overwhelmed with dead and orphans. In 1986 six cases of AIDS were reported to the WHO. They involved homosexual men back from study and work abroad, and two men with IDU from the port. The Thai perception that AIDS was a foreign disease was confirmed by this report, and the exposure elicited little concern when no positive serology was found in clinics. Within 6 years 500,000 people were infected. The AIDS-specific mortality rate in the year 2000 exceeded the mortality from all other causes combined for the young and middle-aged population. In 1989 a “sentinel system” was established that monitored blood sampled from persons with IDU, female commercial sex workers, blood donors, patients at STD clinics, patients at antenatal clinics, and army conscripts. Maps of antenatal clinic prevalence from December 1991 to 1994 show the rapid spread from

Bangkok to the northern provinces and southern beach resorts, and over a couple of years to the northeastern provinces (Brown & Sittitrai, 1995, pp. 16–17). Prevalence in northern antenatal clinics in Chiangmai and Chiangrai had then reached more than 10%. In mid-1994, sexual transmission accounted for 75.5% of AIDS cases; of these, 1% were transmitted by homosexual or bisexual men, and 66% by heterosexual men. IDU accounted for 8%, and blood transfusion for only 0.3%. At first AIDS reached peak age incidence in the late 20s, but over a decade this peak shifted upward by 5 years. AZT is widely used today, so mother-to-child transmission now is rare. Such transmission, however, used to cause 8% of cases (see Brown & Sittitrai, 1995, from which the statistics are taken).

In 1995, about 45,000 Thais died from AIDS, more than 10% of them children; that year 65,686 new cases were diagnosed, 7% of them children. Because children die more quickly, usually in a year, they do not accumulate in the prevalence rates. There were then more than 140,000 people living with HIV/AIDS, and more than 177,000 who had already died. Thailand is said to have “reversed” the epidemic regarding new infections and otherwise to be stabilized at less than 1.5% of adults seropositive. In 2007, 850,000 people were living with HIV/AIDS, and more than 60% of these people were receiving antiretroviral treatment. Deaths had dropped to fewer than 50,000 a year. Less than 1% of pregnant women were HIV-positive and more than 95% of them received AZT treatment.

In Thailand, HIV has been mostly transmitted by heterosexual sex. Because infected men infect their wives, transmission to children through pregnancy and nursing originally constituted a large component of the epidemic. In rural hospitals in northern Thailand, more than 30% of women in a few antenatal clinics tested positive for HIV. In 1993, military conscripts were 3.7% positive. The economic consequences in loss of labor, numbers of orphans, and medical care were devastating. The Thai government had made good general health care available to most of the people, but by 2000 the cost of hospitalization for AIDS alone was more than 1 billion baht (U.S. \$50 million). In Chiangmai, more than a third of hospital beds were occupied by patients with AIDS. As in Africa, there are great micro-level geographic variations in parental death rates. These reflect the spatial diffusion expressed in the local times of arrival of the virus. The impact was first and greatest on northern Thailand. As in Africa, the father usually dies before the mother because he was infected first. The widows have an urgent need for counseling and support. Family- and community-based care are critically important and in need of strengthening.

There were also important differences, however, between stricken countries in Africa. Thailand had a vigorous economy and a rapidly rising average standard of living. Although it was still a predominantly rural and agricultural country, the Thai people had already lowered their fertility to replacement level. In some ways this increased the economic and demographic impact of AIDS morbidity and mortality. Studies in Africa have shown that most extended families can absorb one or two deaths before economic circumstances decline drastically, but in small, more nuclear Thai families, the mortality of working adults had immediate dire consequences. Most care was given by wives who often had to stop work themselves, putting the family standard of living into a downward spiral. The costs of orphaned

children to grandparents (who are usually very poor) forced their withdrawal from school and intensified the problem of child labor abuse, including child prostitution. On the positive side, the vigorous family planning effort made condom usage available and familiar to all. After a few years of initial government denial and concern for its tourist economy, the Thai people had elected a new government that put the leadership responsible for the famously successful family planning efforts in charge of AIDS prevention.

The calamitous and extremely rapid spread of HIV in Thailand concentrated attention on its sex trade (Brown & Sittitrai, 1995; Fairclough, 1995; Gould, 1993). To these studies, surveys, and commentaries, one of us (Meade) adds interpretations from 30 years' experience with the country. Visits to brothels were indeed pervasive, common, and casual in Thai culture as experienced in 1966. Most women accepted the view that it was the nature of men to seek variety and stimulation, and that their casual and occasional pleasure did not necessarily diminish their devotion as husbands and fathers. Nevertheless, most use of prostitutes fell into two categories: Single young men were introduced to sex and taught by experienced older women; and the brothels entertained on special occasions or provided hospitality to visitors such as government officials. Police reassignments and commemorative parties, military recreation, and business conventions provided routine occasions for such entertainment. Some ethnic Chinese merchants also sometimes had "minor wives" or concubines of indeterminate permanence. Most Thai men once married, however, stayed with their families in their villages and had little opportunity or desire to hurt their wives. The brothels that existed throughout the countryside were of a different nature from those today, too. For example, one brothel in a northeastern district town of 6,000 that I knew well had three prostitutes in 1966. Two were widows whose children slept together in a back room; the third was a mentally retarded woman in her 20s whom the widows took care of and called "little sister." Their biggest source of income was the district police, several of whom (as civil servants of the central government) had been stationed there away from their families.

Both the nature and dimension of commercial sex in Thailand were transformed by the Vietnam War. During the war only about 40,000 Americans were based in Thailand, an American ally, at any given time; however, hundreds of thousands came from Vietnam for "rest and recreation." The demand for prostitutes and the flood of money in base towns and Bangkok sent "merchants" to villages far and wide searching for girls. At first deception was important. Village girls were solicited to be hotel maids and restaurant waitresses or dishwashers. Soon more money than a farmer would see in years—enough money to pay off a usurious loan and save his land or to send his other children to college—was offered for a future debt bond on a daughter. Most daughters were proud and happy to be able to help their impoverished parents and siblings. Northern Thailand was especially targeted because women in the northern mountains had long been considered the most beautiful in the country because of their lighter skin. The main railroad trunk line ran between Chiangmai and Bangkok. The north had become a tourist destination for those seeking Thai silk, teak, and temples. People in the poorest region

(the northeast), who were mostly ethnic Lao or Khmer, usually resisted the solicitations, although they suffered more from unreliable rainfall and crop failure. When drought destroyed their livelihood, 1 million desperate people could flood into Bangkok looking for work. In normal years tens of thousands of men went during the dry season, when they could not work on their farms, to work in construction or transportation. Their biggest squatter slums grew up around the port (where they remain). Youth of all kinds migrated to the primate city for educational, training, and entertainment opportunities.

There was a crisis in several sectors of the Thai economy when the Thai government asked the U.S. military to leave after the war ended. As trade relations grew with booming Japan, Taiwan, Singapore, and Germany, however, a diversified tourism became a more important source of foreign exchange and employment. Thailand offered a beautiful and unique culture. The religious sculpture and temples were beautiful; the people were smiling and friendly (because Thailand had never been a colony, its people had developed little resentment or inferiority complex); delicious new food awaited on beautiful beaches. In a few years the brothels and recruitment network found more commercial opportunities than before, as more than 5 million tourists a year visited the kingdom. After "Visit Thailand Year," 1987, advertising for "the land of sex" was taken over by foreign tour operators. A German firm began to import Thai women and rent them out by the month (with returnable deposits) for business. Jumbo jet loads of Japanese businessmen were said to find it cheaper to fly to Bangkok for a weekend, stay in a luxury hotel, play golf all day, and fulfill their fantasies all night than to pay golf fees in Japan.

Thai society has changed in important ways. The country has rapidly industrialized, especially through textile and microelectronics. It is now 35% urban, although still dominated by Bangkok. Values of consumerism seem to have swept aside Buddhism. A generation of men has grown up, especially in the north, who seem to think it normal to treat women as a commodity. A few women have become very wealthy, and many view commercial sex work as an easier job—often with air conditioning—than working in a rice field in the sun. The main alternative, when available to them, is working 60-hour weeks in a sweatshop. By 1990 (as HIV spread) middlemen commonly made "green loans" (e.g., advances for rice crops at planting time) and built dream houses for fathers, understanding that when their daughters in primary school finished their mandatory education, they would be indebted. At the same time, the opium growing of the Golden Triangle (which once was almost entirely for export), together with the civil war against the central Burmese government hegemony, fostered the addiction of increasing numbers of hill tribesmen and has corrupted law enforcement at all levels. Bewildered Burmese tribal girls, sold by fathers or seeking refuge from Burmese soldiers, were pressed into brothels in Bangkok and today in Cambodia.

Therefore, within a couple of years of the relocation diffusion to Bangkok, HIV had moved from those with IDU to commercial sex workers, and within another year the third phase had exploded in the male clients. As the businessmen, soldiers, college students, and laborers went home, the fourth wave burst in their wives and girlfriends. By 1994, just 7 years after first detection, 7.6% of AIDS cases were

infants. Northern Thailand was the first region to face the tragedy. As commercial sex workers became more infected, the industry extended to solicit (sometimes by force as actual slaves) new and younger women. The utterly destitute women of Cambodia were induced to fill the brothels that Thai businessmen set up for the United Nations troops that brought elections to the killing fields.

Between 1974 and 1978, the Khmer Rouge killed and otherwise (through untreated malaria and starvation) brought about the deaths of almost 2 million Cambodians, a third of the population. The impoverished and traumatized children who survived grew to struggle in a population that in 2000 was 2% positive for HIV. International efforts have finally curtailed the runaway abuse of women and children for sex and, following Thailand's model, reduced the seropositive rate to 0.5% in the current population of 15 million.

New York City

As reported by the New York City Commission on HIV/AIDS (2005), the HIV/AIDS epidemic in New York City has increasingly affected blacks, Latinos, and women. In 2003 the percentages of cases were 44%, 32%, and 31%, respectively, compared with 31%, 25%, and 12%, respectively, in 1987. Black men are nearly three times more likely (the report says) to be living with HIV/AIDS than other New Yorkers and seven times more likely if age 40–54 (the population in Wallace's discourse, below). Since 2001 the percentage of new HIV diagnoses that are among MSM increased almost a third, so that the proportion of MSM cases among all males increased from 42 to 52%. In 2003 one-third (up from one-tenth) of people newly diagnosed were women; 90% of them were black or Latina. Those who contracted the infection through IDU decreased from 54% in 1993 (Wallace's study, below) to 34%. Although most cases now occur in people under age 25, the number and proportion of people living with AIDS who are 50 or older has increased sixfold and now constitute more than a quarter.

Once introduced to the U.S., the virus went quickly to the top of the urban hierarchy—New York City (regionally known simply as “the city”). The costs of the epidemic there, from the loss of talent to the expense of care, took a heavy toll and received a lot of attention at the time. The virus progressed from infecting gay men to those with IDU to the girlfriends of users, and then to those who prostituted themselves to support a cocaine habit, and thence to the heterosexual clients from the suburbs. What have been less noticed, however, are the regionalization and expansion of the epidemic within the city and what these show about spatial expression of social processes in the human geography of this country in general. The following ideas and analysis are based primarily on the work of Rodrick Wallace (Gould & Wallace, 1994; Wallace, 1990, 1991, 1993; Wallace & Fullilone, 1991; Wallace, Huang, Gould, & Wallace, 1997; Wallace & Wallace, 1995).

The AIDS epidemic in New York City in the 1990s was well described by Wallace as a “synergism of plagues.” The prevalence of HIV was spatially associated, or regionalized, with drug abuse, violent death from homicide and suicide, low birthweights, high unemployment, crowded housing, arson, population mobility,

and displacement. Wallace analyzed how the process began in the Bronx in the late 1970s, in what had been viable and famously vibrant ethnic communities. Rental housing became overcrowded even as employment and the tax base moved out and education in public schools deteriorated. During the 1980s, national policies of “benign neglect” of the cities became local policies of “planned shrinkage” within New York. The incidence of arson, drug pushing, street crime, drug/alcohol abuse, and suicide, as well as low birthweight and infant mortality, increased in the southern Bronx; fire, police, and medical services were withdrawn. A “meltdown of the sociogeographic structure” and social networks followed, hollowing out a burned-out section and displacing people to surrounding areas in an expansion–diffusion of social pathology. Medical service became gridlocked. Schools further deteriorated, as did employment opportunities.

After discussing the improper policy role of government in triggering the syndrome, Wallace examined the critical thresholds of investment that would be required for “community recrystallization”—which were, of course, much higher than would have been required to prevent meltdown. As the citywide population shifts (e.g., in school pupils) diffused, an “interrelated nexus of behaviors adversely affecting both public health and public order” also transferred HIV to new marginalized communities. The poverty rates of communities surrounding the city became highly predictive of diffusion. The commuter network of the city is extensive; people from Long Island commonly commute more than 2 hours each way. First via commuting homosexuals and people with IDU, but, as the pathologies spread, increasingly through the heterosexual sex trade, HIV expanded to the suburbs. In other cities, too, the spread of AIDS from the “hollowed-out center” of “physically and socially devastated, politically and economically abandoned high-density minority neighborhoods to the suburbs as a single, spatially extended disease ecosystem” was confirmed. “Our empirical results contradict the conclusions of a recent National Research Council report that AIDS will be largely confined within marginalized urban populations. In reality U.S. urban apartheid, particularly its continuing disruption of minority social structures, has markedly accelerated the diffusion of AIDS into suburban communities” (Wallace & Wallace, 1995, p. 333).

One of the most important insights from this research is the foreshadowing of how future epidemics may spread. Regarding the current greatly increased housing and service problems in the city and emergence of multiple-drug-resistant tuberculosis, Wallace et al. (1997, p. 946) observed:

The devastation from that first burnout contributed materially to the return of TB, having an impact like a war. The social disintegration and forced migration of population resulting from another wave of burnout will overwhelm existing TB control programs—including the recently implemented program of Directly Observed Therapy—resulting in markedly increased numbers of patients who fail to complete their arduous six-to-nine month program of antibiotic therapy. This is the prescription for generating multiple-drug-resistant TB, as patients repeatedly enter the hospitals, are partially treated, and then are lost in the maelstrom of burning neighborhoods. [Note: As of 2009, the DOT tuberculosis control program was actually a great success.]

New York City's resistant TB strains can be expected to hopscotch between urban ghettos along the national travel net—most cities did something equivalently stupid to New York's ghetto fire service cuts—and then to ooze out along the commuting field into the suburbs: New York City greatly dominates the nation's urban hierarchy.

North Carolina

As the AIDS epidemic—which is what it was before widespread serology for identification of HIV and development of antiretroviral medicines—spread in the United States, the virus was introduced to communities that were almost universally ill equipped to treat the multiple diseases unsuppressed by an infected immune system. Wilmington, North Carolina, is both a port city and a beach community attracting large numbers of summer tourists from all over the East Coast. It is now located at the end of Interstate 40, but even 20 years ago Wilmington was only a hundred miles from Interstate 95, the main Canada-to-Miami thoroughfare, and even closer to the Camp Lejeune U.S. Marine base. In the 1990s an unexpectedly high incidence of AIDS was reported in the rural, poor, eastern coastal plain counties of North Carolina. There was confusion about how AIDS was spreading. One theory was that, as in Africa, truck drivers and drug runners were introducing it to prostitutes and drug abusers working the truck stops along the interstate, and that this accounted for the higher incidence in the rural area. Another was that the military was spreading it. Some studies of northern cities had indicated that people with AIDS went home to die; this suggested that a return migration of infected patients to their rural and small-town origins might account for the “new” cases being counted there, especially in pleasant beach communities near their original homes. Perhaps international sailors and tourists from the port were spreading the virus in Wilmington, and workers were then taking it home to nearby counties. Or was it simply spreading among a gay community attracted to the beaches? Probably HIV had been introduced, had spread, and had emerged as cases of AIDS in all these ways.

More importantly, how were these sick people getting treatment? Where? How were people with AIDS regarded by the good people of Wilmington, and was their reception any different if they were returned “home boys,” local residents, or new arrivals? How did their treatment and conditions evolve over time, and what was their impact on Wilmington? What was it like to have AIDS or to care for those who did in a southern, conservative, rural area? Kim Elmore (2006a, 2006b) interviewed 63 service providers, people with HIV/AIDS, and other community members and leaders over the course of a year. Maps, she thought (2006, p. 226), “tended to emphasize the virus, but not the people with the virus.”

Because of the stigma, it was initially hard to obtain treatment from local doctors or clinics, or from any service provider in a private setting. The sole gay organization in the region had to provide transportation to the state medical school/research hospital (the facility of last-resort) in Chapel Hill. This allowed the marginalized people to be literally “distanced” and relieved local physicians of the need to deal with the problem. Not only did many health service providers regard AIDS as

the result of “the sins of the homosexuals,” but infected people were even ashamed before other people with HIV/AIDS, whom they themselves considered stigmatized. People who volunteered to work with or provide care for those infected with HIV would face families and coworkers who asked why they worked with “those people.” Back-doors testing for other STDs (much less stigmatized), even support groups in drug treatment centers (also less stigmatized), and similar subterfuges were undertaken. None of this, of course, promoted blood testing, early identification, prevention of dissemination, or provision of the best treatment of the time. But support groups did form. New clinics did open, although anonymity might be unattainable. People who personally knew someone with AIDS began to offer comfort more broadly. Indeed, positive aspects of rural life were the slower pace of life and the “neighborly” attitude. The worst aspects were the lack of privacy, the gossip, and the stigma shadowing people everywhere. The problems of young men with “coming out” to their families, seeking information on prevention or protection, and obtaining anonymous testing and accessible treatment continue to facilitate exposure of susceptible individuals and long periods of unrecognized infectiousness, and so to promote the spread of HIV/AIDS to a new generation.

OTHER EPIDEMICS

Maybe nothing is more inherently “geographic” than the study of the diffusion of disease at a variety of appropriate scales. This chapter has looked at how disease agents diffuse through settlement systems with social contexts. Many epidemic infections that need study have not been considered. Intervention in the recurrent epidemics of viral meningitis in West Africa come to mind. Avian flu and H1N1 are new targets. Contagious diseases have an ecology, but their characteristics, impact, and prediction are especially appropriate for spatial analysis.

REFERENCES

- Abler, R., Adams, J., & Gould, P. (1971). *Spatial organization: The geographer's view of the world*. Englewood Cliffs, NJ: Prentice-Hall.
- AIDS Action Council. (1997). *AIDS forum*. Washington, DC: C-SPAN.
- Bailey, N. J. T. (1957). *The mathematical theory of epidemics*. High Wycombe, UK: Charles Griffin.
- Bailey, N. J. T. (1975). *The mathematical theory of infectious diseases and its applications*. High Wycombe, UK: Charles Griffin.
- Bian, L. (2003). A conceptual framework for an individual-based spatially explicit epidemiological model. *Environment and Planning B*, 31, 381–395.
- Bian, L. (2007). A network model for dispersion of communicable diseases. *Transactions in GIS*, 11(2), 155–173.
- Bian, L., & Lieberman, D. (2005). Simulating spatially explicit networks for dispersion of infectious diseases. In D. Maguire, M. F. Goodchild, & M. Batty (Eds.), *GIS, spatial analysis and modeling* (pp. 245–265). Redlands, CA: ESRI Press.
- Brown, T., & Sittitrai, W. (1995). *The impact of HIV on children in Thailand*. Honolulu: East West Center Program on Population, University of Hawaii and Thai Red Cross Society Program on AIDS.

- Brownlea, A. A. (1972). Modelling the geographic epidemiology of infectious hepatitis. In N. D. McGlashan (Ed.), *Medical geography: Techniques and field studies* (pp. 279–300). London: Methuen.
- Centers for Disease Control and Prevention (CDC). (2008). Subpopulation estimates for the HIV Incidence surveillance system—United States 2006. *Morbidity and Mortality Weekly Report*, 57(36), 985–989.
- Cliff, A., Haggett, P., & Smallman-Raynor, M. (1993). *Measles: An historical geography of a major human viral disease from global expansion to local retreat, 1840–1990*. Oxford, UK: Blackwell.
- Cliff, A. D., & Haggett, P. (1982). Methods for the measurement of epidemic velocity. *International Journal of Epidemiology*, 11, 82–89.
- Cliff, A. D., & Haggett, P. (1983). Changing urban–rural contrasts in the velocity of measles epidemics in an island community. In N. D. McGlashan & J. R. Blunden (Eds.), *Geographical aspects of health* (pp. 335–348). New York: Academic Press.
- Cliff, A. D., & Haggett, P. (1988). *Atlas of disease distributions: Analytic approaches to epidemiological data*. Oxford, UK: Blackwell.
- Cliff, A. D., & Haggett, P. (1989). Spatial aspects of epidemic control. *Progress in Human Geography*, 13, 315–437.
- Cliff, A. D., Haggett, P., & Ord, J. K. (1986). *Spatial aspects of influenza epidemics*. London: Pion.
- Cliff, A. D., Haggett, P., Ord, J. K., & Versey, C. R. (1981). *Spatial diffusion: An historical geography of epidemics in an island community*. Cambridge, UK: Cambridge University Press.
- Currie, W. (1792). *Historical account of the climates and diseases of the United States of America*. Philadelphia: Dobson.
- Currie, W. (1811). *A view of the diseases most prevalent in the United States of America*. Philadelphia: Humphreys.
- Curtis, A., Mills, J. W., & Blackburn, J. K. (2007). A spatial variant of the basic reproduction number for the New Orleans yellow fever epidemic of 1878. *The Professional Geographer*, 59(4), 492–502.
- Elmore, K. (2006a). The migratory experiences of people with HIV/AIDS in Wilmington, North Carolina. *Health and Place*, 12(4), 570–579.
- Elmore, K. (2006b). Southern discomfort: AIDS stigmatization in Wilmington, NC. *Southeastern Geographer*, 46(2), 215–230.
- Fairclough, G. (1995). A gathering storm. *Far Eastern Economic Review*, 158, 26–30.
- Farmer, P. (1992). *AIDS and accusation: Haiti and the geography of blame*. Berkeley: University of California Press.
- Farmer, P., Connors, M., & Simmons, J. (Eds.). (1996). *Women, poverty and AIDS*. Monroe, ME: Common Courage Press.
- Ferguson, N. M., May, R. M., & Anderson, R. M. (1997). Measles: Persistence and synchronicity in disease dynamics. In D. Tilman & P. Kareiva (Eds.), *Spatial ecology* (pp. 137–157). Princeton, NJ: Princeton University Press.
- Gallo, R. (1986). The first human retrovirus. *Scientific American*, 224, 88–98.
- Gilbert, M. T. P., Rambaut, A., Wlasiuk, G., Spira, T. J., Pitchenik, A. E., & Worobey, M. (2007). The emergence of HIV/AIDS in the Americas and beyond. *Proceedings of the National Academy of Sciences USA*, 104(47), 18566–18570.
- Gould, P. (1993). *The slow plague: A geography of the AIDS pandemic*. Oxford, UK: Blackwell.
- Gould, P., & Wallace, R. (1994). Spatial structures and scientific paradoxes in the AIDS pandemic. *Geografiska Annaler*, 76B, 105–116.
- Gould, R. R. (1969). *Spatial diffusion* (Resource Paper No. 4). Washington, DC: Association of American Geographers Commission on College Geography.
- Grenfell, B. T., Bjornstad, O. N., & Kappey, J. (2001). Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414, 716–723.
- Grmek, M. (1990). *History of AIDS: Emergence and origin of a modern pandemic*. Princeton, NJ: Princeton University Press.
- Hagerstrand, T. (1952). *The propagation of innovation waves*. Lund, Sweden: Gleerup.

- Hagerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association*, 24, 7–21.
- Haggett, P. (1976). Hybridizing alternative models of an epidemic diffusion process. *Economic Geography*, 52, 136–146.
- Huet, T., Cheynier, R., Meyerhaus, A., Roelants, G., & Wain-Hobson, S. (1990). Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature*, 345, 356–359.
- Hunter, J. M., & Young, J. (1971). Diffusion of influenza in England and Wales. *Annals of the Association of American Geographers*, 61, 627–653.
- Johnson, N., & Muller, J. (2002). Updating the accounts: Global mortality of the 1918–1920 “Spanish” influenza pandemic. *Bulletin of the History of Medicine*, 76(1), 105–115.
- Joint United Nations Programme on AIDS (UNAIDS). (2008). *2008 report on the global AIDS epidemic*. Retrieved from www.unaids.org
- Kalipeni, E., Oppong, J., & Zerai, A. (2007). HIV/AIDS, gender, agency, and empowerment issues in Africa. *Social Science and Medicine*, 64(5), 1128–1150.
- Kearns, R. (1996). AIDS and medical geography: Embracing the other? *Progress in Human Geography*, 20, 123–131.
- Khan, K., Arino, J., Hu, W., Raposo, P., Sears, J., Calderon, F., et al. (2009). Spread of a novel influenza A (H1N1) virus via global airline transportation. *New England Journal of Medicine*, 361, 212–214.
- Lam, N. S. N. (1996). Use of space-filling curves in generating a national rural sampling frame for HIV/AIDS research. *Professional Geographer*, 48, 321–332.
- Lam, N. S. N., Fan, M., & Liu, K. B. (1996). Spatial-temporal spread of the AIDS epidemic, 1982–1990: A correlogram analysis of four regions of the United States. *Geographical Analysis*, 28, 93–107.
- Lowen, A. C., Mubareka, S., Steel, J., & Palese, P. (2007). Influenza virus transmission is dependent on relative humidity and temperature. *PloS Pathogens*, 3, 3151.
- Lowen, A. D., Steel, J., Mubareka, S., & Palese, P. (2008). High temperature (30° Centigrade) blocks aerosol but not contact transmission of influenza virus. *Journal of Virology*, 82(11), 5650–5652.
- Loytonen, M. (1991). The spatial diffusion of human immunodeficiency virus type 1 in Finland, 1982–1997. *Annals of the Association of American Geographers*, 81, 127–151.
- Loytonen, M. (1998). Time geography, health and GIS. In A. C. Gatrell & M. Loytonen (Eds.), *GIS and health* (pp. 97–110). London: Taylor & Francis.
- MacKellar, L. (2007). Pandemic influenza: A review. *Population and Development Review*, 33(3), 429–451.
- Mann, J. M., & Tarantola, D. J. M. (Eds.). (1996). *AIDS in the world: II. Global dimensions, social roots, and responses*. New York: Oxford University Press.
- Marlink, R., Kanki, P., Thior, I., Travers, K., Eisen, G., Siby, T., et al. (1994). Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science*, 265, 1587–1590.
- Nada, P. (1989). AIDS viruses of animals and man. *Los Alamos Science*, 18, 54–89.
- New York City Commission on HIV/AIDS. (2005, October). *Report of the New York City Commission on HIV/AIDS*. New York: Author.
- Pyle, G. F. (1969). The diffusion of cholera in the United States in the nineteenth century. *Geographic Analysis*, 1, 59–75.
- Pyle, G. F. (1984). Spatial perspectives on influenza inoculation, acceptance, and policy. *Economic Geography*, 60, 273–293.
- Pyle, G. F. (1986). *The diffusion of influenza*. Totowa, NJ: Rowman & Littlefield.
- Shannon, G. W., & Pyle, G. (1989). The origin and diffusion of AIDS: A view from medical geography. *Annals of the Association of American Geographers*, 79, 1–24.
- Shannon, G. W., Pyle, G., & Bashshur, R. L. (1991). *The geography of AIDS*. New York: Guilford Press.
- Smallman-Raynor, M. R., Cliff, A. D., & Haggett, P. (1992). *London international atlas of AIDS*. Oxford, UK: Blackwell.

- Smyth, F., & Thomas, R. (1996). Preventative action and the diffusion of HIV/AIDS. *Progress in Human Geography*, 20, 1–22.
- Stephenson, J. (1995). AIDS data animation maps evolving US epidemic. *Journal of the American Medical Association*, 274, 784.
- Stock, R. F. (1976). *Cholera in Africa*. Plymouth, UK: International Africa Institute.
- Thomas, R. (1992). *Geomedical systems: Intervention and control*. London: Routledge.
- Thomas, R. (1996). Modelling space–time HIV/AIDS dynamics: Applications to disease control. *Social Science and Medicine*, 43, 353–366.
- Trevelyan, B., Smallman-Raynor, M., & Cliff, A. D. (2005a). The spatial dynamics of poliomyelitis in the United States: From epidemic emergence to vaccine-induced retreat, 1910–1971. *Annals of the Association of American Geographers*, 95(2), 269–293.
- Trevelyan, B., Smallman-Raynor, M., & Cliff, A. D. (2005b). The spatial structure of epidemic emergence: Geographical aspects of poliomyelitis in north-eastern USA, July–October 1916. *Journal of the Royal Statistical Society A*, 168(Pt. 4), 701–722.
- UNAIDS. (2009). *Thailand country profile on HIV/AIDS, 2007*. Available at www.who.int/countries/tha/en.
- Wallace, R. (1990). Urban desertification, public health and public order: “Planned shrinkage,” violent death, substance abuse, and AIDS in the Bronx. *Social Science and Medicine*, 31, 801–813.
- Wallace, R. (1991). Traveling waves of HIV infection on a low dimensional, sociogeographic network. *Social Science and Medicine*, 32, 847–852.
- Wallace, R. (1993). Social disintegration and the spread of AIDS: II: Meltdown of sociogeographic structure in urban minority neighborhoods. *Social Science and Medicine*, 37, 887–896.
- Wallace, R., & Fullilone, M. T. (1991). AIDS deaths in the Bronx 1987–1988: Spatiotemporal analysis from a sociogeographic perspective. *Environment and Planning A*, 23, 1701–1724.
- Wallace, R., Huang, Y.-S., Gould, P., & Wallace, D. (1997). The hierarchical diffusion of AIDS and violent crime among U.S. metropolitan regions: Inner-city decay, stochastic resonance and reversal of the mortality transition. *Social Science and Medicine*, 44, 935–947.
- Wallace, R., & Wallace, D. (1995). U.S. apartheid and the spread of AIDS to the suburbs: A multicity analysis of the political economy and spatial epidemic threshold. *Social Science and Medicine*, 41, 333–345.
- Webster, N. (1799). *A brief history of epidemic and pestilential diseases*. Hartford, CT: Hudson & Goodwin.
- Welsh, M. J. (1994). *Peer educators’ geographic range of effect: An analysis of an AIDS intervention in the Dominican Republic*. Unpublished doctoral dissertation, University of North Carolina.
- Welsh, M. J., Puello, E., Meade, M., Komes, S., & Nutley, T. (2001). Evidence of diffusion from a targeted HIV/AIDS intervention in the Dominican Republic. *Journal of Biosocial Science*, 23, 107–119.
- Westvelt, J. D., & Hopkins, L. D. (1999). Modeling mobile individuals in dynamic landscapes. *International Journal of Geographic Information Science*, 13, 191–208.
- Williams, J. S., & Rees, P. H. (1994). A simulation of the transmission of HIV and AIDS in regional populations within the United Kingdom. *Transactions of the Institute of British Geographers*, 19, 311–330.
- Wood, W. (1988). AIDS north and south. *Professional Geographer*, 40, 266–279.
- Worobey, M., Gemmel, M., Teuwen, D. E., Haselkorn, T., Kunstman, K., Bunce, M., et al. (2008). Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*, 455, 661–664.
- Yeboah, I. E. A. (2007). HIV/AIDS and the construction of sub-Saharan Africa: Heuristic lessons from the social sciences for policy. *Social Science and Medicine*, 64(5), 1128–1150.

FURTHER READING

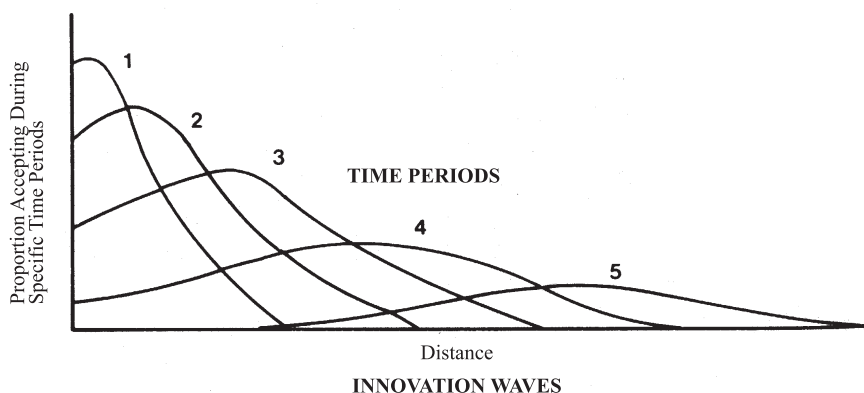
- Adesina, H. O. (1984a). The diffusion of cholera outside Ibadan City, Nigeria, 1971. *Social Science and Medicine*, 18, 421–428.
- Adesina, H. O. (1984b). Identification of the cholera diffusion process in Ibadan, 1971. *Social Science and Medicine*, 18, 429–440.
- Anderson, R. M., & May, R. M. (1991). *Infectious diseases of humans: Dynamics and control*. Oxford, UK: Oxford University Press.
- Angulo, J. J. (1987). Interdisciplinary approaches in epidemic studies: II: Four geographic models of the flow of contagious disease. *Social Science and Medicine*, 24, 57–69.
- Angulo, J. J., Haggett, P., Meghale, P., & Perderneiras, C. A. (1979). Variola minor in Braganca Paulista County, 1956: A trend–surface analysis. *American Journal of Epidemiology*, 105, 272–280.
- Burnet, M., & White, D. O. (1974). *Natural history of infectious diseases* (4th ed.). Cambridge, UK: Cambridge University Press.
- Cohn, S. E., Klein, J. D., Mohr, J. E., van der Horst, C. M., & Weber, D. J. (1994). The geography of AIDS: Patterns of urban and rural migration. *Southern Medical Journal*, 87, 599–606.
- Dutt, A. K., Monroe, C. B., Dutta, H. M., & Prince, B. (1988). Geographical patterns of AIDS in the United States. *Geographical Review*, 77, 456–471.
- Gardner, L. I., Jr., Brundage, J. F., McNeil, J. G., Visintine, R., & Miller, R. N. (1989). Spatial diffusion of the human immunodeficiency virus infection epidemic in the United States, 1985–87. *Annals of the Association of American Geographers*, 79, 25–43.
- Greenwood, B., & DeCock, K. (1998). *New and resurgent infections: Detection and management of tomorrow's infections*. Chichester, UK: Wiley.
- Kwofie, K. M. (1976). A spatio-temporal analysis of cholera diffusion in western Africa. *Economic Geography*, 52, 127–135.
- Mertens, T. E., & Low-Beer, D. (1996). HIV and AIDS: Where is the epidemic going? *Bulletin of the World Health Organization*, 74, 121–129.
- Morse, S. S. (1992). AIDS and beyond: Defining the rules for viral traffic. In E. Fee & D. M. Fox (Eds.), *AIDS: The making of a chronic disease*. Berkeley: University of California Press.
- Nijkamp, P., & Reggiani, A. (1996). Space–time synergetics in innovation diffusion: A nested network simulation approach. *Geographic Analysis*, 28, 18–37.
- Okeyo, T. M., Baltazar, G. M., & Johnston, A. (1996). *AIDS in Kenya: Background, projections, impact and interventions*. Nairobi, Kenya: National AIDS and STDs Control Programme, Ministry of Health, and National Council for Population and Development.
- Osmanov, S., Heyward, W. L., & Esparza, J. (1996). HIV-1 genetic variability: Implications for the development of HIV vaccines. *Antibiotics and Chemotherapy*, 48, 30–38.
- Pan American Health Organization. (1996). New emerging and re-emerging infectious diseases. *Bulletin of the Pan American Health Organization*, 30, 176–181.
- Panos Institute. (1992). *The hidden cost of AIDS: The challenge of HIV to development*. London: Author.
- Pollitzer, R. (1959). *Cholera*. Geneva: World Health Organization.
- Pyle, G. F., & Furuseth, O. J. (1992). The diffusion of AIDS and social deprivation in North Carolina. *North Carolina Geographer*, 1, 1–10.
- Smallman-Raynor, M. R., & Cliff, A. D. (1990). Acquired immunodeficiency syndrome (AIDS): Literature, geographical origins and global patterns. *Progress in Human Geography*, 14, 157–213.
- Thomas, R. W. (1996). Alternative population dynamics in selected HIV/AIDS modeling systems: Some cross-national comparisons. *Geographical Analysis*, 27, 108–125.
- Volberding, P. A., Sande, M. S., Lange, J., & Greene, W. C. (Eds.). (2008). *Global HIV/AIDS medicine*. Philadelphia: Saunders Elsevier.

VIGNETTE 10.1. Diffusion Waves or Stochastic Simulation of Individual Contact?

An innovation or new infectious contagion spreads one person to another, and it is a matter of chance who sits next to whom or whether on a particular day a child visits a neighbor or a grandmother. These individual movements and contacts of diffusion can be modeled stochastically (i.e., by probability), but not too many of them or too far (the networks of individuals at difference scales of Bian's stochastic modeling [see Plate 14] may in the future take individual modeling much further, but not yet). At a higher scale, probabilities of contact become related to distribution in space of those individuals. A kind of summarization of the likelihood of interaction can be made by population potential (Vignette 11.1); networks of individuals in space can be contained in social activity areas and summarized spatially with elliptical cells (Vignette 11.1). Like particles of light, however, at a higher scale of aggregation in space contact diffusion often behaves like waves through space and time.

DIFFUSION WAVES

It is often convenient to think of diffusion as waves of innovation and acceptance spreading geographically (Hagerstrand, 1952). These innovation impulses tend to lose their energy with distance from the source of the innovation. If we plot the acceptance of a new idea, or the onset of a disease contagion, for a series of time periods against the distance from the source, we can see how the innovation gradually fades with increased distance (Vignette 10.1, Figure 1). During each successive time period, the locus of greatest initial acceptance of contagion is further from the



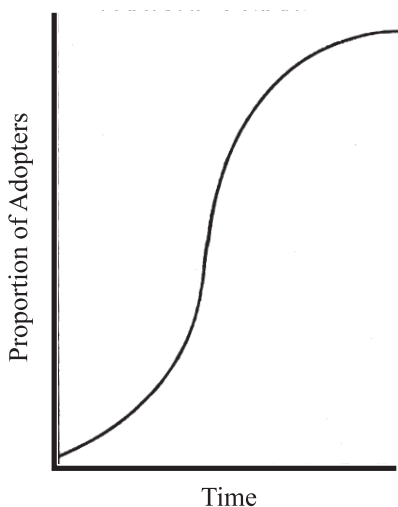
VIGNETTE 10.1, FIGURE 1. Innovation waves. From Gould (1969, p. 11). Copyright 1969 by Association of American Geographers. Adapted by permission.

source. During the first several periods, the total volume of acceptance increases; after that, the number of new acceptances decreases with each successive period. The summation of all these time curves across space and through time results in a pair of bell-shaped curves. The geographic bell is centered over the original innovation point and identifies a declining share of the total population's ever accepting the innovation with increasing distance. The second bell graphs the overall pattern of volume of acceptance (or contagion) from a small number of innovators through the great bulk of acceptors to a final few laggards.

The course of a diffusion process may be described with a logistic, or S-shaped, curve (Vignette 10.1, Figure 2). The curve is described by the formula

$$P = U/(1 + e^{a-bT})$$

where P is the proportion of adopters, T is the time at some point in the diffusion process, U is the upper limit of the adoption (10 if everyone accepts the innovation), e is the base of natural logs (2.7183), and a and b particular values that describe the location and shape of the curve (a identifies the height above the time axis where the S-shaped curve starts, and b how quickly it rises). Because a and b define a particular curve so that it is like no other, they are parameters (constants in a relational expression that determine how two or more variables change together). The logistic curve can be thought of as a cumulative frequency curve derived from the bell-shaped curve of innovation acceptors.



VIGNETTE 10.1, FIGURE 2. Logistic curve. Many diffusion studies have identified the characteristic S-shaped curve of volume of adoption through time. From Gould (1969, p. 20). Copyright 1969 by Association of American Geographers. Adapted by permission.

PROBABILITY SIMULATION AT THE INDIVIDUAL LEVEL

Within a small group, the probabilities of contact (and thus of infection) can be nearly random. However, the probability of contact between groups is often related to distance. How can this idea of a contact field, with the likelihood of infection decreasing with increasing distance from the source of the infection, be turned into an operational model to predict future diffusion?

Hagerstrand (1952) used the bell-shaped curve (i.e., the normal curve, but in three-dimensional space instead of on a sheet of paper it becomes a “bell”) of probabilities of contact to determine a *mean information field* (MIF), or an area in which contacts might occur (Vignette 10.1, Figure 3). The greatest likelihood of contact is within the central cell fitted directly over the innovator (this is like the center of the normal curve; it contains the highest frequencies). Distance decay in the chance of contact is equal in all directions. Hagerstrand used the MIF to initiate a stochastic simulation of the diffusion process. To prime the MIF, the range of numbers from 0 to 9,999 was assigned to each of its 25 cells on the basis of the cell's probability. Thus, each corner cell would receive 96 of the 10,000 digits, while the central grid gets 4,432. The cells are then numbered continuously from 0 to 9,999 by consecutively adding the probability of each cell. For instance, the top line of cells would read: 0–96; 97–236; 237–404; 405–544; 545–640. The next line of cells would continue by adding a range of 140.

The printed MIF can be used as the basis for the simulation process. The driving force that powers the model is a table of random numbers from 0 to 9,999. In a simple form of the model, one innovator exists. The MIF, which can be thought of as a grid floating over the underlying population, is centered over that innovator. He or she has the opportunity to pass the innovation to two other individuals. Those acceptors will be somewhere under the MIF grid. A random number is drawn to determine the grid location of the first acceptor. There is a nearly 45% chance that the acceptor will be in the same cell as the innovator, and a less than 1% chance that the acceptor will be in one of the corner cells. Let us assume that the first acceptor is located under the cell immediately to the left of the innovator, and that the second acceptor is within the central cell. Thus, after the first generation (the first set of transfers), two individuals in two cells of the map now have the innovation. In each subsequent generation each individual with

.0096	.0140	.0168	.0140	.0096
.0140	.0301	.0547	.0301	.0140
.0168	.0547	.4432	.0547	.0168
.0140	.0301	.0547	.0301	.0140
.0096	.0140	.0168	.0140	.0096

VIGNETTE 10.1, FIGURE 3. Mean information field (MIF).

the innovation becomes an innovator and can pass it to two other persons. In generation 2, then, the floating MIF centers over each of the new acceptors, each time with a random number drawn to identify the locations of the next acceptors. In generation 3 these four will join the early acceptors, four individuals passing the innovation to eight others.

The ultimate pattern of diffusion may vary greatly from simulation to simulation. For example, if the first acceptor was to the left of the innovator, this might pull the pattern in that direction, which is known as *direction bias*. If the patterns resulting from many simulation exercises are combined, however, their overall pattern would approximate the bell-shaped curve with its high point centered over the innovator.

This Monte Carlo simulation model can be modified in many ways to better approximate reality. Is the innovator an especially powerful transmitter of the innovation? Then we can allow that individual several new contacts each generation, while the other generations of innovators have only one. Is the disease very infectious? We can allow every innovator and acceptor to infect several new individuals each generation, with time periods being set by the incubation period of the agent. Suppose the underlying population is unevenly distributed. This population can be used to create a set of normalized probabilities, with each cell's probability based not only on distance from the innovator but also on its share of the total population. Barriers can be incorporated into the model as well. A permeable barrier (such as immunity levels) might mean that only every fourth attempted passage of the innovation into a cell beyond the barrier is successful. Other types of barriers can be similarly incorporated. Brownlea (1972), as described in the chapter text, used the real diffusion of hepatitis compared to the "random walk" in concentric circles to locate where the disease spread more quickly or met resistance. He then adjusted probabilities for each cell according to the type of clay or sand, the proportion of children or elderly people, the presence of new schools, and other variables. Next he used his new probabilities for a stochastic simulation of diffusion until it matched the reality. Then he tried to estimate the diffusion in another place entirely, adjusting probability for age distribution, soil type, and other factors in space.

The opportunities for modifying the basic model to more precisely simulate reality are nearly limitless. However, the goal should be to model the essence of the diffusion process and to reduce to a minimum the number of modifications to the MIF model. The critical essence of the model is that it enables identification of these primary influences and suggests some approximation of their relative importance. Stochastic simulation of diffusion has never had much potential in the past because even a simple model is difficult to run by hand. Computer power today makes feasible the great volume of arithmetic manipulations in a simulation modeling.

CHAPTER 11

Health Care and Promotion

The Dartmouth Atlas of Health Care (Dartmouth Medical School, 1998) proclaimed over a decade ago in its opening overview (p. 2) that in health care, “geography is destiny.” The amount of care consumed by Americans “is highly dependent on where they live—on the capacity of the health care system where they live, and on the practice styles of local physicians.” The results of the research for this atlas are discussed later in this chapter, after the role of distance, central place efficiencies and scale issues of equity, and spatial rationalization by region are presented. The spatial patterns identified by the Dartmouth medical faculty and research staff have exposed differences of enormous importance not only to improving health outcomes, but to controlling health care costs. With this atlas series, the geographic study of health service delivery has come full circle.

As discussed in Chapter 1’s history of the middle modeling years, the first American geographic studies on locational analysis and regionalization of health services began with a few initial studies by Pyle, Morrill and Earickson, and Shannon and Dever in the 1970s. This study bloomed in the 1980s into health service research that received the majority of all grants and dominated the annual medical geography paper sessions. It is difficult to appreciate today how little was known about America’s “nonsystem,” which had “just grown,” “like Topsy.” One of us (Meade), then a master’s student in geography at Michigan State University, was asked by John Hunter in 1970 to assist him in examining a problem that had been presented to him. He was hopeful of getting Meade interested in a possible question for her doctoral dissertation. The problem was that follow-up second and third vaccinations for children on public assistance in Detroit had fallen drastically over the course of a few years, and doctors at the public health clinic that administered these vaccinations had no idea why. To keep a long story short, most of the welfare recipients had been moved to a group of several new high-rise housing projects, but they still could only get the free vaccinations from the one public health clinic. It now

took three bus changes for the women (often with four or five children, and often traveling in a Michigan winter) to reach the clinic, and the same to return. No one had ever thought about location, or thought to ask the women who made the effort to get their children the first vaccination about accessibility—where they came from or how long it took to get to the clinic. (No doctoral question here.) When Pyle began his studies of patient reallocation in Chicago hospitals (discussed shortly), and when researchers in disaster emergency medicine at UCLA Medical School first began to think about distribution of blood supply and accessibility if overpasses should collapse in an earthquake, they both found that before they could address their research question they first had to locate the hospitals. Not only did no one know where patients came from, where blood was collected or stored, or where doctors were located (with or without specialties); there was not even a map of where hospitals and clinics were located or what their bed capacities were. There were no spatial data whatsoever.

Forty years later, locational analysis in health service delivery per se has almost disappeared from the geographic research menu. Most of its research questions have been solved and turned into applications for specific problems; its methodological struggles and innovations have been turned into computer programs that planners can use. The remaining needs and questions have mostly become problems of regionalization. The availability of digital data, and the capacity of a geographic information system (GIS) to manage them, analyze them, and present them for visualization, have changed everything. Using a GIS, a geographer can do in a blink what took a professor months to accomplish. Now, at Dartmouth, medical doctors, statisticians, and epidemiologists can map the country by hospital referral areas; can use hospital and Medicare data to examine medical specializations, distributions of specialists, uses of procedures, outcomes, and expenditures; and can address the big questions: Which rate of use is right? Why do some regions spend more than others? What is the value of procedures, given differential outcomes in hospital areas? Other questions, however, have morphed into higher policy issues, have been focused on different social ends, and/or have moved off into the area of health promotion. In the two previous editions of this text, Meade's coauthors Wilbert M. Gesler and Robert J. Earickson, respectively, wrote the (multiple) chapters on health service—their own specialist interest. What should we now do with one chapter in this edition? How do we address the current emphasis on health promotion and social approaches?

The changes beginning in Western society as researchers have come to understand better the development of the chronic, noncommunicable diseases of the older population (see Chapter 8) are analogous to the commencement of public health services in the 19th century. We are reminded of the wisdom of René Dubos, whose very influential book *Mirage of Health* has been discussed briefly in Chapter 2. He refused to accept that the age of infectious disease was over—that medical science had triumphed, and that medicines and such had ended the scourge of communicable illnesses. Health, you will recall, is adaptability. It was not antibiotics that lowered tuberculosis or typhoid before World War II; nor was it germ theory that did away with leprosy or bubonic plague in Europe in the 19th century. Asclepius, the

Greek god of medicine with his snake-entwined staff, has his realm: his rules, his successful treatments, and his followers. However, his sister Hygeia—goddess of hygiene, prevention of disease, and (one presumes) promotion of health—is again gathering human spirits to transform the status of health by changing cultural interactions with the environment.

So this chapter proceeds with three major sections. The first is concerned with Asclepius and summarizes the questions geographers have asked and what they have learned about health care provision. This is not a recitation (or referencing) of many hundreds of articles, but a general identification of some of the key researchers, some of the issues they raised for analysis, and what they and others found. It especially looks at the role of distance. The middle section, “Asclepius Unrobed,” looks at concerns of cultural geography about both health and health care, as well as at alternative views of these. Finally, the section on Hygeia addresses (too briefly) the social issues of health promotion and health status—in other words, the human behavior that creates health, maintains it, and prevents disease to begin with.

ASCLEPIUS: PROVISION OF MEDICAL CARE

It is not hard to identify many of the components of a health care system. There are the providers of care (types of physicians, nurses, dentists, pharmacists, paramedics, and all their assistants); there are the hospitals, clinics, pharmacies, medical offices, and all their testing equipment and technicians for analyzing everything from the activity of brain structures to urine composition; there are home health aides, social workers, nursing homes, and (today) hospices. Most of these components are shared in some form by alternative medical systems, as we shall see. Not actually part of the health care system, but forming it, surrounding it, and struggling to adapt and preserve it (perhaps its god is Vishnu), are the political, economic, and social structures of society and the beliefs of its people. Most of the following discussion of health service research revolves around the unique conditions of the U.S. system, so let us examine it and its history first.

The Health Care System of the United States

Health care delivery in the United States is unique among large industrialized nations in that it is provided through an entrepreneurial system. The main source of health care for most citizens is in a private market of thousands of independent medical practitioners, pharmacies, clinics, and so on, subject to significantly less government regulation than in other developed countries. Policies developed in other industrialized countries to provide efficient and equitable health care have entailed either direct government provision of health services or significant government involvement in various schemes of national health insurance, planning, and regulation. In the United States, however, there has always been a reliance on private markets as the favored instruments for allocating health care in all but a few categorical programs. One result of this political and economic philosophy

has been fragmentation—a collection of systems and subsystems that are based on local and institutional initiatives. There have been attempts at national, state, and local planning—hence all that geographic modeling—but for the most part each segment (e.g., hospital services, community health services, emergency medical services, or pharmaceutical supplies) has gone its own way. As in other sectors of life, in the United States a wide range of special-interest groups has had an impact on health care policy. These groups include major industrial interests; organized labor; federal, state, and local health care planning agencies; hospital organizations; professional physician groups; and the insurance industry. Attempts to bring order out of this chaos have met with limited success. As this is written, reform is being attempted again.

Several federal legislative acts have influenced the organization of the U.S. health care delivery system. The Hill–Burton Act of 1946, today universally agreed to have been “landmark legislation,” was the first of these. One of President Truman’s initiatives, the main objective of Hill–Burton was to provide federal funding for hospital construction in “needy” areas. Need was defined simply as a bed-to-population ratio; specifically, the bill mandated that there should be 4.5 beds for every 1,000 people, without discrimination (but with a separate-but-equal provision that wasn’t struck down until 1963). The federal money had to be matched by state and municipalities. It also came with conditions: States were obliged to identify general hospital service areas of one or more counties, to assume responsibility for needs assessment and planning, and to provide a “reasonable volume” of uncompensated care for those who could not pay to use a facility. Data were obtained at the county level and within metropolitan areas at the census tract level. This led to the establishment of many (usually voluntary) state health planning agencies. There was much variation among states and cities regarding adherence to the provisions of the act, given the lack of any regulatory or oversight provisions. After being extended a couple of times, in 1975 the Hill–Burton Act was transformed into Title XVI of the Public Health Service Act, with the addition of a few regulatory mechanisms.

In the 1960s the major emphasis of national health policy turned toward increased expenditures for the poor and aged. Medicaid was established to provide medical care to the poor who received government payments, and to others unable to pay basic medical costs. Medicare made a similar provision for the elderly. Both programs were and are administered by the states, with financial support from the federal government. Both pay for outpatient treatment, physician fees, hospitalization, surgery, and equipment such as walkers and pacemakers. Hospitals were originally paid their costs (without restriction) plus 2%, and lower-cost substitutes to hospitals were not covered. Over the years, health care costs rose dramatically, many-fold. Some analysts thought that insufficient controls on these programs were to blame.

Comprehensive health planning and regional medical programs (RMPs) were also begun in the mid-1960s. The former program provided funding for both state and metropolitan-area planning agencies, as well as for health care programs within institutions of higher education. RMPs were established to focus on degenerative

diseases, mainly heart disease, stroke, and cancer. The geographic boundaries of RMPs were not clearly defined, so territorial conflicts resulted. Both programs were opposed by the major medical organizations, which saw them as forms of government interference within their bailiwicks, and they were eventually abandoned as more conservative political forces prevailed at the national level.

Another landmark in U.S. health care legislation came in 1974, with passage of the National Health Planning and Resources Development Act. This act was an attempt to update and revise existing programs and to foster cooperation among the national, state, and local health care planning agencies. The new program called for each state to create clearly defined health systems agencies. These agencies' populations were to number between 500,000 and 3 million. Their boundaries were to coincide with other administrative units as much as possible. The health systems agencies became the basic units for determining health care needs, administering federal funds for approved programs, and reviewing and evaluating programs. Each state formed a board of consumers and providers, all appointed by the state's governor for the purpose of health services planning.

The 1974 act by no means overcame the problems presented by earlier programs. Although consumers were supposed to constitute a majority on state planning boards, many felt that providers had the dominant voice in planning decisions. Providers continued to resent government interference and to complain about the bureaucratization of health care delivery, the costs of maintaining health service agencies, and the long review process. Nevertheless, these agencies have continued to exist, as shown by their use to analyze data collected for the Dartmouth atlas series introduced above. Health care costs have escalated dramatically in the United States over the last several decades, from a mere \$12.7 billion in 1950 to almost \$1 trillion today. The U.S. population has aged; technology and medicines have become enormously more powerful and complex; and cost-based payments have provided little incentive for efficiency. Uniform payments to all hospitals since the 1980s have been based on predetermined rates for each of the 468 diagnosis-related groups. This change was a response to findings that there were wide differences (spatially) among hospitals in treatment charges, types of treatment, lengths of stay for similar problems, and uses of ancillary resources.

Perhaps the most dramatic development in the medical care field in the United States since 1990 is the emergence of *managed care*. Basically, managed care refers to the efforts of organizations such as insurance companies and large health maintenance organizations to hold down medical spending by "managing" the financial aspects of the practice of medicine. This translates into companies and organizations specifying how much and what form of health care will actually be allowed to be dispensed by providers. The majority of people in the United States depend on third-party payers (insurance) to cover most of their medical care costs. Supporters of managed care claim that "unnecessary" drugs, tests, procedures, and welfare programs are what drive up costs. Opponents claim that the managers are interfering with the proper practice of medicine and endangering the lives and health of patients. Supporters claim that keeping people well through early detection and promoting healthy behaviors in exercise and eating will keep costs down by pre-

venting serious illness from developing. (That was the great hope in the beginning, at least.) Detractors claim that the way to maximum profit is not to let people who are sick or at risk into a managed care organization in the first place—to “cherry-pick” healthy clients who will demand little care, and leave the others to be without care. The practice of managing health care is still relatively new and is heavily influenced by politics and corporations that have a great economic stake in the outcome of the system. Under these conditions, it is inevitable that inequities have arisen and will continue to arise (Anderson & Poullier, 1999).

Distribution and Issues of Equity

Many graduate departments of geography were abuzz in 1974 as economic geographers and their graduate students encountered Shannon and Dever's (1974) *Health Care Delivery: Spatial Perspectives*. The cover of the book featured a picture of an Ohio town with a banner stretched across Main Street that said: “Stop! We need a doctor.” Inside, with suitable descriptive discussion, were a series of crude SYMAPs (the products of an economical computer-generated mapping system that used an ordinary printer to print maps, with overlays of letters to create graduated shadings of gray density for the key). For the first time geographers saw state-level maps of the distribution of general practitioners, black doctors, female doctors, pediatricians, cardiologists, midwives, and a dozen other categorizations. Midwives and general practitioners were concentrated in southern states; black physicians were mainly in coastal states; female physicians were few and primarily in West Coast states; and specialists were in northeastern and West Coast cities, apparently. Any geographer who saw those maps had to ask: *Why* are they distributed like that? Economic/urban/social geographers thought they had theories and methodologies that could explain; some began to study health service delivery.

Medical personnel, clinics, beds, and the like are regarded as *resources*. There are two main questions about resource distribution: Are there enough? And are distributions fair? The first great question, though, is simply this: Where are the resources? As explained above, nobody knew the answers to this, by county or city or street. Then came more queries: How many of what? At what scale? All those cardiologists mapped at state scale might have been in the biggest city, with no one at all available in the rest of the state; or all the black physicians, so few, might have been in one university hospital. Once some information about distribution was established, there came a further question: Is this distribution of resources supply- or need-driven? Why are they located where they are?

Resources like registered nurses, pharmacies, pediatricians, and hospital beds were mapped “per 1,000” or “per 100,000” at various levels: state, county, city, and minor civil divisions. Ratios of physicians to populations were calculated for specialties and populations at risk for health service areas and other regionalizations (see Chapter 12, Table 12.1). Distributions were examined for equity by mapping ratios called *locational quotients* and values of the *index of segregation*. They were summarized for ethnic and economic groups, using *Gini coefficients* from *Lorenz curves*

(all of these are standard distributional measures and basic spatial statistics). They were compared with standardized mortality ratios and age-adjusted death rates (demographic/epidemiological measures that have been described earlier in this book, but that then had first to be learned). It was gradually realized that subgroups of the population (ethnic/racial groups, age categories, families vs. singles, etc.) were not distributed uniformly, and thus that the specialists who served them shouldn't be either. Normative measures of distribution were developed. Some places had many times more physicians than others; small towns and rural areas were losing doctors rapidly.

Why did doctors locate where they did? Many researchers started with the assumption that money was the main cause, but apparently it wasn't at the top of the list because physicians had a good living and community respect everywhere. Money differentials mattered mostly at the state and regional scales. Racial composition was not a factor for location at the state or city scale, but sometimes influenced location of offices within cities. The ability to pay and the needs of patients were both poor predictors of location. What mattered most were the social needs of the physicians: proximity to other professionals, place of training (which helped establish an information field), and amenities for relaxation from stress during leisure time. Physicians were concerned about the quality of the schools their children would attend, about their ability to stay up to date on medical developments, and (for those attracted to group practices) about the ability to obtain relief and share the 24-hour demands with at least one other physician. The best advice given to those small towns that wanted doctors was to build a golf course!

At a different scale, distance seemed important. Physicians seemed to cluster around hospitals where they had residence and practice privileges. In some rapidly growing suburban areas, location of offices depended on where professional office space could be built or was available in the urban morphology. Midwives and dentists in general practice might be found in small towns, but rarely a medical specialist. Clearly other locational factors were at work. Before an optimal distribution could be promoted, the reasons for the existing distribution had to be understood better.

There are two dimensions of study for this subject: the *macro scale* and the *micro scale*. Macro-scale studies, because they deal with aggregates, should be used to describe rather than to explain general patterns of resource distributions. Attempting to extrapolate from the macro scale to smaller areas, and vice versa, can easily lead to committing the ecological fallacy. Also, some variables associated with resource locations may be correlated among themselves, and it is often difficult to ascertain which variables are the most important ones. Furthermore, the relative importance of these variables may change over time. In some developing countries the data for macro-scale analysis are often simply not available.

Micro-scale studies are the preferred approach to analyzing resource location factors; at the micro level the actual mechanisms behind resource distribution decisions are clearer. A micro-scale study, for example, might focus on practitioners' motives for selecting a location.

Locational Analysis

Given the existence of inequalities in health care resource distributions, no matter how one chooses to measure that inequality, the next logical step is to try to devise means of redressing imbalances. If they're not all in the best place, where should they be located? Having a medical resource present abundantly at some scale does not make it available at another. *Accessibility* depends on many things besides physical presence, of course. Accessibility can depend on ability to pay; on membership (as in certain facilities for union workers, veterans, actors); on race (especially in the segregated South, where in the 1950s even children attacked by polio had to go to different places); and on many other factors. Social science has learned that *social distance* is also important. The language used by doctors, the ability to understand and be understood, and the attitude of medical staff toward patients matters a lot. People want to avoid being humiliated or treated as commodities. Accessibility is sometimes measured in waiting time, which can be hours in a medical reception room for some people, and even days for people in the developing world who have already walked with sick children for many miles. Just recently in the United States, a woman with a history of mental health problems went into convulsions and died on the floor of a hospital reception room where she had been waiting to be seen for 30 hours. This event at least made the national news. People without health insurance have so overloaded the emergency rooms with their advanced problems that teenagers coming in with merely broken legs may have to wait hours to be seen. People without health insurance can be, and are, turned away even from the emergency rooms of private hospitals unless their problems are life-threatening. Certainly these are all issues of accessibility.

The important measure of accessibility for *locational analysis* is distance. Distance, too, can be measured in various ways. The most common way, and the one used below, is linear or "crow-flying" distance. Today any GIS program measures this instantly as a vector. To measure distance to multiple points, there used to be "string distance" as linear distance was marked off on a string secured with a pin to a central point and then pivoted around. A wheeled instrument, or a computer cursor today, can be run along roads to measure the actual road distance the birds don't fly. For a while there was a lot of research interest in asking about travel time. The women in Detroit who needed those three bus changes to get vaccinations for their children weren't far away from the clinic as the private automobile drove; however, the bus trip took them 2 hours instead of 10 minutes. For rural people without cars, such as elderly women needing to keep medical appointments, arranging a ride from a neighbor or relative who has child care or daytime work obligations can take a long time and be unreliable. Then, too, driving distance in the winter snow on mountain roads is not at all the same thing as the same driving distance on the coastal plain in summer. So distance can be measured as a linear vector, as road travel, and as travel time. Despite such cases and special conditions, however, it has been repeatedly found that distance measures are so intercollinear that it usually does not matter which measure is used in statistical analysis. Map vector, "crow-flying" distance is much easier to collect than interviews for large groups.

Factors in Provider–Consumer Contacts: Utilization and the Role of Distance

The most important concept for geographic accessibility and locational analysis is that of *distance decay*, or the concept that “distance has friction.” Figure 11.1 illustrates three lines of decay. The X-axis is the frequency of use—how many people attend a hospital, dentist, or optometrist. The Y-axis is distance from the health resource. Because distance has friction, facilities and services of all kinds show a dropoff in use with greater distance to overcome. Distance decay for health care is not quite as linear as that for movie theaters or fast-food outlets, however. Figure 11.1 shows a hospital/clinic facility. Line A shows a steep distance decay, illustrating that only people living (or perhaps working) near the facility use it. This may be true for getting a flu shot, checking out a chronic cough, or getting a mammogram. Line B shows a low gradient: Use of the facility is much less, but people come from long distances. Someone who cuts a toe off with a lawn mower will get to the facility. Greater or lesser distance is not then an issue. Line C represents the interesting case of a steep distance decay with convenience mattering, followed at a greater distance by renewed effort to get help as people fear they may not be able to if they wait too long. So someone with food poisoning (e.g., an *E. coli* variant), may seek help quickly if it is conveniently accessible; may use over-the-counter household drugs and hang around the toilet if there are many miles to go; but may fear being unable to get to help if vomiting, dizziness, and high fever develop, and so may choose to drive there while doing so is still possible. Medical records from physician offices, clinics, hospitals can be classified by condition (giving birth, chest pains, surgery, etc.), and then patients can be matched by address to streets and blocks, or at least zip codes. Classes of such conditions have predictable distance decay curves, we now know. Hospitals or other facilities sometimes use such knowledge to “steal” rival facilities’ patients, as in taking actions to induce obstetrical patients to go a little farther to them and fill their special beds. Unusually long distance decay curves, by category, indicate underserved populations and sometimes a need for other doctors, dentists, or resources.

Accessibility and distance issues are most simply summed up in the words of a Massachusetts doctor sent by the state legislature in 1851 to investigate the new asylum for the insane and to determine why there seemed to be an increase in insanity. He traveled all around the state, visited several facilities and many more towns,

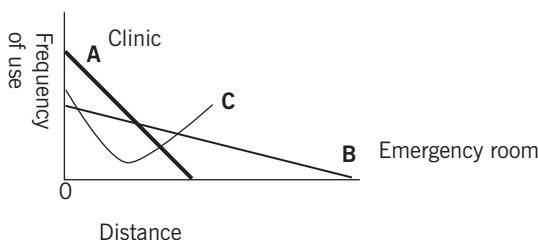


FIGURE 11.1. Three slopes of distance decay.

and wrote up his findings (Jarvis, 1851–1852). “Jarvis’s law” states, simply if crudely, that “in the vicinity of insane asylums there are more lunatics.” Thus the concepts of distance decay and medical accessibility entered the literature. Hunter, Shannon, and Sambrook (1986) followed up on Jarvis’s observations, realizing that otherwise there are almost no “admissions data” from before germ theory. They traveled to and copied admissions and address records at most of the pre-Civil War asylums that still existed (e.g., the Dorothea Dix Hospital in Raleigh, which closed only in 2008). Figure 11.2 shows the distance decay curve they calculated for the mental hospitals in 1824–1855, for eight states. Figure 11.3 shows the data mapped out for four states in the Southeast: North Carolina, Virginia, Tennessee, and Kentucky. There clearly appear to be more lunatics in the vicinity of the asylums, but we know today that this is only because distance has friction and affects utilization.

The great ranges of environmental conditions, size of states, densities and patterns of population settlement, and distances in accessibility in the United States all have profound meaning for efforts to develop national health care. As the population decreases and ages in place in the Great Plains states, for example, the distance a registered nurse may travel to visit patients recovering at home—change dressings, take blood, give antibiotics, monitor complications after surgery, or simply measure blood pressure—has increased in recent decades to over 250 miles a day. The nurse can only reach a few people a day even without winter storms. With gasoline prices increasing, the health care costs of reaching those few people in a workday are much greater than performing the same services for a retirement settlement in the Southeast or an apartment complex in the Northeast. What national health insurance plan can serve them all with the same provisions?

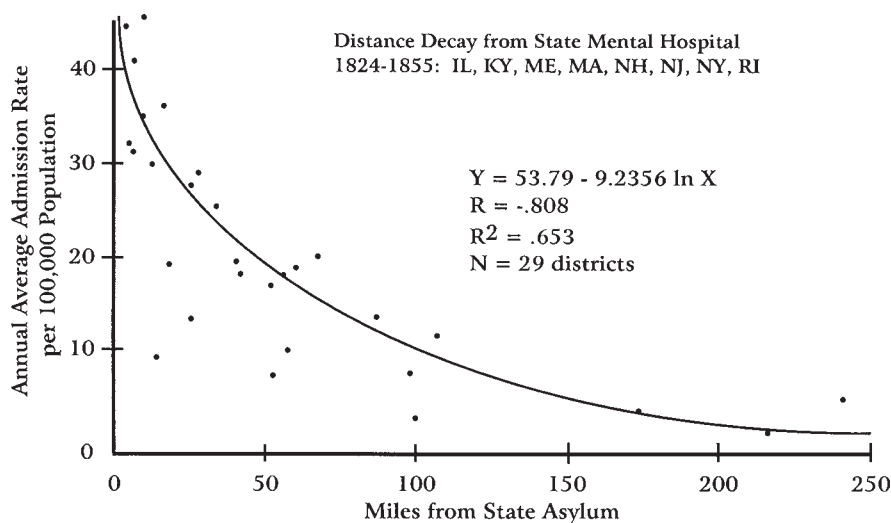


FIGURE 11.2. A distance decay curve of admissions to state mental hospitals in selected states in the 19th century. From Hunter, Shannon, and Sambrook (1986, p. 1046). Copyright 1986 by Elsevier Science. Reprinted by permission.

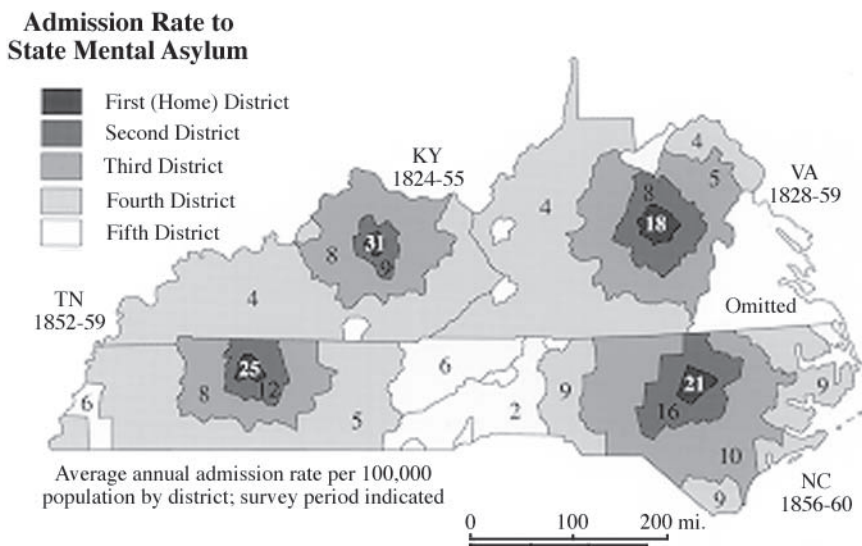


FIGURE 11.3. The effect of distance on state mental hospital admissions in selected southeastern states in the 19th century. From Hunter, Shannon, and Sambrook (1986, p. 1041). Copyright 1986 by Elsevier Science. Reprinted by permission.

Looking at the obstacle of distance from the other side, Patricia Gober (1997) looked at the role of access in state abortion rates. The number of clinics available per 100,000 women varies enormously from state to state, of course. She looked further at the impact of a new federal law requiring a 24-hour waiting period after first attending a clinic and being advised. In Philadelphia at that time, a young woman would have had a choice of going to any of 14 clinics, going home overnight, and returning the next day. In North Dakota a young woman might have needed to travel over 200 miles to Sioux Falls, the only choice, and probably stay overnight there—a prohibitive expense for her, probably, even without that winter snow.

For help in addressing some of its medical care shortfall areas, the United States has come to rely, even at state level, on the importation of foreign-trained physicians (international medical graduates, or IMGs). Aside from the issue of the quality of their training, are IMGs more likely to be found in underserved areas? Mick, Lee, and Wodchis (2000) analyzed American Medical Association data and found a positive correlation between disproportionately high numbers of IMGs and low doctor-to-population ratios.

Quite a few ways have been developed to examine accessibility in hopes of improving it. Always measuring distances from the residence does not address how people circulate (see Chapter 5). People commonly perform chores and access facilities on the way home from work (e.g., a dentist's appointment over lunch hour, a run to the pharmacy on the way home), so workplaces may be the relevant points of origin. People also make multiple-purpose trips. Someone who goes to a pharmacy, a supermarket, an exercise gym, and a dry cleaner on a weekend circulation may seek out a physician or dentist in the same activity area. Gerard Rushton started

using *small-area analysis* (subcounty, city block) population data to create *population-sheds* (i.e., population circulation areas, analogous to how the rain falling on many steep hillsides flows down eventually to one river in its basin—a *watershed*). Such population-sheds, also known as *catchment areas*, have become part of market analysis of many kinds of medical facilities—from dental care in rural Iowa to information centers for inner-city Hispanic teenagers about diabetes prevention.

Central Place Hierarchies

Central place and urban hierarchies have been explained in Chapter 10. These concepts are invoked again here because they grow out of and provide order to the distribution of goods and services. In the United States all the functions of urban hierarchies do not coincide; instead, they are usually separated functionally. Thus the top of the market spatial hierarchy, the top of the educational spatial hierarchy, the top of the government spatial hierarchy, the top of the medical spatial hierarchy, and the lower-order places overlap but seldom coincide. In Europe, the equivalents of U.S. Capitol Hill, Wall Street, Harvard, and the medical facilities of Johns Hopkins, the Mayo Clinic, and the Centers for Disease Control and Prevention would all be in Paris or London. Instead, where are they?

The origins and structures of central place hierarchy have been explained in Chapter 10. Contagion may cascade down the urban hierarchy, but we need to think of health services as being gathered up in the hierarchy. Low-level services (e.g., general practitioners, pediatricians, obstetricians, pharmacists, dentists), which need to be used by many people frequently, should theoretically be widely dispersed. High-order services, such as medical specializations and regional facilities (e.g., those for severe burns), are used rarely and so need a centralized location that can draw upon more potential patients to provide the demand. A study of Phoenix, Arizona (Gober & Gordon, 1980), did indeed show that specialists, more than primary care physicians, tend to cluster around hospitals. In any system that is top-heavy with specialists, patients must travel farther for specialized care.

Figure 11.4 shows the “demand cones” for different types of hospitals: local (satellite), suburban, city, and major research-university hospitals. Think of the center of the largest cone as 0 on a graph of distance decay (Figure 11.1) for the Y-axis, and the curves of demand that form the cones are then the distance decay curves of usage. The X-axis represents the distances people travel. The cones overlap as people go to different facilities for different needs, taking rare conditions or problems needing special equipment or specialized physicians to higher-order places. The central major research hospital in Figure 11.4 draws people from across the state, perhaps for laser eye surgery or chemoprophylaxis and radiation for cancer; however, the satellite and suburban hospitals are more cost-effective and more accessible for ordinary births (premature births are transported up the hierarchy), heart attacks, appendixes, and broken bones.

As can be seen with the array of hospitals and clinics from small and local (sometimes multiple), to regional and research hospitals drawing from an entire region or the whole nation, it is easy to link central place theory to the functional

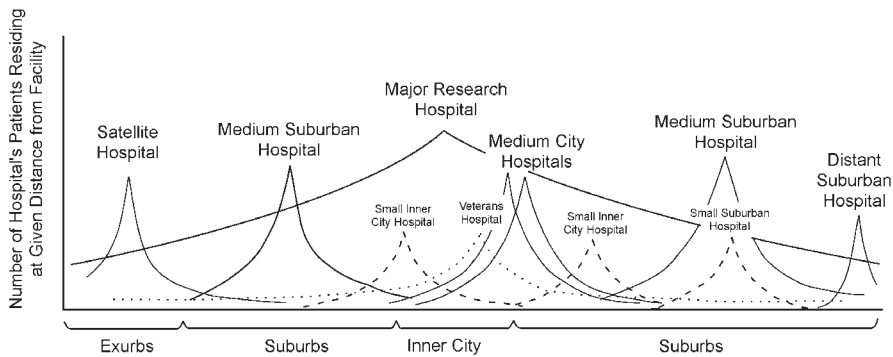


FIGURE 11.4. Distance demand cones for different types of hospitals in the typical eastern American city. Data from Pyle (1979, p. 220).

and spatial organization of an idealized health care system. Actually applying this theory, however, must take into account many factors that distort the ideal. These factors include physical barriers, transport networks, and income distributions. The result is a system that usually retains its functional nature, but whose spatial configuration rarely resembles the nested hexagons of theory. An ideal spatial organization is harder to realize in practice than is a functional hierarchy.

Central place organization in the ideal sense was attractive to several developing countries as they became independent and strove to develop a national health service system. The public health care hierarchy of several developing countries (especially former British colonies) recognized the need to get primary care to villages, even if only by mobile clinic. Therefore they tried to build a hierarchy in which more specialized care became as near and accessible to people as possible, while facing the practical need of serving enough people at each level to fully utilize their precious and too-rare medical resources. The hierarchy can be portrayed as a system ascending from village to rural health clinic, district health center, regional health center, and national health center, picking up specialized personnel/equipment and medical beds as use rises.

Figure 11.5 shows the plan for development of health care made by the Malaysian government after independence. At the bottom, in every village there was to be at least one government-trained (i.e., not traditional, although she might have been that once) midwife who understood the importance of sterilization of equipment, the danger of germs, and the symptoms of difficulty that required medical monitoring or intervention. Every five midwives were to be part of a subclinic located in the largest of the villages. Here there would be a nurse practitioner who could treat skin infections such as scabies, infected wounds, and children's earaches; take blood and sputum, and send it to be checked for malaria or tuberculosis; vaccinate babies; and provide continuing medicine for tuberculosis or leprosy. The subclinic also included a drug dispenser, as well as a public health field worker who could chlorinate wells, promote sanitation, and provide mosquito nets. The nurse practitioner acted as a screener, telling patients whom the nurse was unable to treat or

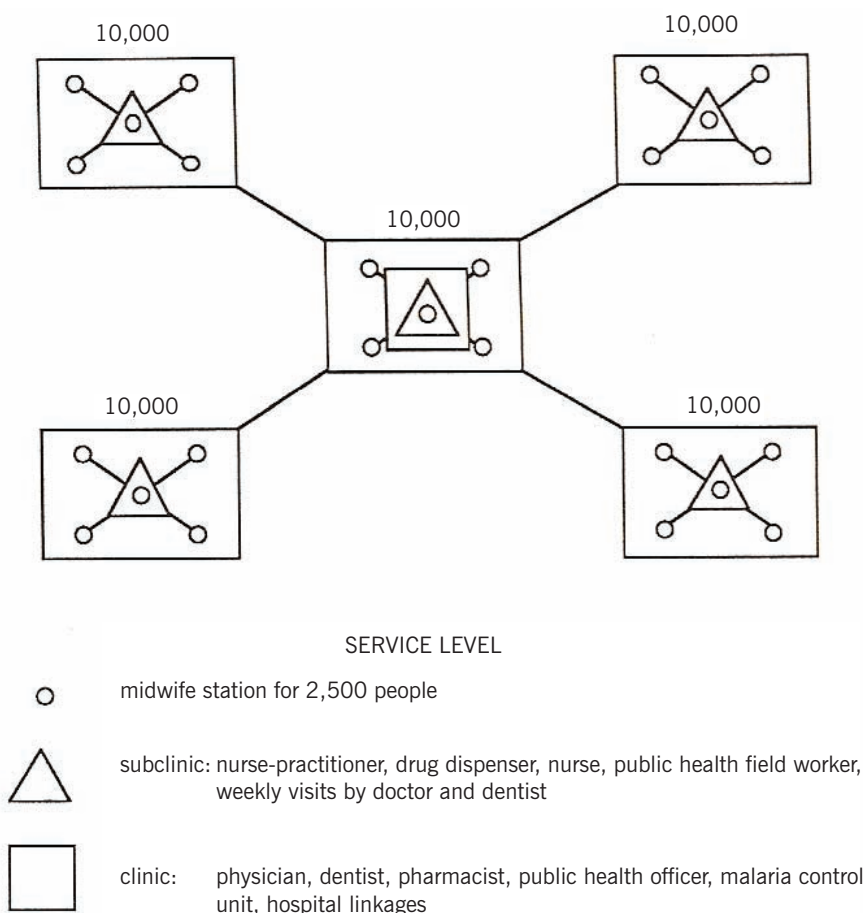


FIGURE 11.5. The ideal distribution of health care, according to the central place hierarchy planned in Malaysia.

who weren't responding well to return on Friday (1 day a week), when there would be a doctor present from the clinic—the district health center in the largest nearby town. (In Meade's experience, a dentist spent a day at the subclinic twice a month, complete with his foot-treadle drill!) At the district clinic level there were a doctor and dentist in residence (beside the ones who traveled "on circuit"), a pharmacist, the district malaria control office, and in the town a small hospital with its medical staff. The town of 10,000 thus had, by plan, five midwives; a subclinic for quick treatment and screening of the town population; a dispenser and a pharmacist; and (with hospital, resident, and traveling personnel) 10 physicians, including surgeons and specialists in infectious disease and heart disease. For treatments and illness beyond the district clinic level, patients would be referred to the medical center in the state capital, which for Selangor State was also at the time the national capital, Kuala Lumpur, with its medical school and medical research institute.

This was operationalized reality, not just the idealized plan. Of course, not everything developed so smoothly. Changes were needed. Chapter 5 has described research on a land scheme (a settlement) of the Federal Land Development Authority, and also a traditional village that was studied as a control. People on the scheme had a midwife, as did the control village, which was also host to the subclinic. To travel to the subclinic with a sick child, or one who needed vaccination, a person had to travel 7 miles along a winding, paved road through forest—and then perhaps come back another day for the doctor. It was a long walk, unless someone with a motorcycle or an official with a car could be convinced to miss work time and transport the person and the small child (or children). Beyond the village subclinic, it was another 12 miles to the town district clinic, but there was also a hospital there. For the first 15 years after independence, the Ministry of Health had thought the hard parts would be finding and training enough personnel for the subclinics and midwives, and then fully staffing the plan through villages in settled and well-populated agricultural areas as well as the new settlements being pushed into the rain forest. Over time, ministry officials realized that there was a great demand for primary care in the villages: Midwives needed to be “trained up” to provide vaccination, take care of those ear and skin infections, and send sputum and blood samples to the lab (rather like “barefoot doctors” in China then). People often preferred to bypass the subclinic, however. If they were able to get transport out of their settlement, they preferred to go all the way to the clinic in town where there were doctors and direct referral to a hospital was possible. In other words, there was a barrier to getting on the road at all, but once that barrier was overcome by need, the greater distance to higher services and expertise, less waiting or return time, and simply greater help was eagerly overcome. Gerard Rushton found the same pattern of bypassing the middle to seek higher-quality in care in India, and it has since become a generally recognized adjustment of the system’s morphology.

If these were the distortions in the “ideal” resulting from rather minor distance and time delays in higher-quality service, could a central place system be envisioned to work at a scale of hundreds of miles and more? Moreover, the market efficiency approach of planning hierarchies of medical care in a central place framework can be challenged as unjust and immoral. Access to more than primary-level health care is a need and a right; why should a rural person’s life and pain be less important than an urban person’s? Doesn’t everyone need access to X-ray radiology and mammography for diagnosis and screening? Shouldn’t C-sections and appendectomies be available to all, and clearing of clogged arteries as well? Perhaps not heart transplants or treatment for first-degree burns, but laser surgery for the eyes? Surely incubators for prematurely delivered babies should be immediately accessible. Often this issue is phrased in terms of money. For the British National Health Service, there was the “inverse care” argument that those with the greatest medical needs received the lowest spending, while the well-off middle class received attention and investment to meet its perceived needs.

There is a very large geographic dimension to this dilemma of justice. In many parts of the world there are very large, quite sparsely populated areas with people

who live hundreds of miles away from any city. Think of the tundra country in Canada, Alaska, Russia, and Scandinavia; the Australian Outback; the Great Plains and western mountains of the United States; the rain forest of the Congo; and the thousands of islands of Indonesia or other archipelagic nations in the Pacific Ocean. On the one hand, there is the problem of transporting people over vast areas of very difficult terrain to medical care—the broken bones to X-ray, the obstructed births to surgery, the delirious fevers to diagnosis and treatment. On the other hand, however, there is the problem of maintaining highly skilled, specialized personnel and equipment in small, remote centers of health care under difficult conditions for themselves and their families. These personnel are then underemployed and treat fewer than 1% of those they could serve elsewhere. This dilemma is not simply a matter of market greed, but of the spatial function of a central place hierarchy.

Location–Allocation

“Whoever said the human race was logical?” That old *Star Trek* line has seldom been applied better than to U.S. health care utilization. The “nonsystem” is a crazy quilt—private and public; preferred provider and walk-in; government Medicare, Medicaid, veterans’ care, and Indian Health Service; insurance paid for by employer, self, the government, or no one at all. In addition, medical technology and digital technology are changing rapidly (or, in the case of medical records, are not); ethnic composition and structure are changing; and suburban, exurban, and new urban settlement patterns are changing. Furthermore, apparently Americans were never simply inclined to use the nearest facility. Location–allocation methodologies in medical geography were initially premised on a rational system, at least in terms of a rational (economic) person and friction of distance. After first using the methods to plan a logical system, geographers learned to use them instead to determine and measure the other influences upon it. For a decade location–allocation planning employed the most medical geographers with master’s degrees, dominated the annual professional meetings with days of papers, and filled whole books and courses. Gradually, as its algorithms were solved, computer models took over—and then GIS made it so easy that even nongeographers could do it. It is still important today for geographers concerned with health care to understand the logic behind the planning, however.

The Chicago regional studies of the late 1960s included efforts to plan future health care provision in accordance with the community health areas mentioned earlier in this chapter (Morrill & Earickson, 1968, 1969). The question that Gerald Pyle (1971) tried to address was where future hospital beds and specialized personnel/treatment would be needed 10 and 20 years into the future of Chicago. The population was moving out to the suburbs. Race and ethnicity were shifting territorially, as were age structure, poverty/welfare, and taxpayer money. Could migration and population studies in geography predict where young, middle-aged, and older populations, or richer and poorer populations, would be located in the next decade or two? More to the point, could they predict where the future heart attacks, childbirths, and dialysis needs would be? If millions of dollars were to be spent on

new hospitals, clinics, scanners, and medical training of laboratory workers and physician specialists, where should they be located to meet future needs?

This was ambitious research. Whole methodologies had to be developed. As explained above, at first no one knew where the existing hospitals, specialized beds, physicians, or any patients were located. Hospitals had to be mapped, and their types and capacities of beds and personnel determined. Hospital discharge data were used to identify patients by types of problems, but they had to be located, and distance decay calculated and their demography projected. Computer simulation models were developed (at a time, remember, when desktops did not exist and the university mainframe had to be programmed to do procedures that took hours to solve). Pyle plotted where people originated for each hospital, using line maps. Before he could simulate where new capacity would be needed, it was necessary to analyze where capacity was already less than needed, and where there was surplus. Each patient for a hospital (by types such as heart disease, cancer, or stroke) was located and then “allocated” to the nearest bed of that type. Identifying in the models where facilities were filled, and overspilling patients had to be allocated to the next nearest facility, would identify the underserved demand. Future care needs for dialysis and heart attacks were in fact simulated, but what attracted a lot of attention from geographers was the relatively large proportion of the population that did not go to the nearest facility—sometimes even going far away. The difference between ideal location–allocation and what utilization existed helped identify social distance, as well as the use of culture-specific facilities (Catholic and Jewish hospitals, an actors’ guild hospital, a union hospital, and of course one public hospital serving the low-income and welfare-receiving public). Morrill and Earickson found that reallocating hospital space could substantially decrease patient travel, but they pointed out that the same objective could be achieved by relaxing existing constraints of income and race. One disadvantage of these early models was that they left out nonspatial aspects of health care delivery, such as disease diagnosis and the complicated rules imposed by Medicare and Medicaid. As such, they were too simplistic to be applied to planning applications today, which require socioeconomic modeling of health service delivery systems.

People or needs can be located and allocated to a facility, once its capacity is known. Think of trying to move patients in nursing homes out of the path of a major hurricane. Where is an empty bed that can be used by someone with hips in traction? Who needs oxygen, or merely careful 24-hour watch, and which facility is a match? These patients are located in this nursing home here and those in that, and where are the nearest available safe beds to which they can be allocated? One of the first uses of location–allocation was in zoning ambulances. If a person having a stroke was picked up from this apartment, or a burn victim from that road crash, where would be the nearest appropriate-for-that-service-level facility to which the patient should be transported? After a few years, the question was reversed. If ambulances should reach all heart attack victims within 20 minutes, where would emergency facilities have to be located in a sprawling urban area to do that without wasting tax money on superfluous centers?

Time went by, and the topics became closing hospitals in cities, or how utilization and unmet needs would shift when chains bought up public hospitals and converted them to use by health care organizations or religious groups. Issues of scale abound. Resource inequality is consistently found within administrative jurisdictions. Ill health related to the poverty common to U.S. inner-city neighborhoods is invisible when national, state, or even county statistics are viewed. In most countries rural dwellers have been at a disadvantage in obtaining basic primary health care, much less specialty services, because medical care resources tend to be concentrated in urban or suburban places—a vexing problem to Canada as well as the United States. As medical students in the 1960s began to shift into medical specialties (which usually are associated with higher-order centers in central place hierarchies), and as rural physicians began to age and retire, there was a great drop in the number of physicians establishing practice in rural and nonmetropolitan areas. By 1988, 111 counties across the United States were without a single physician. As a result, studies of medical personnel distribution were initiated; underserved areas were identified; and federal programs such as the Community Health Centers Initiative (which brought specialists to rural centers 1 day a week or on a similar schedule, while building and maintaining the centers) were developed (Ricketts & Cromartie, 1992). Geographers were active in the work of rural health services research centers in North Carolina and the Pacific Northwest (Hart, Pirani, & Rosenblatt, 1991), and in studies developing increasingly sophisticated location-allocation models for specific practitioner need and supply.

Regionalization

The dilemma of how to make needed health services accessible to all people, and yet maintain cost efficiencies and quality of personnel and service when people are very few and far apart, was introduced to American geography before there were any who studied health services—before even Shannon and Dever (1974). This dilemma was first examined in graduate seminars in economic geography, where students read Godlund's (1961) account of it as posed by the Swedish government to geographers at the University of Lund. The government was constructing its socialized medical system and wanted the geographers to regionalize the service areas so that everyone would have access to needed care. They were to use the existing medical resources (clinics, hospitals, nursing and medical schools, etc.) and analyze what other facilities needed to be constructed and where to provide these for the nation. Immediately the Swedish geographers confronted the central place dilemma ("on steroids," as we might say today). Their major hospitals, medical schools, and other facilities were located in southern Sweden on the arable plains, along with most of the population. High-order personnel and services, and a large portion of the population, were located around Stockholm. Some of the population, however, lived in the mountains, even north of the Arctic Circle. These people were not just reindeer pastoralists. Fishing vessels sailed out of villages in the fjords. Fishing (with its tackle, lines, and machinery) was dangerous work even without stormy weather and subzero temperatures. Men lost hands and feet. And

how could a woman with an obstructed birth, or broken bones, or a heart attack be contacted and transported 500 miles to Stockholm? How could screening for tuberculosis, various cancers, or even dental caries be brought to people there accessibly enough? It seemed obvious that one or more regional centers were needed, but how did one regionalize that? And who had to go live there doing what? Much of the development of “medical geography as health service delivery” as recounted above has involved addressing that dilemma and developing the concepts, methodologies, and techniques to do so.

Defining Regions

A *region* may be defined as a bounded area that is internally homogeneous by one or more criteria. Regions may be precisely defined by researchers using numerical measures, or they may be simply mental constructs, such as the perception of the extent of a neighborhood by its residents. Before quantitative, statistical methodology became widely used, regions were mostly *formal*, defined by nominal characteristics in common (e.g., language, religion, or government administration). In recent decades, thinking about the spatial structure of society has involved examining interrelated activities, usually with core areas and interconnections, which geographers have gradually constructed into spatial systems. Such systems may include the advertising market catchment areas of local television broadcast, newspaper delivery, and shopping center attraction; they may also include the attendance of children at various schools according to census characteristics of race, median income, welfare, location/capacity of facilities, numbers of buses, and connecting roads. The historic lack of geographic planning in the U.S. health care system is clearly shown by the numerous different definitions of regions. In North Carolina, for example, there have been the following types of multicounty regions: U.S. Public Health Service health service areas, American Medical Association service areas, Department of Veterans Affairs patient service areas (a federal regionalization for hospitals), comprehensive health planning areas, and hospital service areas (a federal regionalization related to Medicare). When all these regions are mapped and overlaid, of course there is almost no correspondence of boundaries. Nothing can be compared between regions or across the state. Health service areas sometimes incorporate more than one comprehensive health planning area, but at least they share some boundaries. Even the most basic rates are difficult to construct, as census data have until very recently been aggregated and made available only by counties or their subdivisions, not by combinations. Therefore income, education, poverty, housing, or other health-relevant data couldn't be combined to characterize a health service region.

There are both simple and complex methods for constructing service regions. Because the only geographic identifier in patient data is often the postal (zip) code of a patient's residence, a popular method of regionalizing medical service areas is to consider a medical facility's primary service area to be those zip code areas that contribute the highest proportions of patients to the facility's load. Although this method is based on distance decay, the fact that zip codes were never associated

with census data or based on any size of population, and that they are not constant but change to suit the postal load, has made them almost useless to associate with any theory or explanatory data. Potential and gravity models have been used (Chapter 10), as well as efforts to identify market areas based on those elliptical cells of social activity (see Figure 8.8 and Vignette 12.3). There is a large and growing literature on the methodology of regionalization and GIS for emergency use and for health care.

There are at least three ways to rationalize health care delivery spatially: through management of patient flows (e.g., several hospitals funnel their routine outpatient surgery cases to one large free-standing facility); through the allocation of special services, wherein expensive high-tech medical equipment is centered in a few major hospitals, or a hierarchy of clinics and hospitals is established within a region; and through managing the location of health care personnel, which follows the same philosophy as the allocation of facilities (Ricketts, Savitz, Gesler, & Osborne, 1994). There are several issues to consider in creating any regions. Geographic scale is of course important; regions in densely populated areas are necessarily smaller than those in sparsely populated areas. In small-area investigations (county or metropolitan area), data may be collected at the block or neighborhood level, whereas in rural areas only county-level data may be available or meaningful. Theoretically, a region should be geographically contiguous. However, supply and demand in health services do not always conform to spatial contiguity. Patients sometimes bypass many providers to seek care with a distant facility or doctor, as noted earlier; also, indigent patients do not always conveniently live within the same neighborhood or the same city. Using existing political or postal boundaries as medical service regional boundaries may not be realistic either. Because demographic data are grouped within political/administrative jurisdictions, in which taxes are collected and funds are spent by state and local governments as well, there is a tendency to structure service regions within these same borders. However, people often tend to ignore city, county, and state lines in their daily life transactions (remember activity areas?). Population-sheds, or market catchment areas, and the preconditions for illness certainly are not so constricted. Furthermore, political jurisdictions often lack social and economic homogeneity. For example, one large postal (zip) area in the Baltimore, Maryland, metropolitan area extends from within the city into the county. It contains neighborhoods of different racial, economic, and age structures. Chronic disease rates differ significantly in magnitude and kind from one socioeconomic class to another (see Chapter 8).

The distribution and utilization of emergency medical services have a clear impact on the well-being of society. They are good examples of hierarchical regionalization. Peters and Hall (2000) demonstrated that ambulance resources, which are an important component of emergency health care, are often insufficient, resulting in performances below those required to respond reliably to emergency calls from demand areas. They showed how to use a GIS to assess ambulance response performance and to plan changes in services. The capacity of a GIS to comprehend multiple factors, variously combine them, and select areas for which they are shared and/or different is powerful for regionalization.

As researchers began to understand better how those variables identified earlier (education, race, income, occupation, health insurance status, etc.) worked to “enable” utilization of health care services, and how variables such as knowledge of existing services, health behavior practices, and family size acted to “predispose” people to using them, they increasingly found location–allocation modeling simplistic. More attention turned toward spatial interaction modeling, as in those activity areas of old, only now equipped with *network nodes* for prevention sites. (This too was paralleled by developments in modeling disease diffusion; see the discussion of Ling Bian’s work in Chapter 10.) Recently Yassenovskiy and Hodgson (2007) have combined location–allocation modeling and hierarchical spatial system modeling (i.e., central place) with spatial interaction modeling to create “hierarchical location–allocation with spatial choice interaction modeling.” Location–allocation models applied to three levels of health service can optimize objective functions and still respect constraints on the system (such as health insurance requirements). The bypass of “optimal” locations noted above in the description of the Malaysian system has been frequent enough to seriously discount the least-distance factor, while adding attractiveness of size of facility or adjustments for multipurpose trips. Yassenovskiy and Hodgson address the problem by incorporating a spatial interaction model based on competing destinations’ attractiveness and distance, but accommodating multiple destinations and other complexities of health care utilization behavior. Thus they powerfully synthesize hierarchical, interaction, and location–allocation modeling.

GIS Analyses and Simulations: Regionalizing and Rationalizing

The availability of GIS and the ability to use it to run many kinds of simulations have both put the previous years of research and conceptualization to good use and made many of the most laborious operations routine. One such operation developed a pair of models—one access-based (i.e., distance to care) and one demand-based (i.e., efficiencies of central place)—to evaluate and address inadequacies of areas of community hospital access (Messina, Shortridge, Groop, Varnakovidia, & Finn, 2006). Like the Swedish government as described at the beginning of this section on regionalization, the state of Michigan’s Department of Community Health approached geographers at Michigan State University “with questions about how spatial analyses might be employed to develop a revised community hospital approval procedure” (p. 1). The three objectives were (1) visualizing the spatial patterns of population in relation to 136 community hospitals; (2) developing a “clear, defensible” methodology to quantify access to existing hospitals (assessing distance, travel time, road networks, etc.); and (3) contrasting the existing spatial distribution with a theoretical optimum configuration. Plate 16a shows the results of the access-based model. The areas with limited access—by travel time, most importantly—covered much of the Upper Peninsula (but notice the mostly forested and swampy center of the peninsula, with community hospitals in settlements on the two lake shores). When criteria of contiguity and minimum population density were applied, the map (not shown) changed. Only the southeastern area of limited

access (north of Detroit), and the smaller area across the state to the west (around Grand Rapids), remained. The Upper Peninsula was too sparsely populated, and areas of limited access were too fragmented.

Plate 16b shows one of the two maps of locations that would maximize coverage, based on the demand model of allocation of existing patient bed-days. The three new locations, each of which would potentially service 25–137 beds per day, were all on the northern edge of metropolitan Detroit when this 16.1-kilometer maximum distance was used. When the maximum distance was increased to 32.2 kilometers, the resulting three new locations shifted northward (map not shown). Then they were in the northeastern corner of the Michigan “mitten,” southwest of that in the center, and in the southern protuberance of the Upper Peninsula near the Wisconsin border—all near villages too small to find in an atlas. The main limitation described by Messina et al. for the access-based models was that the road network was not complete, since in Michigan all “M” class roads have been digitized, but the state still ignores federal and private roads. The main, but important, limitation for the demand-based model was that it had to be based on aggregating demand by zip code. This made disaggregating analysis impossible, and made the model sensitive to aggregation error. Despite the casual sophistication of the GIS modeling, the outcome is still striking: Remote land areas will continue to have lack of accessibility to a hospital, and population demand for facilities and services is still aggregated near a large (the largest) central place, which already has a large cluster of hospitals. It is especially striking to us, however, that the state is now seeking spatial analyses, with its hospitals mapped and its population access modeled. This is a far cry from Detroit’s request as described at the beginning of this chapter.

Hospital Privatization and Closure

This basic inequality examined above between rural and urban areas—the rural areas’ vast territory, remoteness, low population density, small or only periodic central places, poverty, and distance to health resources that those in urban areas, with transportation networks and accessible shopping facilities appropriate to their density and affluence, take for granted and demand more of—is only exaggerated at the international scale. Centrally planned, socialist political systems have sometimes promoted equal accessibility to medical resources in all areas or regions—“territorial justice”—at considerable cost to the central government, and generally a lowering of the best quality achieved. When market forces enter into these systems, the rationality of efficiency may improve the quality and accessibility of care for those in the population centers, but tends to devastate the basic care level once available in peripheral or rural areas. In the United States in the late 1980s/early 1990s, there was a surge of hospital privatization as high-cost public hospitals were bought by health management organizations of various kinds (mostly insurance companies, some church-related groups) for the use of their memberships, or by investors simply for profit. A spate of studies followed analyzing where displaced patients went, whether the region’s health care overall became more or less rational, and who got left out of or was denied what services.

The effects of privatization internationally were much more severe. The health service system of the Soviet Union collapsed almost totally with the rebirth of Russia. As discussed in Chapter 8, epidemics of diphtheria, tuberculosis, and HIV, as well as sharp increases in alcoholism, fetal alcohol syndrome, and heart attacks, increased mortality and decreased life expectancy for the first time in an industrialized country. As mortality soared, fertility fell even lower. The worst conditions have abated, but the rate of natural increase is still negative. In the cone of South America, the changes in government and market forces in Chile and Argentina drastically affected the whole system of health care, including urban–rural differences in accessibility, as Scarpaci (1989, 1990) and others studied.

In China the changes are still expressing themselves as several hundred million Chinese leave rural areas and move not only to cities (see Chapter 5), but in particular to the cities of Guangdong and other rapidly industrializing areas affected by globalization. Under the Mao Zedong regime, doctors were forced from urban hospitals to rural and sometimes remote commune hospitals (county hospitals) for a year or more of service. The top layer of medical school education was removed so that more “barefoot doctors” and other paramedics could be educated about Western medicine (especially germs, sanitation, sterilization, and vaccination). One of the aphorisms memorized by a billion people was “In matters of health, put the emphasis on rural areas.” People everywhere were extremely poor (the per capita gross national product at the time was barely U.S. \$200), but people had access to both traditional Chinese and Western medicine as practiced by doctors and in hospitals, paid for by their communes. Later, when they were so strongly encouraged to have only one or two children, they felt secure not only that their child would be likely to live but that they would be supported in their old age by the commune pension system. Under Deng Xiaoping’s agricultural responsibility system, the communes were dismantled. As the market influences have increased and migration has left hundreds of millions unconnected to the cities in which they live and work, affluence has become important for both patients and care providers. One picture we authors had no plate space left to include in this book was of a beautiful little well-equipped clinic that the government built. Emch’s photo showed that it had no patients and no personnel, since doctors were all making money in private practice, whereas the poor peasants could not afford even the government charge. The growing inequality of health care in China has been documented by Smith (1998; Smith & Fan, 1995), who explains that the urban–rural gap has been reinforced by two other trends: the preference among trained medical personnel to live and work in cities rather than rural areas, and the tendency for individuals who might otherwise have chosen a medical career to choose a more profitable one.

The Dartmouth Atlases and Their Truths

In the early 1990s the medical and statistical scholars of Dartmouth University Medical School (Hanover, New Hampshire) began studying the patterns of national utilization of Medicare funds for procedures in hospitalizations mapped and analyzed by hospital service areas. Their investigation of health care utilization and provi-

sion, using 306 hospital referral regions and Medicare expenditures, produced *The Dartmouth Atlas of Health Care 1996*. The atlas was so productive of insights, spatial pattern analyses, and findings important for policy analyses that the atlas project has continued for well over a decade now, with each atlas centering on a specialized topic of interest. The atlases are national, but there are regional and even some state volumes. They are for sale, but can be downloaded as .pdf files. Recently more interactive capacity has been added to the project's website (www.dartmouthatlas.org) so that typing in specific zip codes can bring up hospitals; search means have been added; files can be downloaded; and a tremendous research data bank has been created.

The atlases have shown, according to the website's description of the 1998 edition, that

the amount of care consumed by Americans is highly dependent on where they live—on the capacity of the health care system where they live, and on the practice styles of local physicians. Variations in the intensity of use of hospitals, the striking differences in the way terminal care is delivered, and the idiosyncratic patterns of elective surgery raise significant questions about the outcomes and value of health care. The fundamental questions posed by the Atlas are Which rate is right? How much is enough? And What is fair?

The atlases have found great variation among regions in medical resources available, but even more in practices and resource use. The national atlas edition tracking the care of patients with severe chronic illness (Wennberg, Fisher, Goodner, & Skinner, 2008) examined especially resource use during the last 2 years of life. Surprisingly, rates of elective surgery showed the least variation in pattern and were nearly random among hospitals and regions. Cardiologists, in contrast, performed nearly twice as many heart procedures in the Hospital Service Area of Miami as in that of Minnesota, and more people spent more time in intensive care in the last months of life in Miami. In Minnesota, there were greatly increased rates of knee and hip surgeries, low rates of heart procedures, and greater use of hospices. Higher-spending regions compared to lower-spending regions had a 32% higher per capita supply of hospital beds; had 75% more internists and 26% fewer family practitioners; and generally had *worse* outcomes. Differences were minor regarding rates of major elective surgery, but higher-spending regions had more hospital stays, physician visits, specialist referrals, imaging, and minor procedures. In outcomes for these regions, however, mortality over a period of up to 5 years was slightly *higher* following acute myocardial infarction, hip fracture, and colorectal cancer diagnosis.

In summation, reports on the spatial patterns found that the most intensive use of hospitals did not produce better outcomes, and urged that the rates of use of the hospitals with the best outcomes be established as normative. There was no difference in functional status as an outcome, despite the differences of intensity in treatment. Physicians' "perceptions of quality" in the most intensively served areas were more likely to include poor communication among physicians and inadequate con-

tinuity of patient care. Patients where the per capita ratios were highest were more likely to complain that specialists were inaccessible. Studies of variation “provide evidence that populations in low cost regions are not sicker or in greater medical need than those in high cost regions.” Wennberg and colleagues (2008) concluded: “A system that rewards higher costs by paying them punishes areas that use fewer resources more efficiently and are reimbursed less.”

The greatest surprise, as noted above, was that “more care is associated with higher mortality.” The explanation given was that “all medical care poses some risk, and the more procedures done, the more hospitalized, the more the risk the patient is exposed to. Receiving more care increased the chances of error being committed and complications being encountered.” It was also found that “more care does not provide patients with greater satisfaction or improved functional status.” These findings are important to the current effort (Fall 2009) to reform U.S. health care, since, among other things, they show the possibility for substantial savings and cost containment without sacrificing quality by identifying a “best-practice” model. It is too easy to imagine the opposition, however, claiming that not paying for ineffective and redundant practices in expensive-care regions is “rationing.”

In accessibility to available specialists, importance of distance as an obstacle, use of medical services and expectations of outcomes, and all such measures and judgments, we are really dealing with perceptions—with the subjectivity of people, patients, health service providers, and researchers alike. Yet in the United States there has been some commonality of belief and value from which its citizens can proceed to debate their differences. In contrast, even germ theory, let alone the efficaciousness of the medical procedures and practices analyzed in the Dartmouth Atlas series, has not been comprehended or experienced by many hundreds of millions of people in this world.

ASCLEPIUS UNROBED: CULTURAL ALTERNATIVES AND PERCEPTIONS

Up to this point in this chapter, the analyses of health services and their locations have all centered on the health care system variously called *Western medicine* or *biomedicine*. This system certainly includes aspects of cultural perception, from social distance and what to do about it to differences in usage. In Chapter 8 we have described how women have higher morbidity at every age, but men have higher mortality at every age. It's long been “known” (or at least perceived to be known) that this is at least partly because women more readily seek advice and help, whereas men often do not want to show such vulnerability. It also used to be “known” that members of one ethnic group would come to work with fever and aching bones, and that members of another would likely call in sick with a cold.

Most of our world's understanding of illness, what to do about it, and how to prevent it was developed before germ theory, however. Most traditional treatments are derived from millennia of empirical observation, although religious views certainly have had their input. For billions of people, home remedies and traditional

practitioners with their medicines remain the front line of health care. As the limitations of biomedicine have become obvious and drug resistance has resulted in a renewed search for new drugs, traditional medicine has developed a new attraction. Biomedicine, it is said, is reductionist, is remote in social distance, and speaks an incomprehensible language. It cares about organs, not people. It never listens, and it always hurries. Traditional medicine, in contrast, is seen as holistic: It cares about healing persons, listens to them, talks to their families, works with their life issues, gives them gentle drugs, and lays hands on them. Indeed, traditional medicine has increasingly become not only an alternative, but a complement, to biomedicine. This section examines the beliefs and structures of traditional medical systems that are developing as professional alternatives, especially Chinese medicine. Their use in contemporary Western society seems likely to grow.

Alternative Medicine: A General Description

Although trepanning seems to have been practiced by Neolithic peoples, the first solid evidence of medical practice has been found in the primary culture hearths that produced the earliest written records. The medicine of these ancient civilizations had roots in the supernatural; practitioners were essentially priests. There is evidence, however, of empiricism, practical organization, and the beginnings of public health. Egypt was famous in ancient times for its medical techniques. Physicians and dentists practiced there as early as 2700 B.C. The earliest known legal code, the code of King Hammurabi of Babylon (c. > 90 B.C.), contains laws on malpractice and setting medical fees. Hot and cold running water systems and urban sewerage systems were already in widespread use in the Harappan civilization of the Indus Valley by 2000 B.C. Public health and medical practice in the New World, as reported by the European explorers of the 15th and 16th centuries A.D., were on a par with those in the rest of the world. In particular, the Spaniards highly praised the sewerage systems of the Incas.

Traditional medicine exists in various forms and under a range of names, including *nonprofessional* or *indigenous* as well as *traditional*, although the more inclusive term *alternative medicine* is widely used today. Practitioners of traditional medicine are called *healers*, *native doctors*, *shamans*, or *medicine men*. The study of traditional medical systems has emerged in the discipline of *ethnomedicine*, which involves the study of beliefs and practices stemming from indigenous cultural development (Good, 1987). Four general types of traditional practitioners may be identified: spiritual or magicoreligious healers, herbalists, technical specialists such as bone setters, and traditional birth attendants. The first category, *spiritual healers*, is the most common and includes respected healers and charlatans alike. Religion and magic have always been closely tied to healing in traditional societies. Supernatural beings are believed to affect, among other things, illness and health. The healer mediates among the supernatural, the patient, and the community (Good, 1987). *Herbalists* focus on the use of medicinal plants to cure illness. Knowledge of herbal formulations is sometimes codified, sometimes simply passed down from healers to apprentices.

Efforts to identify and research the nature and potency of herbally derived medicines have become common in Western nations. *Bone setters*, as the term implies, have a somewhat narrow focus, but in some places their practice resembles that of chiropractors. *Traditional birth attendants* assist mothers at childbirth. The necessary skills may be acquired by helping experienced attendants deliver babies, or through modern training efforts sponsored within clinics and hospitals in many communities throughout the world.

Most traditional medical systems are confined to limited areas and specific populations, scattered throughout every continent (Gesler, 1984). Various Native Americans of both North and South America, numerous African tribes, and a variety of groups in Asia (from Mongols to Malays in the Indonesian islands) practice traditional medicine. Traditional healers may be the basic providers of health care for up to 90% of the rural population in South Asia and Africa.

Although it is difficult to generalize about the medical practices of all these people, some ideas and techniques are widespread. For example, diagnosis and treatment can be carried out by immediate family members, other kin, and/or group leaders, as well as by healers. The various treatments held in common include the use of medicinal herbs, prayers, the sacrificing of animals, exorcisms, the wearing of sacred objects, and the transferral of disease from one person to another. Such treatments often depend on an intimate knowledge of intracommunity relations to be successful. They seem especially effective in dealing with mental illness.

A healer may be a judge; many diseases are seen as stemming from violations of the morals and mores of society. A sick person may be seen as having broken relations with the supernatural or with other humans; his or her suffering may be viewed as a social sanction, and the healer's diagnosis as a kind of social justice. Treatment in such cases often involves a cathartic confession. Because village life is very close-knit, tensions must be resolved for the group's survival. Thus the healer also plays a role of creating "psychic unity." Treatment often includes having sick people, their relatives, and other people bring out their ill feelings toward each other. The healer's knowledge of community conflicts is important here. In many African societies, much less distinction is made between material and spiritual worlds than in Western nations. Illness may be viewed as having a social or spiritual origin, in addition to physiological causes. A psychosocial approach to medicine is therefore applied in African medicine, where a person's body, mind, and soul are conceived as an indivisible whole. A medicine may be both a substance administered to treat and prevent illness, anything used to control the spiritual cause of illness, or both. Although biomedicine appears to be eclipsing traditional treatment in sub-Saharan Africa, indigenous remedies have been considered just as effective and much cheaper, more accessible, and more easily understood.

Alternative Professional Systems

Ayurvedic medicine, unani, Galenic medicine, Chinese medicine, and biomedicine are all "professional" systems. They are highly organized and have established

arrays of techniques and codes of conduct. Each has a highly developed pharmacopoeia and a long and progressive history, and they all serve very large populations today.

The origins of *ayurvedic* medicine, which means “the science of living to a ripe age,” can be traced to the migration of Aryans into the Indus Valley around 2000 B.C. By the 6th and 5th centuries B.C., the Indian medical system had approached its present form. In this system, disease represents a disequilibrium of four *humors* (wind, bile, mucus, and blood), and cures attempt to reestablish a proper balance. Health is related to *karma*, or the effect of good and evil deeds, in both one’s former and present lives. Practitioners must have a comprehensive knowledge of the pharmacopoeia and must understand well the influences on health of climate and morality. Ayurvedic medicine today is important for both rural and urban Indians. Its core area is north India. The Dravidian culture of south India has many ayurvedic practitioners but has also established the *siddha* system, which has employed substances that purportedly could transform base metals into gold as well as aid in rejuvenating the human body. In diagnosis and treatment, *siddha* generally corresponds with ayurvedic medicine. Indian migrants have carried their ancient medical practices to other parts of the globe.

Galenic medicine, unani, and biomedicine all have their roots in Greek medicine. Hippocrates (c. 460–377 B.C.) represented the culmination of early Greek medicine. His “On Airs, Waters, and Places” (see the quotation that precedes Chapter 1) was the first great classic of medical geography: It associated certain diseases with certain climates and recognized that cultural practices and social institutions can change or temper climatic conditions. Hippocrates denied that disease had supernatural causes, and he stressed careful observation of patients. One of the many statements ascribed to Hippocrates is “Persons who are naturally very fat are apt to die earlier than those who are slender.” By his time, the idea of the four fundamental elements in nature (hot, dry, wet, and cold) and the four bodily fluids or *humors* (blood, phlegm, yellow bile, and black bile) had been developed. A healthy person maintained a proper balance of these humors, which originated from four parts of the body and were associated with the four fundamental elements (see Figure 11.6). Disease was a manifestation of improper balances among the humors and elements. Treatment depended on the nature of that imbalance.

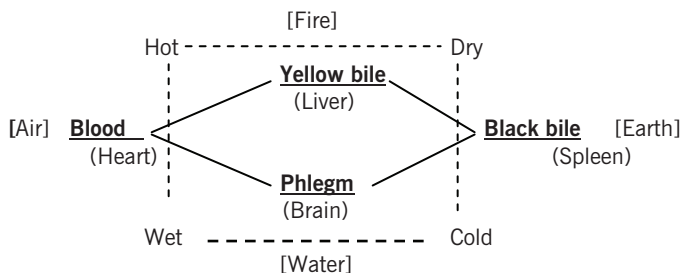


FIGURE 11.6. Hippocratic (Greek) humoral medicine.

After the time of Hippocrates, many humoral and nonhumoral schools of medicine arose in Greece. Only with the coming of Galen of Pergamon (c. A.D. 130–201) did humoral medicine gain the upper hand. Galen, the “father of experimental physiology,” was a Greek residing in Rome; he was physician to the Emperor Marcus Aurelius. His experimental insights became important texts, first for the Islamic Arab civilization and then in the early Renaissance for Western Europe. He modified the Hippocratic idea that an equilibrium of humors could be achieved, by teaching that physical, cultural, and demographic factors could cause one of the four humors to dominate and produce a unique “temperament”: sanguine, phlegmatic, choleric, or melancholic (familiar concepts to anyone acquainted with English literature). The *Galenic* body of medical thought held sway in Europe for 1,500 years. Galenic practice began to lose its importance only over the last 300 years. When Europeans began exploring and colonizing the world after A.D. 1500, Galenic medicine diffused along with many other cultural traits. When American physicians in the early 19th century (before germ theory was developed) performed bloodletting or used leaches, for example, they were trying to reduce severe fever by releasing blood (hot, heart); and when a person was in a black, melancholic mood (depression) or bilious, Galenic physicians recommended walking outside in the fresh air to get the heart pumping and blood circulating, as well as drinking hot tea, in an effort to rebalance the humors.

The Arabs took over the system of Greek medicine in the 7th and 8th centuries A.D. and called it *unani* (“Greek” in Arabic). In common with other early medical systems, *unani* had a strong ethical element; it stressed the importance of the doctor–patient relationship in healing, as well as the influence of beliefs on disease and health. Western incursions into the Arab world in the 19th century led to the decline and eventual stagnation of *unani*, as biomedicine supplanted *unani* to a large extent. However, it is still important in South Asia, where it was introduced by Muslim conquerors and spread throughout the Mughal Empire. *Unani* exists alongside systems like ayurvedic, siddha, and biomedicine in India, Pakistan, Sri Lanka, and other South Asian countries. Some of these countries have established state and private pharmaceutical companies that produce *unani* drugs. *Unani* is also important in other areas with strong Muslim influence, Southeast Asia in particular.

The foundations of Chinese medicine share with Greek medicine such strong associations of elements, organs, influences, and personality states that the roots must go back further than either. They also differ fundamentally. The Yellow Emperor’s medical treatise, written between 2600 and 1600 B.C., codified the thoughts of Chinese medical practitioners on disease causation and treatment. The world’s oldest medical text, it was nevertheless written long after Chinese civilization was expert in both bronze and iron and higher alloy metallurgy. The Chinese added metal to their view of the nature of the cosmos, and so have five elements to the Greek four. As Table 11.1 shows, Chinese cosmology is organized in fives. Besides those in the table, there are also five directions (north, east, south, west, and center), five major planets, and five of many other things. Think of the possible associations that arise from this macrocosm and microcosm. For example, joy is linked with the heart and blood vessels; with fire; with the Chinese color of celebration, red; and with growth

TABLE 11.1. Divisions and Connections of Chinese Medicine

Five divisions of the macrocosm					
Five elements	Five seasons	Five taste qualities	Five colors	Five atmospheric influences	Five stages
Wood	Spring	Sour	Blue	Wind	Birth
Fire	Summer	Bitter	Red	Heat	Growth
Earth	Late Summer	Sweet	Yellow	Humidity	Change
Metal	Autumn	Sharp	White	Drought	Maturity
Water	Winter	Salty	Black	Cold	Storage
Five divisions of the microcosm					
Five elements	Five sense organs	Five structures	Five Fus	Five Tsangs	Five emotions
Wood	Eye	Sinews	Bile	Liver	Anger
Fire	Tongue	Blood vessels	Small intestine	Heart	Joy
Earth	Mouth	Muscles	Stomach	Spleen	Anxiety
Metal	Nose	Hair	Large intestine	Lungs	Sadness
Water	Ears	Bones	Bladder	Kidneys	Fear

and summer. Sadness is associated with lung and nose (congestion) in mature (late middle) age; with drought and sharp metal; and with white, the color of mourning. Anxiety of course is associated with “butterflies” or worse in the stomach, tense muscles, and dry mouth; with the uncertainties of change (in the context of Table 11.1, especially puberty); and with yellow, the color of cowardice (even in English). American English even recognizes five seasons: On the East Coast, which has a climate very similar to China’s East Coast, we know the few weeks of low humidity and clear, sunny, warm days after leaves start to turn as Indian Summer.

Disease is considered in Chinese medicine to be the result of disharmonies within the body, between humans and the environment (the microcosm and macrocosm) and even throughout the universe. The treatment of disease arises from philosophical concepts such as those of *yin* and *yang*, which are the negative and positive principles of the universe. Yin stands for things such as earth, moon, darkness, and femaleness; yang symbolizes heaven, sun, light, and maleness. (They are represented by that common symbol of eastern Asia, the circle with the “S” inside it as the yin and yang half perpetually turn into each other.) Therapies to maintain a balance among the elements and between yin and yang were developed over many centuries and are still widely practiced. They include several forms of massage, physiotherapy, a very complex and developed pharmacy, a system of what might be called sanitariums devoted to the therapy of breathing, and the health promotion through balanced yin–yang of the practice of *t’ai chi chuan* (*Qidong*, still seen in the parks and balconies of China in early morning). The life energy *ch’i* flows through the “meridians” of the body (these are unblocked by *t’ai chi*). The meridians touch the skin at special points that can be stimulated by heat, pressure, or needles in

the healing technique of acupuncture. Of all Chinese therapies, acupuncture is the one that has been most widely accepted in the United States, in good part because its effectiveness can be tested by scientific methods and it can be practiced even in a biomedical setting. It can clearly prevent pain, although many Western doctors are still not certain about that “stimulating the spleen” business. Although China remains the core area of the practice of Chinese medicine, it is practiced wherever Chinese people have migrated. The Nanyang Chinese (Southern Seas, or those in Southeast Asia) have a major school of Chinese traditional medicine and research center on pharmaceuticals in Singapore. Its influence has been greatest in Korea and Japan, part of China’s cultural sphere.

Biomedicine arose out of the Galenic medical tradition. Practitioners of galenic medicine in post-Renaissance times were sympathetic toward experimental physiology, which led to many important medical discoveries (such as the circulation of blood). Biomedicine advanced during the development of scientific method and inquiry; in particular, the 19th-century discoveries in bacteriology gave scientific medicine its present high status. These discoveries, collectively known as *germ theory*, focused on the idea that infectious diseases are caused by microorganisms. Two outstanding bacteriologists were Louis Pasteur (1822–1895) of France and Robert Koch (1842–1910) of Germany. Pasteur identified the organisms that cause body decomposition and fermentation of wine. He also identified those that produce anthrax and fowl cholera, and worked on vaccines for these diseases; he even developed an antirabies vaccine, although rabies is caused by a virus and viruses could not yet be seen. Koch identified the bacteria that cause wound infections, developed better techniques to identify bacteria, and discovered the germ that causes tuberculosis. Several other scientists entered the new field of bacteriology, and between 1875 and 1905 about two dozen disease agents were found. During this period there were also discoveries in serology and immunology; disease vectors such as the tsetse fly and *Anopheles* mosquito were identified; surgery and gynecology advanced; and anesthetics and antiseptic conditions became part of surgical procedures.

Specialization was spurred by the rapid accumulation of vast amounts of new information in diverse areas of medicine. Professionalization tightened ethical and educational standards within established medicine, partly in response to the proliferation of nonscientific, nonestablishment health practitioners. For example, midwifery clashed with professional obstetrical and gynecological groups in the 19th century. Such organizations as the British Medical Association (1832) and the American Medical Association (1847) were formed.

Of all the world’s major professional medical systems, biomedicine has clearly been the most widely diffused. This medical system has become synonymous with the culture of illness and health in many Western countries, and the Third World has adopted it to varying degrees. Each European power introduced biomedical practices to its colonies. The first European physicians to work in most colonial nations were too few in number to care for any but colonial and military personnel, other Europeans, and a limited number of native persons in the upper social strata (Phillips, 1990). When it was recognized that many diseases are contagious,

programs were initiated to treat larger populations that were in contact with Europeans. Despite this, health for most colonial people remained as poor as it had been in precolonial days.

Biomedicine had little real impact in these areas until after World War II, when the colonial governments began to train native assistants in Western medicine, while others went abroad to study medicine. When the colonized countries gained their independence, a native elite had acculturated to Western medicine. This elite perpetuated the Western medical hegemony. Since then biomedicine has played a major role in the epidemiological transition, one of the primary reasons for today's population growth in developing areas (see Chapter 5). In much of the developing world, Europeans tried for decades to suppress traditional medicine as quackery.

Medical Pluralism

In most parts of the world, people use more than one kind of medicine at the same time. In some countries this is officially tolerated; in others all but biomedicine are suppressed; in still others efforts have been made to integrate multiple systems. This latter approach has worked best with Chinese medicine.

In Malaysia the indigenous Malay population has its own traditional medicine (which preceded unani but, as the Malays became Muslims, was strongly influenced by it) and *bomoh* to practice it. The bomoh have drugs from the forest for various treatments, and they go into trances and interact even with a powerful tiger spirit in order to treat mental health problems. This land and language, after all, gave us "run amok." The Indian population of Malaysia (about 11%) is mostly Dravidian but also northern Indian, mostly Hindu but also Christian. It widely uses both ayurvedic and siddha medicine, especially for "women's problems." The Chinese of many dialect groups and several religions constitute about 40% of the population. Chinese pharmacies, with practitioners, are found on many streets in most towns. Practitioners have even formed a Chinese traditional medical school. Several of the students there at night work in biomedical laboratories during the day, reading tuberculosis slides and blood slides and analyzing urine and fecal parasites. They have integrated the two systems in their own personal practice, as patients have. People of all the ethnic groups and medical traditions in Malaysia seek help from both Chinese pharmacists and biomedical practitioners. Figure 11.7 shows a sign outside the main hospital in the national capital, where there are parallel waiting rooms, outpatient hours, and pharmacies for Chinese and Western medicine (although the beds and the remainder of the hospital are for biomedicine only).

Such institutionalized pluralism is not common. One difficult problem with more than one health system is that people often use both at the same time. Charles Good (1987) studied traditional medicine in Kenya, Nigeria, and elsewhere. The attraction of Western medicine was especially the power of the needle: Antibiotics were magic, vaccines worked. When other medical treatments did not bring quick response, or were not helpful for chronic pains, people repaired to their traditional practitioners. Those herbs were also bioactive, however, and the medical doctors



FIGURE 11.7. Hospital sign in Kuala Lumpur, Malaysia, giving alternative waiting and attendance times for Western and Chinese medicine, in three languages: Malay, Chinese, and English.

never knew who was taking what. Treatments were not followed up or required drug prescriptions taken; drug interactions were sometimes serious. There was a danger that people would go to traditional practitioners until they were sure that their advice was not working, and then they would present themselves at the Western clinic with sickness or cancer too far advanced for successful treatment. This is the main argument that biomedical doctors continue to raise in opposition to traditional practices, including versions of quackery today in the United States: When people pursue “miracle cures” and other alternative healing, it can cause them to postpone the more efficacious biomedical treatment. Tumors really can be surgically removed and killed with chemotherapy most of the time, but not so easily after they metastasize. Chinese practitioners in Malaysia said (in response to a survey conducted by Meade) that when their second treatment for an earache failed to cure it, they sent their patients to a Western doctor because they knew they could not help nasopharyngeal cancer—an important cause of death for the Cantonese.

The holistic health movement in the United States started in California (and North Carolina!). It diffused across the nation rapidly, adding variety as it did so. Holistic therapies include massage of various types, acupuncture (by licensed practitioners), and chiropractic (by licensed practitioners with degrees), as well as things like aromatherapy and advice on healing colors (from Chinese medicine?). The universal complaint against biomedicine was (and still is) that it was reductionist, impersonal, and inhumane. It dealt not with bodies but organs, and increasingly with molecules. People were processed, seldom listened to, and rarely touched. Tra-

ditional medical practitioners of whatever kind, it was argued, focused on individuals, took their time, listened, touched, and asked about family and life stresses.

The professionalization of standards promoted by the American Medical Association in order for individuals to be licensed to practice medicine really did eliminate a lot of “quackery” by get-rich-quick, ignorant, incompetent crooks selling “snake oil.” (Today some of them are returning via the Internet.) Practitioners of biomedicine have recognized many of the truths in the arguments for holistic medicine, however, and have started to change some things. The most prominent among these has been the response to the findings of the Dartmouth Atlas, that less can be more, and better. “Preference-sensitive care” is being promoted (especially in choosing procedures and hospital stays), as is listening to the concerns and observations of patients and their relatives. Obstetrical rooms are being redesigned to look more familiar; pain care has become a priority; and simple practices to reduce hospital infections have been renewed, including washing hands with soap and using alcohol wipes instead of antibiotics and detergents. Space has been made to relieve the stress on staff members, and so to improve their own health and the care they render. For example, break areas with water features and plants, and rooms in healing colors, are made available. Hospitals and other medical facilities have also become more open to less traditional therapeutic practices. Many now include lessons in meditation and visualization, or even humor therapy.

One new form of therapy works at several levels in a community. It’s been known for more than a decade as *pet therapy* (see Wilson & Turner, 1998). Petting a dog can lower a person’s blood pressure (the dog’s, also!), and companionship and the need to care for another living creature provide deep solace for many people. Numerous studies have shown that people who have pets recover faster from surgery and have better survival rates than those without pets (though some of those studies are poorly done—not even considering, for example, how long a person has had a pet, or whether it promotes walking and other healthy activities or just cuddles). Well-done studies in hospitals have shown, by monitoring hormones and stress chemicals, that patients who get visitors are benefited over patients with no visitors, but that visits with a dog produce better results than visits with only a human. Now puppies are taken from shelters out to visit weekly in nursing homes, which helps socialize both the puppies and the residents. People make the effort to train their dogs as Good Citizens and visitors—keeping them under control (no jumping up or dangerous exuberance); teaching them to tolerate wheelchairs, loud hydraulic noises, and crutches; and helping them to ride elevators without fear. They bathe and groom the dogs. Then they spend hours every week volunteering their time, along with the dogs, visiting in children’s hospitals, cancer wards, rehabilitation facilities, and other places they’re now welcome. Some nursing homes, following the so-called “Eden philosophy,” have resident dogs that circulate with the medicine dispatchers and hang out in the courtyard. The big problem, it seems, is jealousy among patients over feeding and visiting time!

So it takes a community to visit a patient, with canine help as a tool of our material culture. This requires a joint effort by animal shelters and their workers, dog

trainers, and volunteers who use their own valuable time out of care for others. It also requires changed institutional rules and development of new procedures. The result, however, is that people who were previously feeling alone, certainly marginalized, and even outcast are now embraced. It is part of a social change that brings people with various disabilities and infirmities into mainstream activities and places of daily life. It changes *place* to be more supportive, caring, and health-promoting.

There are several directions research can go from here. There have been a few (very few) efforts to look at the spatial dimensions of traditional medicine. There is a hierarchy of expertise and professional reputation among Chinese practitioners in Malaysia, for example, which draws clients from varying distances (catchment areas). In addition, where do practitioners of various forms of traditional medicine get their pharmacopoeia? Where is it grown, how marketed? How are new practitioners recruited and trained? Now that practitioners and clients are all moving to the city, what is happening to all these traditional spatial arrangements? By and large, however, research has not gone in that direction. Instead, what has emerged is the importance of place for healing.

Therapeutic Landscapes, Broadly Considered

Among many other things, *place* as a “living construct” is now used to provide a framework for understanding health (Kearns & Gesler, 1998; Kearns & Moon, 2002). Now that healing is again clearly recognized as mental and spiritual as well as physical, place is *where* mind and body get linked up. Probably the clearest expression of the roles of subjective feeling and meaning in getting well can be found in looking at *therapeutic landscapes*—a concept developed by Wilbert Gesler (1992, 1993, 1996, 2003, 2005), a coauthor of the first edition of this book, whose own research grew out of health service delivery. While emphasizing the physical, mental, and spiritual transformation found in places like Epidaurus and Lourdes, he also examines historical contingency and the discourses over contested space. In regard to Lourdes, for example, Gesler (1996) states: “Healing takes place at Lourdes because the site is a focus for meaning.” Pilgrims have faith foremost, and a lifetime of learning the experience of others; they “come expecting transformation, and authentic experience”—which they often have. They “understand the meaning of the rituals,” which are more intense in this place, and the feeling of group *communitas*, “the field of care” experienced at Mass. They view Lourdes as a place of peace. Gesler notes that the place has been contested over the centuries in many ways: laity versus church, church versus state, believers versus skeptics, profit and tourism versus pilgrimage and spiritual retreat, the profane versus the sacred.

The meaning of *place* in this sense is gradually developing and changing. Williams (1999) has expanded the concept to include positive places of healing, like children’s camps or new housing. Shelters for the homeless are obviously such places, as are bars and other places important for communication as “information nodes.” Andrews (2003) writes of “placing consumption” when writing of places sought out

for alternative (now called complementary) medicine by older people. They value such medicine for its holistic attitude to health, the time taken with them, the gentleness; and they value the places where such medicine is practiced for their traditions of alternative lifestyles and the open, scornless reception they receive there. They also, however, value these places for the wide variety of therapies and levels of skill available there—as “central places” of alternative medicine, so to speak.

Other geographers are writing about sites of power in “spaces of sobriety.” Hospitals are often criticized as being not places of healing, but impersonal, dehumanizing places. This criticism has changed many hospitals in ways alluded to above, as health service providers seek to enhance them as healing places. One such development has involved hospice care during the last weeks of life. Most machines are turned off and tubes removed, and people are invited in to read, talk, or pray. Often the terminally ill patients are removed from the hospital to a community place where perceptions, feelings, and meanings are those of family and community, and where larger connections are possible. One of the hospitals in Chapel Hill has created an amazing butterfly garden in an interior courtyard. It is used mostly by staff and distraught relatives as a place of relaxation and solace, but increasingly the lucid terminally ill are requesting to be wheeled into it for their last hours of life.

HYGEIA: HEALTH PROMOTION

This final section on medical geography as health promotion is short partly because this text, like all things, must come to an end. More importantly, it is a dimension of medical, or health, geography that needs its own book for today’s developments.

Geographers concerned in their research with health promotion have worked to address the space, and place, of persons with the disabilities. Whether disabilities are mental or physical, research on accessibility involves more about space than distance. Place is local, and information coming from interviews and focus groups and other qualitative methodologies is filling it with meaning. The journal *Health and Place* recently published a whole issue on the geographies of intellectual disability (Metzel & Philo, 2005). The articles in this issue deal with social poverty and service dependency, entangled geographies of social exclusion, residential instability in deprived areas, and places changing as rural areas get restructured. The few geographic studies on mental health have been mostly concerned with the old question of whether mentally ill persons (especially those with schizophrenia) are clustered in areas of old cities because the blighted conditions and treatments made them ill, or gathered there from elsewhere for the amenities and tolerance or support of such areas. There have been studies of the “not in my backyard” syndrome in regard to halfway houses, zoning laws, and such. Dear and Wolch (1987) made an impact in this area of study with their *Landscapes of Despair*. The present book deliberately has not dealt with mental health issues because of both lack of space and the relative lack of research. Looking toward the future, medical and social geographers can be engaged in studies of health promotion and the many economic, social, demographic, and locational issues and needs of mental health.

Transforming the Landscape for Health Promotion

There is a greater, larger movement abroad in developed countries as the processes of cultural ecology work on changing the cultural behavioral interactions with the habitat that have proved so harmful for human health. Our readers may remember that this has happened before. René Dubos wrote of how infectious diseases had been decreased and even eliminated without knowledge of germs by transformations of the built, social, and natural environments. Nutrition and the body's ability to fight infection and survive had been improved by transporting, storing, and (later) refrigerating food and making it available in the cities in the winter, even to the poor (who were then most of the people). Rising incomes and standards of living in the West had improved housing (away with mildew, dampness, chills, and darkness) and clothing and bathing (away with lice, fleas, and encrustations of filth); chimneys and (later) central heating had provided warmth (away with internal wood smoke, cold, and winter peaks of mortality). The "garden city" movement, as governments and industrialists faced the horrendous housing conditions and squalor of the new industrial cities, led to planned gardens (e.g., New York City's Central Park), planting of trees, and regulations on architecture (to lessen crowding and increase light and ventilation). Sanitation systems were created, and chlorinated water was provided for drinking, cooking, and hygiene. The worst of the coal dust was removed from the air. City children were sent to camp in fresh air to exercise and build healthy bodies. In New York City and elsewhere, the upper class mobilized to provide and deliver milk to the most impoverished children of new immigrants. When germs were discovered, vaccines were created as cultural buffers; behaviors that avoided germs were created; practices of hygiene and whole infrastructural systems of public health were developed. Infectious diseases were greatly reduced, controlled, or even eliminated before antibiotics. Remember how the concentrated livers of fish produced relief from bone disease and pelvic malformation even before vitamins were known? The adaptability of humans changed their environment and behavior in ways that created better states of health.

We humans are at it again. Now our main causes of death are noncommunicable, usually degenerative diseases, as well as the violence of automobile accidents, suicide, and murder. Nutrition again has gotten out of balance, now producing obesity and higher risk of disease. We live in a chemical soup, and much of it is consumed with our food. Today a major public health campaign has begun to change our diet—to get the fats, sugars, and chemicals out of it. School lunch programs are being overhauled. Farmers' markets and other local means of purchasing organically grown, fresh produce are increasing exponentially. Genetically modified crops have their own dangers, but they promise to almost remove pesticide and fungicide spraying from the environment. Energy change will be removing the particles, heavy metals, and benzene from the air, along with the carbon. Cigarette smoke is not gone, but is on its way out. People are urged to exercise, even just to walk. The consequences of all this can be seen in falling rates of heart disease and stroke—for decades the major causes of death.

This is only the beginning, we think. Urban forms and means of transportation are being changed in the New Urbanism. Some of these changes are intended to facilitate walking again, and to re-create places at a walkable, human scale. Low-rise buildings have some of their own problems, but at an expense of many millions of dollars many of the old welfare high-rises have largely been demolished and people are being moved to what are believed to be more humane housing conditions, more supportive of community. Greenways are being built in large cities and small towns—some to connect places and functions, some just to promote exercise in healthy, outdoor, more therapeutic landscapes. The question “Could smart urban design keep people fit and trim?” is abroad in the land. People are learning yoga and t'ai chi chuan. New sport activities are being invented, such as agility training and flyball for socializing and bonding between dogs and humans (individuals, neighbors, and the community at large). Such changes in the built environment, nutrition, behavior, and cultural buffers can only be a human ecology of building better states of health. There is a lot of research to be done in this area of medical/health geography.

REFERENCES

Asclepius

- Akin, J. S., Guilkey, D. K., Griffin, C. C., & Popkin, B. M. (1985). *The demand for primary health services in the Third World*. Totowa, NJ: Rowman & Littlefield.
- Anderson, G. F., & Poullier, J.-P. (1999). Health spending, access, and outcomes: Trends in industrialized countries. *Health Affairs*, 18, 178–192.
- Bannerman, R. H., Burton, J., & Ch'en, W. C. (Eds.). (1983). *Traditional medicine and health care coverage*. Geneva: World Health Organization.
- Cloutier-Fisher, D., & Joseph, A. E. (2000). Long-term care restructuring in rural Ontario. *Social Science and Medicine*, 50, 1037–1045.
- Curtis, S. (1989). *The geography of public welfare provision*. London: Routledge.
- Curto de Casas, S. I. (1994). Health care in Latin America. In D. R. Phillips & Y. Verhasselt (Eds.), *Health and development* (pp. 234–248). London: Routledge.
- Dartmouth Medical School, Center for the Evaluative Clinical Sciences. (1998). *The Dartmouth atlas of health care 1998*. Chicago: American Hospital Publishing.
- Dartmouth Medical School, Center for the Evaluative Clinical Sciences. (1999). *The Dartmouth atlas of health care 1999: The quality of medical care in the United States*. Chicago: American Hospital Publishing.
- Ebdon, D. (1985). *Statistics in geography* (2nd ed.). Oxford, UK: Blackwell.
- Eyles, J., & Woods, K. J. (1983). *The social geography of medicine and health*. London: Croom Helm.
- Feldstein, P. J. (1992). The changing structure of the health care delivery system in the United States. In M. M. Rosenthal & M. Frenkel (Eds.), *Health care systems and their patients* (pp. 21–36). Boulder, CO: Westview Press.
- Francis, R. L., Lowe, T. J., Rushton, G., & Rayco, M. G. (1999). A synthesis of aggregation methods for multi-facility location problems: Strategies for containing error. *Geographical Analysis*, 31(1), 67–87.
- Garnick, D. W., Luft, H. S., Robinson, J. C., & Tetreault, J. (1987). Appropriate measures of hospital market areas. *Health Services Research*, 22, 69–89.
- Gesler, W. M., Hartwell, S., Ricketts, T. C., & Rosenberg, M. W. (1992). Introduction. In W. M.

- Gesler & T. C. Ricketts (Eds.), *Health in rural North America*. New Brunswick, NJ: Rutgers University Press.
- Ghosh, A., Rushton, G. (Eds.). (1987). *Spatial analysis and location-allocation models*. New York: Van Nostrand Reinhold.
- Gober, P. (1997). The role of access in explaining state abortion rates. *Social Science and Medicine*, 44, 1003–1016.
- Gober, P., & Gordon, R. J. (1980). Intraurban physician location: A case study of Phoenix. *Social Science and Medicine*, 14D, 407–417.
- Godlund, S. (1961). *Population, regional hospitals, transportation facilities, and regions: Planning the location of regional hospitals in Sweden* (Lund Studies in Geography Series B, Human Geography No. 21). Lund, Sweden: Department of Geography, Royal University of Lund.
- Gould, M., & Moon, G. (2000). Problems of providing health care in British island communities. *Social Science and Medicine*, 50, 1081–1090.
- Hall, G. B., & Joseph, A. E. (1988). Group home location and host neighborhood attributes: An ecological analysis. *Professional Geographer*, 40, 297–306.
- Hart, L. G., Pirani, M. J., & Rosenblatt, R. A. (1991). Causes and consequences of rural small hospital closures from the perspectives of mayors. *Journal of Rural Health*, 7, 222–245.
- Hunter, J. M., Shannon, G. W., & Sambrook, S. L. (1986). Rings of madness: Service areas of 19th century asylums in North America. *Social Science and Medicine*, 23, 1033–1050.
- Jarvis, E. (1851–1852). On the supposed increase of insanity. *American Journal of Insanity*, 8, 333–364.
- Kearns, R. A., & Joseph, A. E. (1993). Space in its place: Developing the link in medical geography. *Social Science and Medicine*, 37, 711–717.
- McLafferty, S. L. (1989). The politics of privatization: State and local politics and the restructuring of hospitals in New York City. In J. L. Scarpaci (Ed.), *Health services privatization in industrial societies*. New Brunswick, NJ: Rutgers University Press.
- Messina, J. P., Shortridge, A. M., Groop, R. E., Varnakovidia, P., & Finn, M. J. (2006). Evaluating Michigan's community hospital access: Spatial methods for decision support. *International Journal of Health Geographics*, 5, 42–60.
- Mick, S. S., Lee, S.-Y. D., & Wodchis, W. P. (2000). Variations in geographical distribution of foreign- and domestically-trained physicians in the United States. *Social Science and Medicine*, 50, 185–202.
- Mohan, J. F. (1998). Explaining geographies of health care: A critique. *Health and Place*, 4, 113–124.
- Morrill, R. L., & Earickson, R. J. (1968). Variation in the character and use of Chicago area hospitals. *Health Services Research*, 3, 224–238.
- Morrill, R. L., & Earickson, R. J. (1969). Location efficiency in Chicago hospitals. *Health Services Research*, 4, 127–145.
- Peters, J., & Hall, G. B. (2000). Assessment of ambulance response performance using a geographic information system. *Social Science and Medicine*, 49, 1551–1566.
- Pyle, G. F. (1971). *Heart disease, cancer, and stroke in Chicago: A geographical analysis with facilities, plans for 1980* (Research Paper No. 134). Chicago: University of Chicago, Department of Geography.
- Pyle, G. F. (1979). *Applied medical geography*. New York: Wiley.
- Ricketts, T. C., & Cromartie, E. (1992). Rural primary care programs in the United States. In W. M. Gesler & T. C. Ricketts (Eds.), *Health in rural North America*. New Brunswick, NJ: Rutgers University Press.
- Ricketts, T. C., Savitz, L. A., Gesler, W. M., & Osborne, D. N. (1994). *Geographic methods for health services research*. Lanham, MD: University Press of America.
- Rosenthal, M. M., & Frenkel, M. (Eds.). (1992). *Health care systems and their patients: An international perspective*. Boulder, CO: Westview Press.
- Scarpaci, J. L. (1989). Dismantling public health services in authoritarian Chile. In J. L. Scarpaci

- (Ed.), *Health services privatization in industrial societies* (pp. 219–244). New Brunswick, NJ: Rutgers University Press.
- Scarpaci, J. L. (1990). Medical care, welfare state and deindustrialization in the southern cone. *Environment and Planning D*, 8, 191–209.
- Shannon, G. W., Bashshur, R. L., & Spurlock, C. W. (1978). The search for medical care: An exploration of urban black behavior. *International Journal of Health Services*, 8, 519–530.
- Shannon, G. W., & Dever, G. E. A. (1974). *Health care delivery: Spatial perspectives*. New York: McGraw-Hill.
- Shannon, G. W., & Spurlock, C. W. (1976). Urban ecological containers, environmental risk cells, and the use of medical services. *Economic Geography*, 52, 171–180.
- Smith, C. J. (1998). Modernization and health care in contemporary China. *Health and Place*, 4, 125–139.
- Smith, C. J., & Fan, D. (1995). Health, wealth, and inequality in the Chinese city. *Health and Place*, 1, 167–177.
- Wang, F., McLafferty, S., Escamilla, V., & Luo, L. (2008). Late-stage breast cancer diagnosis and health care access in Illinois. *Professional Geographer*, 60(1), 54–69.
- Wang, F., & Wei, L. (2005). Assessing spatial and nonspatial factors for healthcare access: Towards an integrated approach to defining health professional shortage areas. *Health and Place*, 11, 131–146.
- Wennberg, J. E., Fisher, E. S., Goodner, D. C., & Skinner, J. S. (2008). The Dartmouth Institute for Health Policy and Clinical Practice, *Tracking the cases of patients with severe chronic illness*. Chicago: American Hospital Publishing.
- Yasenovskiy, V., & Hodgson, J. (2007). Hierarchical location–allocation with spatial choice interaction modeling. *Annals of the Association of American Geographers*, 97(3), 496–511.
- Zhao, Z. (2006). Income inequality, unequal health care access, and mortality in China. *Population and Development Review*, 32(3), 461–483.

Asclepius Unrobed and Hygeia

- Andrews, G. J. (2002). Towards a more place-sensitive nursing research: An invitation to medical and health geography. *Nursing Inquiry*, 9(4), 221–238.
- Andrews, G. J. (2003). Placing the consumption of private complementary medicine: Everyday geographies of older people's use. *Health and Place*, 9, 337–349.
- Dear, M. J., & Wolch, J. R. (1987). *Landscapes of despair*. Princeton, NJ: Princeton University Press.
- Gesler, W. M. (1984). *Health care in developing countries*. Washington, DC: Association of American Geographers.
- Gesler, W. M. (1991). *The cultural geography of health care*. Pittsburgh, PA: University of Pittsburgh Press.
- Gesler, W. M. (1992). Therapeutic landscapes: Medical issues in the light of the new cultural geography. *Social Science and Medicine*, 34(7), 735–746.
- Gesler, W. M. (1993). Therapeutic landscapes: Theory and a case study of Epidaurous, Greece. *Environment and Planning D*, 11, 171–189.
- Gesler, W. M. (1996). Lourdes: Healing in a place of pilgrimage. *Health and Place*, 2(2), 95–105.
- Gesler, W. M. (2003). *Healing places*. Lanham, MD: Rowman & Littlefield.
- Gesler, W. M. (2005). Therapeutic landscapes: An evolving theme. *Health and Place*, 11, 295–297.
- Good, C. M. (1987). *Ethnomedical systems in Africa: Patterns of traditional medicine in rural and urban Kenya*. New York: Guilford Press.
- Kearns, R. A., & Gesler, W. M. (Eds.). (1998). *Putting health into place: Landscape, identity and well-being*. Syracuse, NY: Syracuse University Press.
- Kearns, R. A., & Moon, G. (2002). From medical to health geography: Novelty, place and theory after a decade of change. *Progress in Human Geography*, 26(5), 605–625.
- Leslie, C. (Ed.). (1976). *Asian medical systems: A comparative study*. Berkeley: University of California Press.

- Metzel, D. S., & Philo, C. (Eds.). (2005). Geographies of intellectual disability [Special section]. *Health and Place*, 11(2).
- Phillips, D. R. (1990). *Health and health care in the Third World*. New York: Longman Scientific & Technical.
- Ramesh, A., & Hyma, B. (1981). Traditional Indian medicine in practice in an Indian metropolitan city. *Social Science and Medicine*, 15D, 69–81.
- Williams, A. (Ed.). (1999). *Therapeutic landscapes: The dynamic between wellness and place*. Lanham, MD: University Press of America.
- Wilson, C. C., & Turner, D. C. (Eds.). (1998). *Companion animals in human health*. Thousand Oaks, CA: Sage.

FURTHER READING

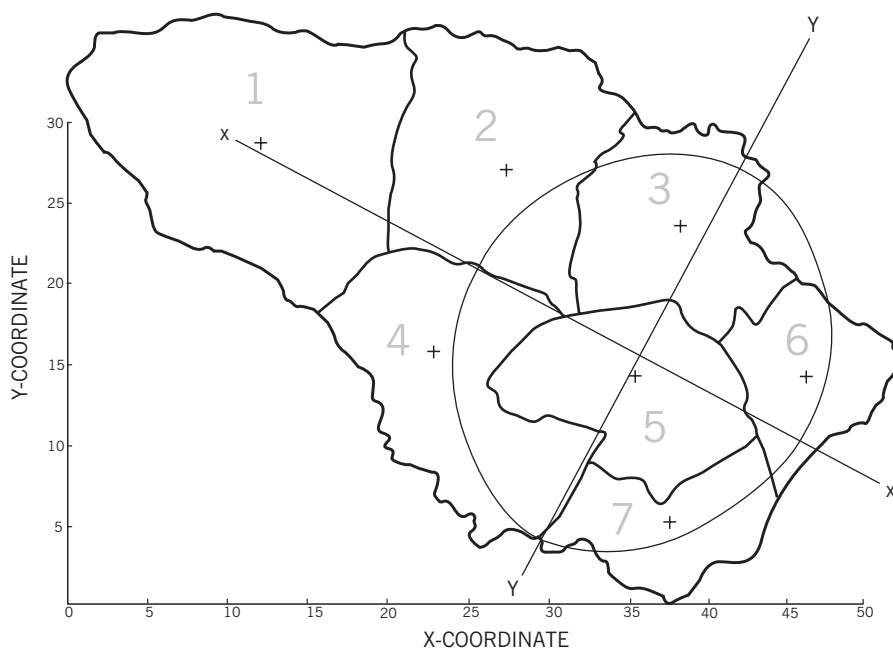
- Buor, D. (2003). Analysing the primacy of distance in the utilization of health services in the Ahafo-Ano South district, Ghana. *International Journal of Health Planning and Management*, 18, 293–311.
- Buor, D. (2004). Gender and the utilization of health services in the Ashanti Region, Ghana. *Health Policy*, 69, 375–388.
- Eyles, J. (1987). *The geography of the National Health*. London: Croom Helm.
- Fulder, S. (1996). *The handbook of alternative and complementary medicine*. Oxford, UK: Oxford University Press.
- Gatrell, A. C. (2002). *Geographies of health: An introduction*. Oxford, UK: Blackwell.
- Joseph, A. E., & Phillips, D. R. (1984). *Accessibility and utilization*. London: Harper & Row.
- Laditka, J. N. (2004). Physician supply, physician diversity, and outcomes of primary health care for older persons in the United States. *Health and Place*, 10, 231–244.
- Loytonen, M., & Gatrell, A. (Eds.). (1998). *GIS and health*. London: Taylor & Francis.
- Madge, C. (1998). Therapeutic landscapes of the Jola, the Gambia, West Africa. *Health and Place*, 4, 293–312.
- Mayhew, L. D. (1986). *Urban hospital location*. London: Allen & Unwin.
- Meade, M. S. (1986). Geographic analysis of disease and care. *Annual Review of Public Health*, 7, 313–335.
- North Carolina Rural Health Program website: www.shepscenter.unc.edu/research_programs/Rural_Program/rhp.html
- Oppong, J. R. (1996). Accommodating the rainy season in Third World location-allocation applications. *Socio-Economic Planning Sciences*, 30(2), 121–137.
- Oppong, J. R., & Hodgson, M. J. (1994). Spatial accessibility to health care facilities in Suhum District, Ghana. *Professional Geographer*, 46(2), 199–209.
- Scarpaci, J. L. (Ed.). (1989). *Health services privatization in industrial societies*. New Brunswick, NJ: Rutgers University Press.
- Smith, C. J. (Ed.). (1988). *Public problems: The management of urban distress*. New York: Guilford Press.
- Stribling, D. E. (1983). *Holistic health: A changing paradigm for cultural geography*. Unpublished doctoral dissertation, University of North Carolina.
- Techatraisak, B. (1985). *Traditional medical practitioners in Bangkok: A geographical analysis*. Unpublished doctoral dissertation, University of North Carolina.
- Thomas, R. (1992). *Geomedical systems*. London: Routledge.

VIGNETTE 11.1. Application of Spatial Statistics to Health Care Delivery

In central place theory, the center of a population is the location of a good or service for consumers within some predefined trade area. The trade area may be defined, if there is only one supplier, by how far the average consumer is willing or able to travel to obtain the good or service. If there are multiple suppliers, the consumers must be divided up among the suppliers, and the trade areas will decrease in size as the number of suppliers increases. Suppose there is a large enough population to support a suburban medical center with about 400 beds and a comprehensive clinic. The most equitable location for that center will be one in which the average travel time from the center to the furthest households is the same in all directions.

MEAN CENTER

Assuming that land values, zoning, and other factors are satisfactory, where should the medical center be located to meet the objective of centrality? To illustrate, we arbitrarily chose a county located in the eastern United States, which has been divided into its seven census subdivisions (see Vignette 11.1, Figure 1). A coordinate system is superimposed on the map. The spacing on the horizontal and vertical axes produces a 1-mile grid. We make a simplifying assumption that the population of each subdivision is rep-



VIGNETTE 11.1, FIGURE 1. A typical U.S. county with seven census subdivisions.

VIGNETTE 11.1, TABLE 1. County Coordinates and Populations

County subdivision	X coordinate	Y coordinate	Population
1	12	29	11,000
2	27	27	10,400
3	35	22	17,284
4	23	16	7,037
5	36.5	14.5	7,769
6	47	16	7,212
7	42	7	34,800

resented by a single *control point* in each area, as shown on the map. The objective is to find the point within the county that lies in the center of the population. This is known as the *bivariate weighted mean center*.

The population of this county is not evenly distributed within its subdivisions, as may be seen in Vignette 11.1, Table 1. Because most of the population is concentrated in subdivisions 3, 5, 6, and 7, the bivariate mean should be “pulled” toward this population mass. Note that the target population may consist of particular age groups or may be gender-specific. In either case, only the number or proportion of people in the target category is used to weight each control point. The weighted mean center of this population can now be found by multiplying the X and Y coordinates for each control point by the population weight associated with that point. The mean of the weighted X coordinates and the mean of the weighted Y coordinates define the position of the weighted mean center. The equations for the weighted mean center are

$$\bar{X}_w = \frac{\sum_w X}{\sum_w} \text{ and } \bar{Y}_w = \frac{\sum_w Y}{\sum_w}$$

where X and Y are the coordinates of the points, and *w* is the population weight that is to be multiplied by the X and Y coordinates, respectively. As each *wX* and *wY* operation is completed, they are accumulated to yield the sums that are divided by the sum of the population weights. In our medical center location example, the coordinates of the mean weighted center are *X* = 36.13 and *Y* = 16.56, as shown in Vignette 11.1, Figure 1.

ELLIPTICAL CELL

It is also instructive to know how the population is dispersed around the bivariate mean. In the previous example, the district control points are dispersed along a northwest–southeast axis; however, the bulk of the population is concentrated in the southeastern end of the county. To ascertain the distribution of our population, we may employ a *standard deviational ellipse*. This measure, as the name implies, summarizes the dispersion of our population in terms of an ellipse, which is balanced on the fulcrum of the mean center, with its long axis in the direction of the maximum dispersion and its short axis in the direction of minimum dispersion. Three quantities emerge from the computation of the ellipse: the lengths of the long and short axes, and the directional

VIGNETTE 11.1, TABLE 2. Preliminary Computations for Mean Center and Standard Deviations

County subdivision	X coordinate	Y coordinate	$X - \bar{X}$	$Y - \bar{Y}$	x^2	y^2	xy
1	12	29	-24.13	12.44	582.3	154.8	-300.2
2	27	27	-9.13	10.44	83.4	109	-95.3
3	35	22	-1.13	5.44	1.3	29.6	-6.1
4	23	16	-13.13	-.56	172.4	.31	7.4
5	36.5	14.5	.37	-2.06	.14	4.2	-.76
6	47	16	10.87	-.56	118.2	.31	-6.1
7	42	7	5.87	-9.56	34.5	91.4	-56.1
					992.2	389.6	-457.2

orientation of the ellipse. It will soon become apparent that this is a complicated process best suited to a computer.

It is first necessary to transpose the coordinate system in terms of the bivariate mean. This involves subtracting the mean from each of the original X and Y coordinates (Vignette 11.1, Table 2). We may then compute the angle of rotation of the original axes as it is oriented to the pattern of control points. This is done with the following trigonometric equation:

$$\tan v = \frac{(\sum x^2 - \sum y^2) + \sqrt{(\sum x^2 - 4\sum xy)^2}}{2(\sum xy)}$$

where $\tan v$ is the tangent of the angle of rotation, and x and y represent the quantities $(X - \bar{X})$ and $(Y - \bar{Y})$, respectively. The first two terms in the numerator of the equation involve summing the squared deviations around the bivariate mean, but $\sum xy$ involves summing the cross-product of the deviations around the mean. Vignette 11.2, Table 2 shows the individual calculations and the sums just mentioned. Angle v is measured clockwise from the original Y -axis, like a bearing from north on a map. The angle of rotation equation can produce a negative value of $\tan v$. If this occurs, ignore the sign when looking up the value in a table of tangents. However, the angle found from the tables must be subtracted from 90° in order to give the correct value of v . In this case, $v = 28^\circ$ for the Y -axis. The X -axis is drawn at a 90° angle to the Y -axis.

The next step is to calculate standard deviations about each of the transposed axes that will define the shape of the ellipse. Beginning with the X -axis,

$$\sigma_x = \frac{(\sum x'^2) \cos^2 v - 2\sum(x'y') \sin v \cos v + (\sum y'^2) \sin^2 v}{n}$$

where $\cos^2 v$ is the square of the cosine of the angle. Similarly, $\sin^2 v$ is the square of the sine of the angle. The equation for the Y -axis is

$$\sigma_y = \frac{(\sum x'^2) \sin^2 v + 2(\sum x' y') \sin v \cos v + (\sum y'^2) \cos^2 v}{n}$$

In this example, $\sigma_x = 11.4$ and $\sigma_y = 13.3$. The resultant standard deviational ellipse is fitted to the county subdivision pattern in Vignette 11.1, Figure 1. Since σ refers to distance away from the mean, the length of the ellipse along its two axes is $2\sigma_x$ and $2\sigma_y$. For a more complete exposition on the various descriptive measures for spatial distributions, the reader is directed to Ebdon (1985, Ch. 7).

POPULATION POTENTIAL

The efficient location of a health care facility should take into account the *relative positions* of patients, care providers, and institutions. A technique based on the precept that population or activity at one location influences spatial interaction is called *population potential*. Social and economic interaction is substantially influenced by the distance between people and the locations of their activity centers. The notion that a medical center should be placed in the center of its service population is based on the fact that the attraction of a service declines with distance. Furthermore, socially and economically homogeneous communities with larger populations have a greater impact on service centers than do communities with smaller populations. The basic concept of population potential is that the magnitude of spatial interaction is directly influenced by the size of the populations at their origins. In this example, we assume that the seven district control points are the origins of the populations living within their borders. Symbolically, the population potential, V , at any control point i is defined as

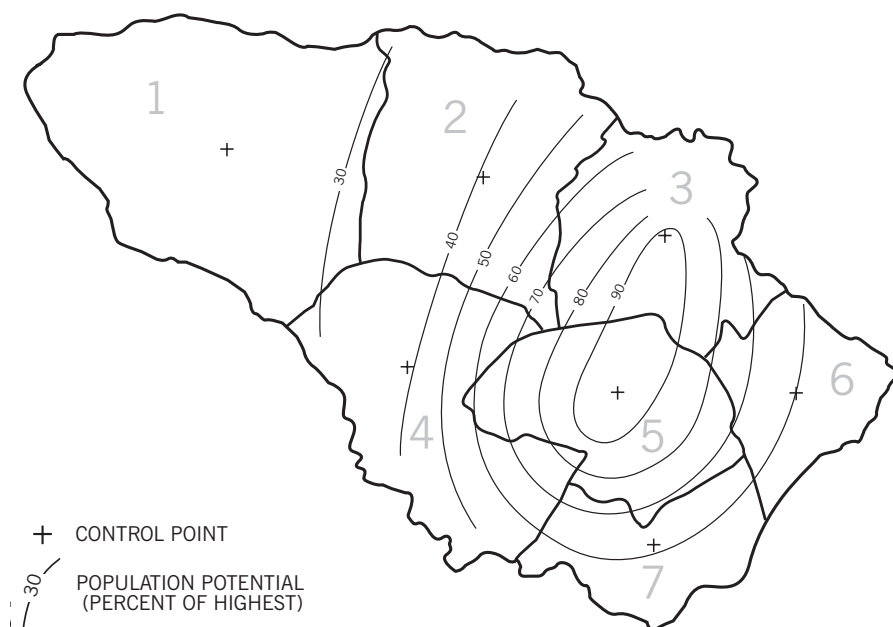
$$V_i = \sum \frac{p_j}{d_{ij}}$$

where there are n origins influencing location i whose populations are given by P_j and whose distances from i are given by d_{ij} .

Applying this to the medical center problem, we compute for each of the seven districts ($j = 1$ to 7) the ratios p_j/d_{ij} and sum them to yield the potential at i . This result is measured in persons per mile. Table 3 shows the potential computations for the seven

VIGNETTE 11.1, TABLE 3. Population Potentials for County Districts

District	X coordinate	Y coordinate	Population	Potential	% highest
1	12	29	11,000	120	24
2	27	27	10,400	206	41
3	35	22	54,800	473	95
4	23	16	12,270	186	37
5	36.5	14.5	90,580	497	100
6	47	16	34,800	300	60
7	42	7	31,750	310	62



VIGNETTE 11.1, FIGURE 2. Isolines of population potential around the population center of a county.

county districts. The column labeled “% highest” is the ratio of each potential ($\times 100$) to the maximum potential, which belongs to district 5 in our example.

These potentials were obtained with the help of a computer program written by Robert J. Earickson, Meade’s coauthor for the second edition of this book. Vignette 11.1, Figure 2 shows an isoline map of the pattern of population potential (in terms of % highest). The map identifies the population core area of the county. This happens to be the site of the county’s major retail center, which was originally constructed in the 1960s. Population potential tells us nothing about the particular characteristics of specific locations. It is a macro-spatial concept, just as are mean center and standard deviational ellipse (Taylor, 1977). The potential at any point on the map reflects only the impact of control point populations and their relative distances to other control points. In effect, it tells us how close (relatively) we are to large population clusters. We could substitute income, disease incidence, ethnic population, or any other characteristic for population. Potential maps of these variables have obvious utility to business or government decision makers.

CHAPTER 12

Scale, Spatial Analysis, and Geographic Visualization

This chapter builds on concepts introduced in Chapter 3, including the definitions of a *geographic information system* (GIS) and descriptions of data structures. It describes some of the cartographic and spatial data integration capabilities of a GIS. GIS has also revolutionized spatial analysis, and this chapter begins to describe some of the spatial-analytical capabilities of a GIS. It offers this overview of spatial analysis within a theoretical context of basic geographic concepts, including *scale*. The purpose of this chapter is to increase awareness of the ways distributions can be analyzed spatially across scale, the ways queries and findings can be portrayed and presented, and the potential for error and abuse that lurks in computerized analysis. There is not enough space to review all the ways spatial analysis has been applied to measuring disease distributions, and we could not do such a review as well as the tour de force of Cliff and Haggett (1988). Nor do we attempt to explain how to make maps or conceptualize and manage a GIS. Those specialized kinds of knowledge have many sequential courses and texts of their own (e.g., Cromley & McLafferty, 2002; Longley, Goodchild, Maguire, & Rhind, 2005). Some of our readers will learn new questions they can ask and ways that these can be addressed. Whether or not we do research, we all see some maps and research results in journals, books, and papers. These are proliferating as the use of GIS spreads in studies of health and disease (see “Further Reading” for a few examples of empirical papers that use GIS and spatial analysis). People dealing with health and disease matters today need to become more spatially sophisticated thinkers, and this chapter is intended to help them do that.

SOME ISSUES OF SCALE

We can think most simply of *scale* as the amount of area being considered, while realizing that the area under consideration is at once a component of a larger area and is itself made up of smaller areas: Space is a continuum. The word *scale* itself originates from the Latin word for *ladder*, which was both a means of ascent and descent and a set of marks used for measuring. A scale is the proportion a map, or other model, bears to the thing it represents.

A scale can be a ratio between the dimensions of a representation and those of the object being represented, as in “1 inch to a mile.” Its technical use in geography, through cartography, is sometimes at variance with common usage. This can be a source of confusion. If 1 centimeter on a map represents 100,000 centimeters on the earth’s surface, the scale is 1:100,000; if 1 centimeter represents 1,000 centimeters, the scale is 1:1,000. Because the former fraction is a much smaller number than the latter, it is referred to as a “small scale” and the latter as a “large scale.” Note, however, that the small-scale map covers a much larger area, and the large-scale map covers a smaller area. Thus a map of the United States is small-scale, and a map of a house’s interior is large-scale. To avoid confusion, it is not uncommon today to speak of “macro” and “micro” studies, referring to the area covered in terms more familiar to common usage (large areas or populations being macro, and small populations or areas being micro).

Things vary over space. The amount of that variation that can be captured and studied is necessarily dependent on the proportion being represented (i.e., on the scale of analysis). This causes both the fascination and the frustration—the inherent challenge—of spatial analysis.

THE ECOLOGICAL FALLACY

The infamous *ecological fallacy* has bedeviled geographic comparisons for a long time. It crops up when statistics are compared across scales. We cannot use a state-level risk factor to predict what will happen to specific individuals, and we cannot interview friends and neighbors to predict national opinion. Even given bona fide cause and effect, measurable statistical association varies differently at different scales. One measure of statistical association is called the *chi-square test*, which we describe in Vignette 12.1, using a medical geography example. Sociologists first confronted the ecological fallacy by aggregating population by trying to analyze its groups. Geographers also confront it by aggregating area to different scales and trying to analyze zones (the *modifiable areal unit* problem).

Variance changes with scale. If information is based on individual questionnaires, we are aware of the diversity of response. When these answers are aggregated to county level, *central tendency leveling* takes place. There will be less variation among 100 counties in a state than there was among the millions of individuals who contributed the information. There is less variance among state economic areas than among counties, and less among states than among state economic areas. These

changes in variance have little to do with cause and effect, and a lot to do with the way we can measure and generalize our findings.

Cleek (1979) and others have pointed out that correlations are especially affected by changes in aggregation. Data are often aggregated by an independent variable. This inflates correlation coefficients, but usually does not affect regression coefficients (see Vignette 12.2 for a description of the statistical technique called *regression analysis*.) For example, as individuals are aggregated into the county in which they live, socioeconomic measures are generalized. Several adjacent counties are likely to be more similar with regard to such generalized characteristics as median income than are the thousands of individuals who compose them. In Cleek's example, the coefficient of variation (standard deviation/mean) for mortality rates from leukemia was 7.0% when states were compared at the national level and 20.9% when counties were compared within the state of Wisconsin. Similarly, the coefficient of variation of mortality rates for cancer of the nasopharynx is 188.9% for Wisconsin at the county level and 24.4% for the United States at the state level. Colon cancer, however, has higher state-level variation (25.8%) than county-level variation (17.3%). The message, Cleek says, is clear: Report regression coefficients, and treat correlation coefficients with care.

What this all means is that patterns of association are different at different scales of analysis, and it is an error to take an association that is true at one scale and infer that it will be true at any other scale. On the other hand, the change in association patterns can itself sometimes be used as an etiological clue to separate the expression of different causal factors. In one case there may be great between-state variation but little within-state, between-county variation; in another there may be great local variation but little state-level variation. These patterns can be used to address different possible causes that relate to individual behavior, broader patterns of water sources or occupations, or still broader regional patterns of atmospheric pollution or population migration.

SCALE OF ANALYSIS AND UNITS OF OBSERVATION

Among the first things a researcher must decide are the scale at which a question will be addressed and the units of observation that will be used. Individuals may be a source of information, but their information must be merged and aggregated. Socially, people may be grouped into families, play or work groups, ethnic groups, and so on. Spatially, people may be aggregated into households (which occupy a living space), census blocks, census tracts, cities or minor civil divisions, counties, state economic areas (which are merely groupings of counties that create a larger areal aggregation), states, regions, and so on. Some of the theoretical issues involved in analyzing area-level distributions are discussed in Chapter 9. Such units of observation or data collection may be split up or put together to make other units. This may be done either deliberately, as when regions are defined for a specific research purpose, or inadvertently, as when census tract boundaries are altered from one decennial census to the next. Regions can be defined that cut across borders, raising an

issue that is sometimes referred to as *modifiable units*. Changes in boundaries often alter the level of analysis, make comparisons difficult, and affect the interpretation of results, as illustrated in the following examples.

The first example shows what might happen when one is looking for the cause of a disease (Figure 12.1). Suppose an industrial chemical is being dumped into a river. This causes people who live along the river and drink its water to contract a certain disease. This disease can also be contracted by breathing car emissions or by breathing the chemical put into the air by a smelter. Because people who live near the smelter, along the river, or in congested parts of this town tend to have similar socioeconomic and demographic characteristics, ethnicity, income, and race may be associated with this disease. If one were to study this disease of unknown etiology at different scales (i.e., using different geographic units of analysis), different associations would be identified.

Study 1 focuses on a minor civil division where people breathe the air of a smelter (a) and car emissions (b), and so a and b are found to be associated with the disease. Study 2 examines another minor civil division where people drink water from a stream (c) and breathe car emissions (b). Study 3 is carried out in a rather rural minor civil division wherein people work in a smelter and breathe its air (a), and are German in ethnic origin (d). In study 4 the rural minor civil division has ethnic German people (d) who drink the stream water (c). These associations are found at the micro scale, but they are puzzling because there is so little consistency. Suppose that study 5 looks at county data, combining areas 1 and 2, and finds that car emissions (b) are important. Study 6 also examines county data, combining

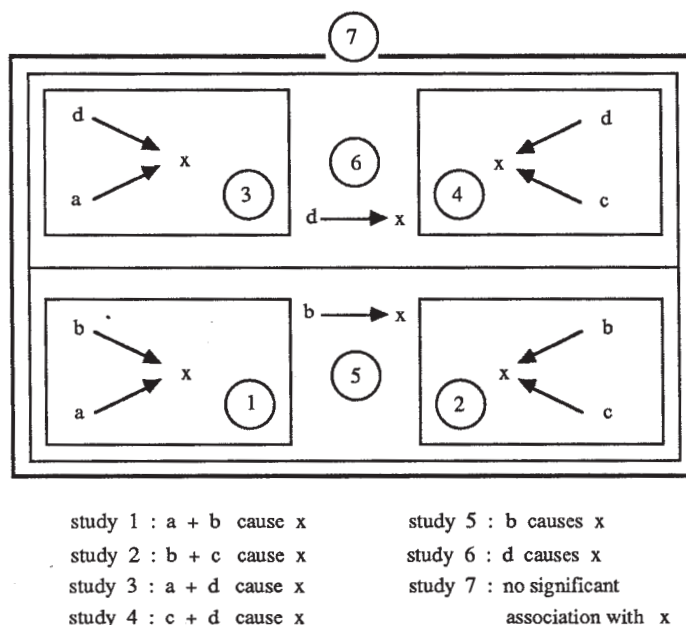


FIGURE 12.1. Disease causation factors at different scales of analysis.

areas 3 and 4, and finds that being German (d) is important. Finally, study 7 uses state data and combines all these areas but finds no important associations.

The second example of modifiable units deals with health care delivery. Suppose a health systems agency consists of 10 counties, as shown in Figure 12.2. Table 12.1 shows the number of physicians and total population for each county. Planners in the agency may wish to organize their counties in different ways, and then compare groups of counties to determine whether they are being relatively well or poorly served by physicians. Suppose they compare a “northern tier” (counties A–E) with a “southern tier” (counties F–J). The northern tier has an overall physician-to-population ratio of 128.8, and the southern tier one of 171.2. Now suppose that there is an interest in comparing the standard metropolitan statistical area (SMSA) (counties D, F, and G) with the other, more rural counties. The SMSA has a physician-to-population ratio of 187.5, and the other seven counties a ratio of 84.9. Choosing the way of grouping the counties clearly changes the assessment of the resource situation. Urban–rural investigations (see the discussion of air pollution in Chapter 7) confront this issue constantly. Do we use the SMSA, urban area, political city, central county, or minor civil division in the analysis? The results can vary significantly. Whatever scale of unit is used for analysis, researchers face this problem: Even census tracts, for example, have spatial variation within them, such that values may be more similar between two blocks across the tract border (often a street) than across the census tract unit of observation. A technique called the *spatial filter/Monte Carlo approach* can be helpful with these problems (Openshaw, Charlton, Wymer, & Craft, 1987; Rushton, 1998).

Because these units of observation exist in spatial relationships with each other (adjacency, distance, etc.), another issue that needs to be addressed is *spatial autocorrelation* (see Vignette 12.3). That is, census tracts or counties when used as units of observation may not be truly independent of each other, as most statistical procedures require (e.g., regression analysis). Adjacent territories usually influence each other, as the property value of one house affects that of property adjacent. Time series analysis also has to confront this problem, as the value of something in one year usually affects the value in the next year, so that using sequential years as units of observation can be statistically invalid because of autocorrelation. Because

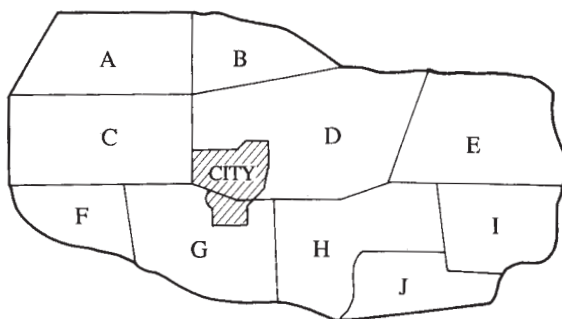


FIGURE 12.2. Ten counties within a health systems agency jurisdiction.

**TABLE 12.1. Health System Agency County
Physician-to-Population Ratios**

County	Number of physicians	Population	Physician-to-population ratio (per 100,000)
<i>A</i>	12	19,122	62.8
<i>B</i>	15	25,639	58.5
<i>C</i>	2	4,553	43.9
<i>D</i>	126	62,798	200.6
<i>E</i>	10	16,014	62.4
<i>F</i>	40	29,528	135.5
<i>G</i>	74	35,670	207.5
<i>H</i>	22	9,608	229.0
<i>I</i>	5	5,926	84.4
<i>J</i>	8	6,321	126.6
Totals	314	215,179	145.9

more scientists worry about time than about space, statistical procedures have been developed to “remove” or adjust for the autocorrelation in time periods, and geographers usually adapt these techniques to their study in space.

These issues of scale dependency are complex but critical to spatial analysis. What is the range of scales at which relationships are alike, and at what scale do things appear and act differently? That is, at what scale does spatial autocorrelation break into randomness and scale independence occur? There are various specialized techniques to determine this, using such things as semivariograms to analyze scale relationships, and fractal analysis to describe the nature and degree of the spatial relationships. Vignette 12.3 presents one example of the way units can be “joined” and tested for spatial autocorrelations, and its usage is discussed in Chapter 10 (on disease diffusion). The very extent of the spatial associations, whether limited just to adjacent units or extending over broad regions, can be used to illuminate etiology. Just as change in variability through scale can be usefully or fallaciously used, so the phenomenon of spatial autocorrelation can be used.

SOURCES FOR SPATIAL DISEASE DATA

There are many different sources of GIS data for health studies. The majority of studies use health data that have some geographic location associated with each observation, and they are georeferenced after the fact. For instance, when you are hospitalized or see a doctor, the hospital or doctor’s office usually has information about where you live. Studies sometimes have address-level data that permit individuals’ houses to be located. Chapter 3 describes how houses can be located by using U.S. Census Bureau TIGER GIS files. The result of an address-matched database of a health outcome would be the household point location of all people with a particular disease. Data at the individual level are not available, however, except

under strict research guidelines governed by university institutional review boards. Individual household locations can also be identified during fieldwork by using a global positioning system (GPS) receiver that measures the X and Y coordinates (e.g., latitude, longitude). Figure 12.3a shows an inexpensive handheld GPS that can be used to identify point locations to the nearest 10 meters. More sophisticated and expensive receivers are accurate to the nearest meter, decimeter, or even centimeter. Figure 12.3b shows a much more expensive model that is accurate to the nearest decimeter. For instance, Figure 12.4a displays the locations of households in a study area of rural Bangladesh called Matlab, where one of us (Michael E. Emch) has conducted extensive medical geography studies. Emch and colleagues collected information about people who contracted cholera in each household, and they could therefore map the location of each cholera case, as shown in Figure 12.4b. Note that the locations of cases in Figure 12.4b are not real, as there are ethical problems associated with identifying the locations of individual households and thus revealing exact locations of where people live who contracted a disease.

VISUALIZING AND SUMMARIZING DISEASE DISTRIBUTIONS

Often when disease distributions are reported, they are aggregated up to an area level. In Matlab, Bangladesh, for example, cholera disease distributions can be shown at the village level. Figure 12.5a shows the total number of cases of cholera by village in Matlab in a choropleth map (see Chapter 3 for a description of choropleth maps and classification methods of defining numerical categories). Figure 12.5b shows the number of cases per population by village.

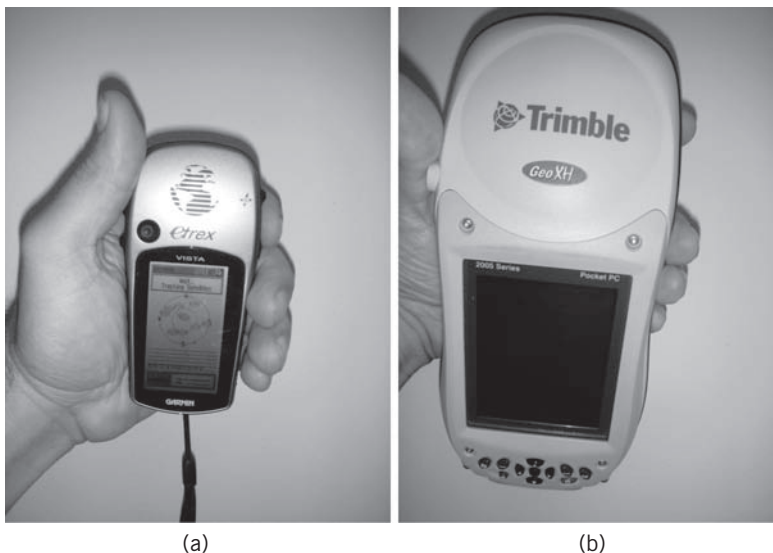


FIGURE 12.3. Handheld GPS receivers.

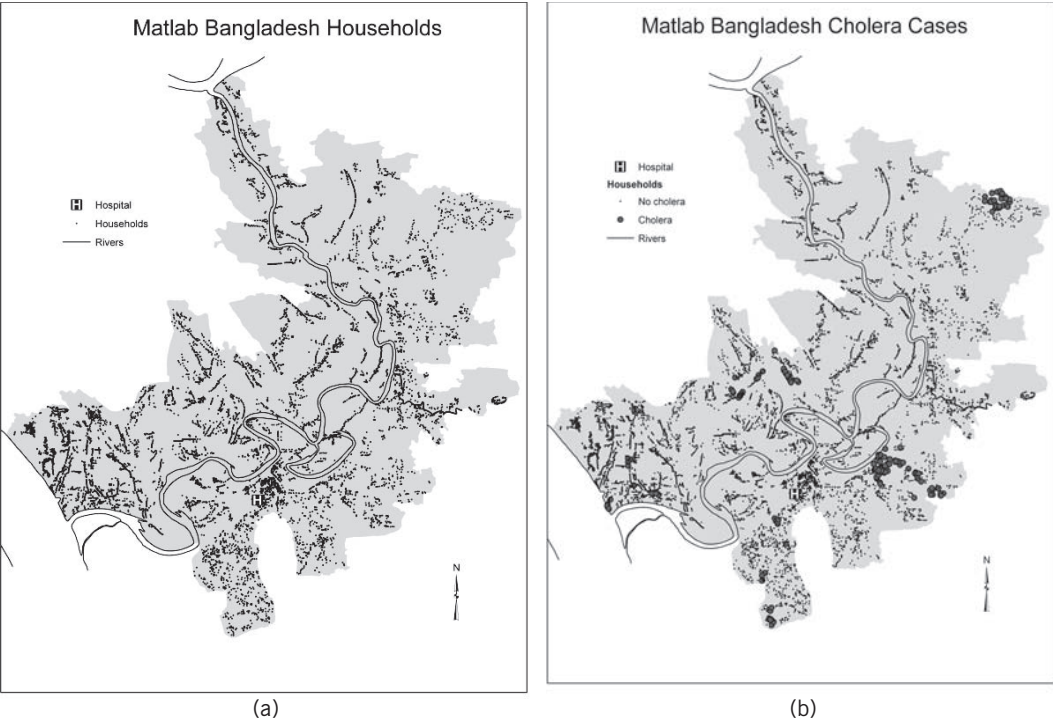


FIGURE 12.4. Household-level maps. (a) Household points for which disease data can be mapped. (b) Cholera cases.

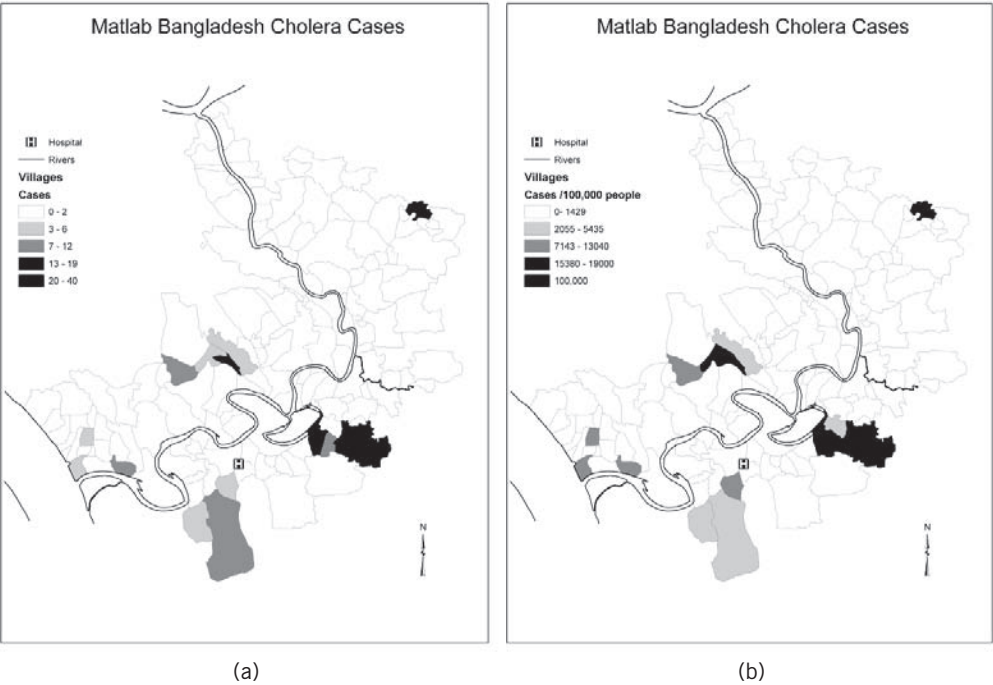


FIGURE 12.5. Area-level maps. (a) Cholera cases by village. (b) Cholera cases per population by village.

Disease distributions can also be summarized by using summary statistics such as *mean centers*. A mean center takes a distribution of cases such as the cholera points shown in Figure 12.4b; the average *X* coordinate (i.e., the longitude or horizontal projected coordinate) and the average *Y* coordinate (i.e., latitude or vertical projected coordinate) are then calculated and represent the center of the epidemic. Figure 12.6a depicts the mean center of cholera cases for the distribution shown in Figure 12.4b. The map shows the original distribution of cholera cases as well. We can also measure whether the disease distribution exhibits a directional trend—in other words, whether cases are farther from the mean center in one direction than in another direction. Figure 12.6b shows the *standard deviational ellipse*, which summarizes the directional distribution of cholera in Matlab within one standard deviation of the mean.

Another tool that can be used to display distributions is the *raster density calculation*. The user starts with a vector point file and sums the points that fall within the search area; that number is then divided by the search area size to get each cell's density value. The search area represents the raster resolution and is the scale at which the map will be produced, and the cell size is the area that each raster cell represents in ground units. Figure 12.7 is a raster GIS map of cholera density in three classes. The map shows the hot spots of cholera within the study area. Notice that the map looks boxy or “pixelated” around the edges. That is because each cell represents 100 meters on the ground. The density calculation was calculated with a

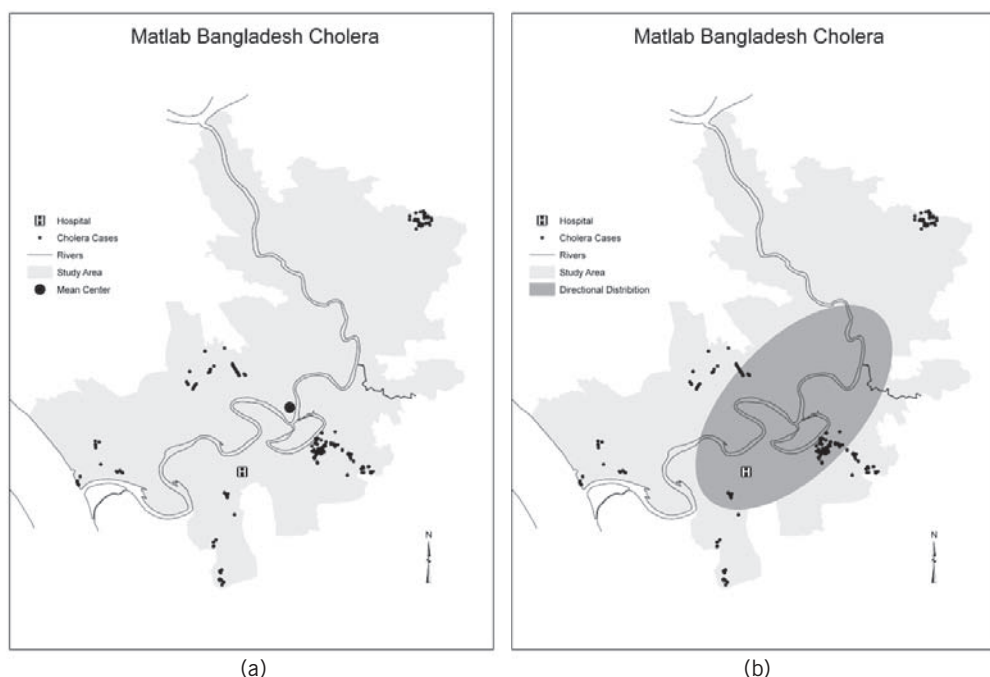


FIGURE 12.6. Spatial summary statistics. (a) Mean center of cholera epidemic in Matlab, Bangladesh. (b) Standard deviational ellipse of cholera.

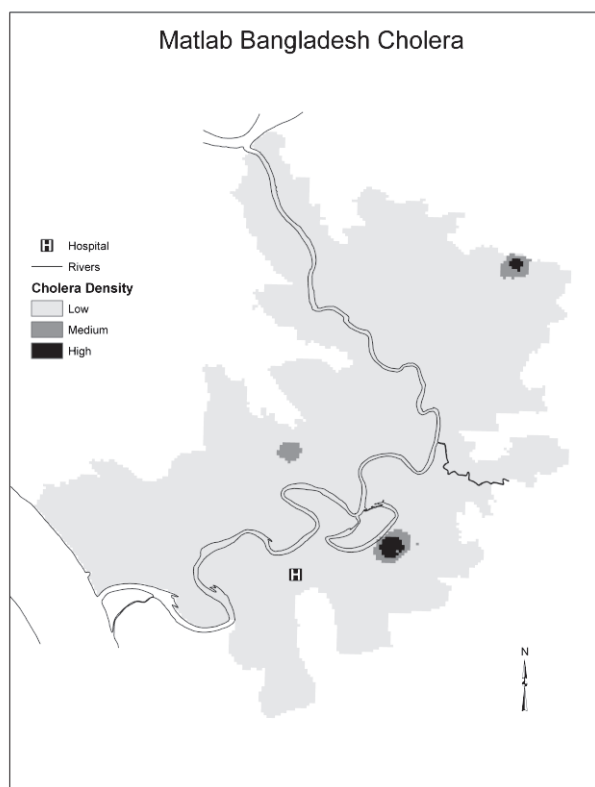


FIGURE 12.7. Cholera density in Matlab, Bangladesh.

500-meter search radius; thus the software counted the number of cases of cholera within 500 meters of each cell, and each cell has a value in the GIS database that is the number of cholera cases in the 500 meters around it. Then all of the pixels in the raster GIS were classified into three categories, which were given the subjective names of *low*, *medium*, and *high*.

Spatial interpolation uses existing observations to estimate values at unsampled sites. Spatial interpolation is based on Tobler's law of geography, which is that things that are close together in space are more likely to have similar values than things that are far apart. Spatial interpolation is a common tool in most commercial GIS packages. The user starts with a series of points such as those shown in Figure 12.8a, which represents socioeconomic status (SES) values for groups of households. The values in this hypothetical example range from 1 to 9, with 1 being very low SES (poor) and 9 being very high (rich). Because people who are poor tend to live next to other poor people, we can guess that if there is a cluster of poor people, a household with an unknown SES value will have a value similar to those for the ones around it. Figure 12.8b is an interpolated SES map. This interpolated map was created using the *inverse-distance-weighted* interpolation method, which determines cell values by using a linearly weighted combination of a set of sample points. The weight is a

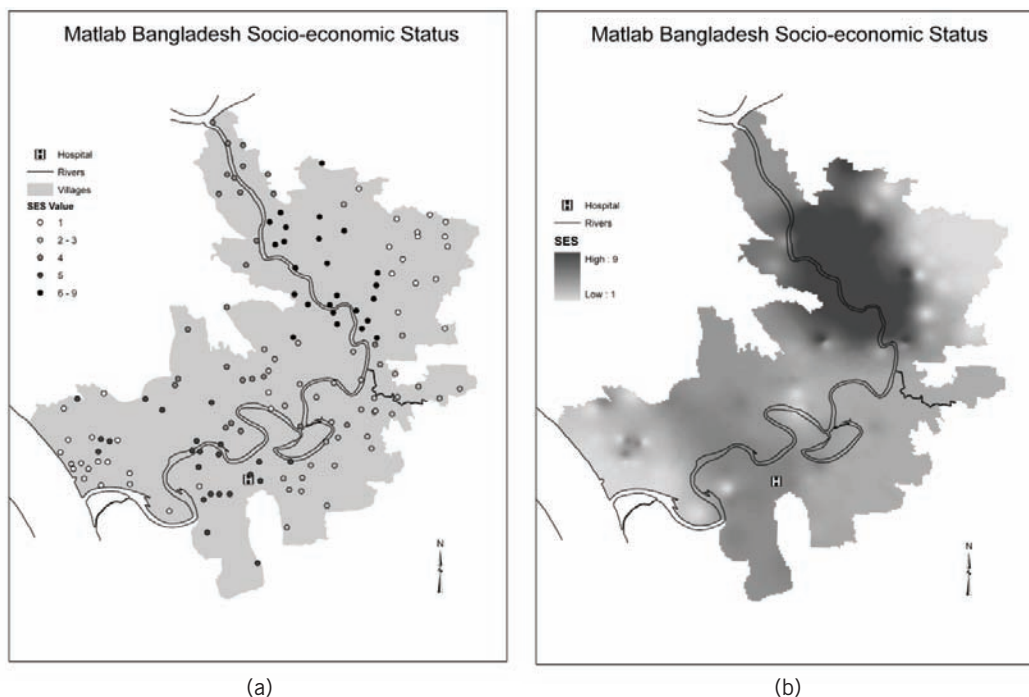


FIGURE 12.8. Interpolated SES map. (a) Known SES observation points. (b) Interpolated inverse-distance-weighted map.

function of inverse distance, which means that known observation points that are closer have more of an influence on the value at the unknown location. The input is a vector point map of known observations, and the output is a raster surface map of interpolated cell values.

SPATIAL ANALYSIS

There are many different types of spatial analyses that cannot be discussed comprehensively in this text (see “Further Reading” for many diverse examples). Much has been written about spatial analysis with a GIS. A few examples are given below and citations are provided in the “References” and “Further Reading” sections. Three types of spatial-analytical operations are discussed here. The first type is *buffer operations*, and the next is the fundamental concept of *overlay analysis*, of which there are many types. The last is *map algebra*, which is really an entire suite of raster spatial-analytical tools.

Buffers are new polygons that are zones of a specified width around a point, line, or polygon. For example, Figure 12.9 is a map showing areas in Lilongwe, Malawi, that were chosen to conduct a malaria vaccine trial. The circles around the

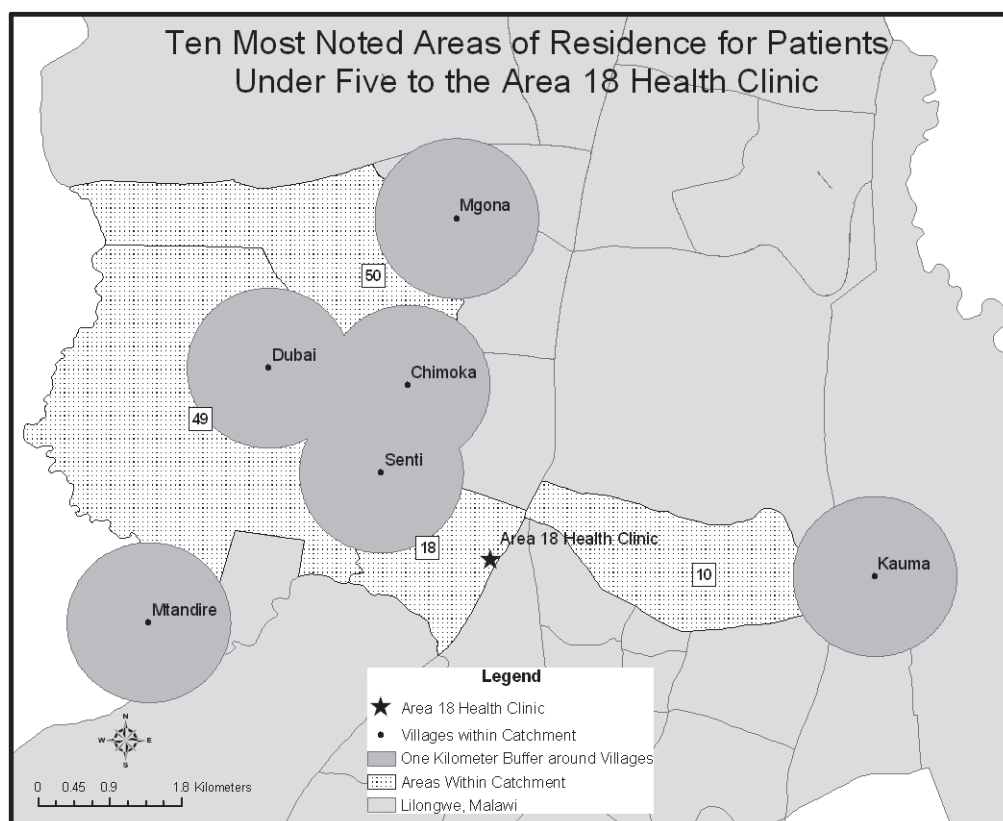


FIGURE 12.9. Patient recruitment around market areas in Lilongwe, Malawi.

market points represent the areas within 1 kilometer of those market centers. These buffer areas were used to define the area to recruit patients in the trial.

Overlay analysis is the integration of two or more map layers to create a new map layer. Different types of mathematical combinations of maps can be used to combine the layers. One type of overlay analysis is in a vector GIS: Two different polygon layers can be combined through a geometric intersection, and only features in the area common to both layers will be preserved. An example of how overlay analysis can be used in a study investigating the human–environment ecosystem drivers of an avian influenza virus (AIV) is offered below. Plate 15a is a map of global H5N1 AIV distribution from January 2003 through August 2008.

As described in Chapter 10, the highly pathogenic H5N1 AIV was first identified in 1997 and has spread across Asia, Africa, and Europe since 2003. Scientists and public health officials are trying to determine where in the world this pandemic flu could emerge next. A GIS approach has been devised to answer the “where?” question about a future strain of pandemic H5N1 AIV by spatially integrating population, environmental, and genetic information (Wan, Chen, Luo, Emch, & Donis, 2007; Wan, Wu, et al., 2007; Wan et al., 2008). Figure 12.10 shows the types of layers

that might be integrated in a GIS to understand what human–environment contexts are associated with appearance of the H5N1 virus and in genetic change in the virus. The idea is that certain land use patterns, population characteristics, and climatic patterns might lead to strains of the virus that are dangerous to humans.

Also as described in Chapter 10, a new genetic variant of the influenza A virus, H1N1, of swine origin appeared for the first time in March and April 2009; because humans had never been exposed to the virus before, it quickly became a pandemic. Plate 15b displays the geographic spread of the H1N1 flu as of July 6, 2009, when it had spread to 61 countries throughout the world. (By the end of July it had spread to more than 70.)

Map algebra is another system of spatial data integration, but it is in the raster data model (DeMers, 2000; Tomlin, 1990). In map algebra each cell in a raster map has a cell value that represents a characteristic of an earth surface feature. For example, in Figure 12.11, map 1 (top left) might represent areas with contaminated groundwater, and map 2 (bottom left) might represent areas with sick people. In map 1 a value of 1 represents a contaminated area, and in map 2 a value of 2 represents an area with sick people. Adding the values of the raster cells that correspond to the same areas in maps 1 and 2 would result in the values shown in map 3 (right). In map 3 values of 3 would therefore represent areas with contaminated groundwater and sick people; values of 2 would indicate areas with sick people without contaminated water; values of 1 would indicate areas with contaminated water and no sick people; and values of 0 would represent areas without contaminated water or sick people.

Map algebra is basically math applied to rasters that are geographically referenced. The language facilitates simple arithmetic combinations of rasters such as in Figure 12.11, or complex combinations of many layers. Arithmetic combinations

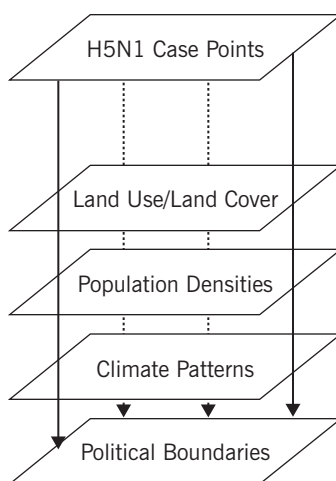


FIGURE 12.10. Human–environment ecosystem factors that can be overlaid with H5N1 virus locations and genetic information.

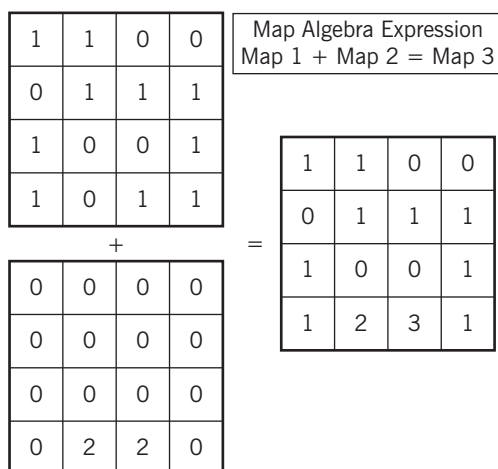


FIGURE 12.11. Map algebra expression.

include addition, subtraction, multiplication, division, and others. An arithmetic operator might be used to convert feet to meters by multiplying a raster in which cells represent elevation in feet by 0.3048, resulting in an elevation raster in meters. Other types of map algebra operators include Boolean and specialized. Boolean operators can be used to make true–false statements, resulting in 1 (true) and 0 (false) values in the output raster map. For example, in Figure 12.11 a Boolean statement could be created in the map algebra language that would result in values of 1 for areas with contaminated water and sick people, and the rest of the cells would have values of 0. Specialized operators are those that do something very particular. For instance, if an input raster represents elevation, the slope operator identifies the rate of maximum change in elevation from each cell. Map algebra is a powerful spatial modeling tool, and the analyst’s combination of tools in innovative ways defines the spatial-analytical operation. Commercial GIS software packages, including ESRI’s ArcGIS, implement combinations of map algebra tools.

SPATIAL STATISTICS

Geographic distributions require specialized statistics. Vignette 12.3 describes the concept of spatial autocorrelation, which is a measure of how similar things are to one another by distance. Vignette 12.2 describes a nonspatial statistical method called regression analysis, which allows the user to relate one distribution to another. For instance, a regression formula might look at how many calories per day people eat and how their caloric consumption is related to how much they weigh. The regression formula will no doubt show a positive relationship, but it will not be perfect because the amount of exercise people do modifies the effect of eating. One of the assumptions of regression analysis is independence of observations. What this means is that one observation cannot be dependent on another. But with spatially

autocorrelated map distributions, units are spatially dependent by definition. For example, if you were doing a study on how pollution affects health and you had data at the census tract level, neighboring census tract pollution values would be related to one another, and therefore the assumption of independence of observations would be violated. To deal with this problem, specialized spatial regression models have been created to deal with the nuisance of spatial autocorrelation (Anselin & Bera, 1998). These specialized models attempt to control for the autocorrelation (Anselin, 2002, 2005). One freeware software package that can be used to implement such spatial regression models is called GeoDa (see www.geoda.uiuc.edu).

Another type of spatial regression analysis called *geographically weighted regression* (GWR) explores the nature of relationships (Brunsdon, Aitkin, Fotheringham, & Charlton, 1999; Fotheringham, Brunsdon, & Charlton, 2003). For instance, the relationship between educational attainment and poverty might be strong in one part of a country and not another. In this hypothetical example, it's possible that poverty explains low educational attainment in areas where there isn't a lot of government spending on education. Therefore, since poorer people do not have money, they would not be able to attend college. In other parts of the country, where the government provides free college education to its people, the relationship might not be as strong because even poor people can attend college. GWR can thus be used as a tool to understand how spatial relationships vary in space and to hypothesize why they vary the way they do. A software implementation of GWR is called GWR3x and is available through the National University of Ireland, Maynooth (see ncg.nuim.ie/ncg/GWR/software.htm).

Spatial clusters of disease are aggregations of disease in space or in space and time. Disease clusters occur when there are more cases than are expected by chance. Identifying disease clusters can be used to identify possible risk factors related to the disease. For instance, as described in Chapter 3, John Snow identified a cluster of cholera near the Broad Street pump in London in 1854. The cluster was attributed to a contaminated water supply, and the pump handle was taken off. There are many different types of cluster methods for different types of distributions. Global spatial cluster methods detect clustering of a disease, but they do not identify the particular locations of the clusters. Local cluster methods identify the particular locations of clusters. The method of *focused clusters* determines whether there is a cluster around a particular location, such as the Broad Street pump.

One problem with spatial cluster analysis is that a user does not always have an idea of the spatial scale at which the cluster exists. The *spatial scan* statistic was developed to allow the user to look for clusters across spatial scales. This statistic detects clusters by gradually scanning an elliptical window across a study area, determining the number of observed and expected cases inside the ellipse around each location (Kulldorff, 1997; Kulldorff & Nagarwalla, 1995). An example is shown in Figure 12.12, which shows spatial clusters of a birth defect called gastroschisis in North Carolina. Notice that two clusters of different sizes were identified (Root, Meyer, & Emch, 2009). A freeware software implementation of the spatial scan statistic is SaTScan, which was developed at the National Cancer Institute (see www.satscan.org).

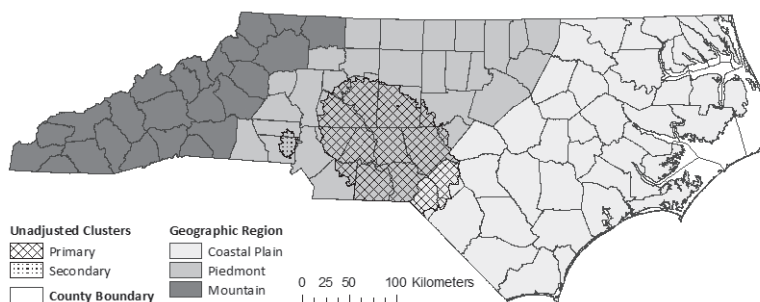


FIGURE 12.12. Spatial clusters of gastroschisis births, identified by using the spatial scan statistic.

CONCLUSION

The careful and accurate specification and analysis of disease data are complex undertakings. Scale is an important property that is often overlooked in disease studies. Much of the most important analysis in geography is of patterns and dynamics in space or over distance. Spatial analysis has some special complications of its own, but also some of the most powerful methods of portrayal and communication. The body of knowledge and much of the wisdom of experience accumulated over centuries in cartography are in danger of being ignored, if not totally lost. When studies do not consider cartographic principles or theoretical issues like scale effects and the ecological fallacy, they are often not well conceptualized. Computer and software development continues to advance, as does the development of technologies like the GPS for spatial data collection. With all of these advances, GIS-based health studies have become quite easy to implement, but there are also many ill-conceived projects that produce spurious findings by unthinking users. And yet the euphoria is real. Health researchers need to proceed with due care, with appreciation of the complexities of scale, but the possibilities are revolutionary.

REFERENCES

- Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3), 247–267.
- Anselin, L. (2005). *Exploring spatial data with GeoDA: A workbook*. Urbana: Spatial Analysis Laboratory, Department of Geography, University of Illinois, Urbana–Champaign. Retrieved from sal.uiuc.edu
- Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237–289). New York: Dekker.
- Bailey, A. J., Sargent, J. D., Goodman, D. C., Freeman, J., & Brown, M. J. (1994). Poisoned landscapes: The epidemiology of environmental lead exposure in Massachusetts children 1990–1991. *Social Science and Medicine*, 39, 757–766.
- Brunsdon, C., Aitkin, M., Fotheringham, A. S., & Charlton, M. E. (1999). A comparison of random

- coefficient modeling and geographically weighted regression for spatially non-stationary regression problems. *Geographical and Environmental Modeling*, 3(1), 47–62.
- Cleek, R. K. (1979). Cancer and the environment: The effect of scale. *Social Science and Medicine*, 13D, 241–247.
- Cliff, A. D., & Haggett, P. (1988). *Atlas of disease distributions: Analytical approaches to epidemiological data*. Oxford, UK: Blackwell.
- Cromley, E., & McLafferty, S. (2002). *GIS and public health*. New York: Guilford Press.
- DeMers, M. N. (2000). *GIS modeling in raster*. New York: Wiley.
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. E. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester, UK: Wiley Europe.
- Griffith, D. A., & Amrhein, C. G. (1997). *Multivariate statistical analysis for geographers*. Upper Saddle River, NJ: Prentice Hall.
- Haggett, P. (1976). Hybridizing alternative models of an epidemic diffusion process. *Economic Geography*, 52, 136–146.
- Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26, 1481–1496.
- Kulldorff, M., & Nagarwalla, N. (1995). Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 14(8), 799–810.
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographic information systems and science* (2nd ed.). Chichester, UK: Wiley Europe.
- Misch, A. (1994). Assessing environmental health risks. In L. R. Brown et al. (Eds.), *State of the world, 1994* (pp. 117–136). New York: Norton.
- Openshaw, S., Charlton, M., Wymer, C., & Craft, A. (1987). A Mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems*, 1, 335–358.
- Root, E. D., Meyer, R., & Emch, M. (2009). Evidence of localized clustering of gastroschisis births in North Carolina, 1999–2004. *Social Science and Medicine*, 68, 1361–1367.
- Rushton, G. (1998). *Improving public health through geographic information systems: An instructional guide to major systems*. Iowa City: Department of Geography, University of Iowa.
- Tomlin, C. D. (1990). *Geographic information systems and cartographic modeling*. Englewood Cliffs, NJ: Prentice Hall.
- Wan, X.-F., Chen, G., Luo, F., Emch, M., & Donis, R. (2007). A quantitative genotype algorithm reflecting H5N1 avian influenza niches. *Bioinformatics*, 23(18), 2368–2375.
- Wan, X.-F., Nguyen, T., Smith, C., Zhao, Z., Carrel, M., Davis, C., et al. (2008). Evolution of highly pathogenic H5N1 avian influenza viruses in Vietnam between 2001 and 2007. *PLoS One*, 3(10), 1–12.
- Wan, X.-F., Wu, X. M., Lin, G., Holton, S. B., Desmone, R. A., Shyu, C. R., et al. (2007). Computational identification of reassortments in avian influenza viruses. *Avian Diseases*, 51, 434–439.
- Ziegenfuss, R. C., & Gesler, W. M. (1984). Geographical patterns of heart disease in the northeastern United States. *Social Science and Medicine*, 18, 63–72.

FURTHER READING

- Abler, R. E. (1993). Everything in its place: GPS, GIS, and geography in the 1990s. *Professional Geographer*, 45, 131–139.
- Ali, M., Emch, M., Yunus, M., & Clemens, J. (2009). Modeling spatial heterogeneity of disease risk and evaluation of the impact of vaccination. *Vaccine*, 27(28), 3724–3729.
- Armstrong, R. W. (1973). Tracing exposure to specific environments in medical geography. *Geographic Analysis*, 5, 122–132.
- Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. New York: Wiley.
- Barnett, J. R., Pearce, J., & Howes, P. (2006). “Help, educate, encourage?”: Geographic variations

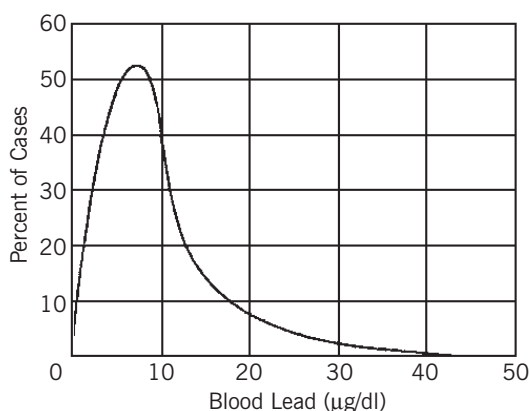
- in the provision and utilisation of diabetes education in New Zealand. *Social Science and Medicine*, 63, 1328–1343.
- Bennett, R. J. (1979). *Spatial time series*. London: Pion.
- Carrel, M., Emch, M., Streatfield, P. K., & Yunus, M. (2009). Spatiotemporal clustering of cholera: The impact of flood control in Matlab, Bangladesh, 1983–2003. *Health and Place*, 15(3), 771–782.
- Cliff, A. D., & Ord, J. K. (1973). *Spatial autocorrelation*. London: Pion.
- Curtis, S., Copeland, A., Fagg, J., Congdon, P., Almog, M., & Fitzpatrick, J. (2006). The ecological relationship between deprivation, social isolation and rates of hospital admission for acute psychiatric care: A comparison of London and New York City. *Health and Place*, 12, 19–37.
- Curtis, S. E., Southall, H., Congdon, P., & Dodgeon, B. (2004). Area effects on health variation over the life-course: Analysis of the longitudinal study sample in England using new data on area of residence in childhood. *Social Science and Medicine*, 58, 57–74.
- DeMers, M. N. (1997). *Fundamentals of geographic information systems*. New York: Wiley.
- Dobson, J. E. (1993). A rationale for the National Center for Geographic Information and Analysis. *Professional Geographer*, 45, 207–215.
- Dobson, J. E., Armstrong, M. P., Cromley, R. G., Goodchild, M. F., Marble, D. F., Peuquet, D. J., et al. (1993). "Automated geography" in 1993. *Professional Geographer*, 45, 431–461.
- Dunn, C. E., Bhopal, R. S., Cockings, S., Walker, D., Rowlingson, B., & Diggle, P. (2007). Advancing insights into methods for studying environment–health relationships: A multidisciplinary approach to understanding Legionnaires' disease. *Health and Place*, 13, 677–690.
- Dunn, C. E., Kingham, S. P., Rowlingson, B., Bhopal, R. S., Cockings, S., Foy, C. J. W., et al. (2001). Analysing spatially referenced public health data: A comparison of three methodological approaches. *Health and Place*, 7, 1–12.
- Fagg, J., Curtis, S. E., Stansfeld, S., & Congdon, P. (2006). Psychological distress among adolescents, and its relationship to individual, family and area characteristics in east London. *Social Science and Medicine*, 63, 636–648.
- Fotheringham, S., & Rogerson, P. (1994). *Spatial analysis and GIS*. London: Taylor & Francis.
- Gesler, W. M. (1986). The uses of spatial analysis in medical geography: A review. *Social Science and Medicine*, 23, 963–973.
- Giggs, J. A. (1983). Schizophrenia and ecological structure in Nottingham. In N. D. McGlashan & J. R. Blunden (Eds.), *Geographical aspects of health* (pp. 197–222). New York: Academic Press.
- Grady, S. C. (2006). Racial disparities in low birthweight and the contribution of residential segregation: A multilevel analysis. *Social Science and Medicine*, 63, 3013–3029.
- Grady, S. C., & Ramirez, I. (2008). Mediating medical risk factors in the racial segregation and low birthweight relationship. *Health and Place*, 14(4), 661–677.
- Haggett, P. (1990). *The geographer's art*. Oxford, UK: Basil Blackwell.
- Heywood, I. (1990). Geographic information systems in the social sciences. *Environment and Planning A*, 22, 849–854.
- Lam, N. S., & Quattrochi, D. A. (1992). On the issues of scale, resolution, and fractal analysis in the mapping sciences. *Professional Geographer*, 44, 88–98.
- Marble, D. F., Calkins, H. W., & Peuquet, D. J. (1984). *Basic readings in geographic information systems*. Williamsville, NY: SPAD Systems.
- Marshall, R. (1991). A review of methods for the statistical analysis of spatial patterns of disease. *Journal of the Royal Statistical Society*, 154, 421–441.
- McLafferty, S., & Grady, S. (2004). Immigration and geographical access to prenatal clinics: A GIS analysis. *American Journal of Public Health*, 95(4), 638–640.
- Meade, M. S. (1980). An interactive framework for geochemistry and cardiovascular disease. In M. S. Meade (Ed.), *Conceptual and methodological issues in medical geography* (pp. 194–221). Chapel Hill: Department of Geography, University of North Carolina.
- Odland, J. (1988). *Scientific geography: Vol. 9. Spatial autocorrelation*. Newbury Park, CA: Sage.
- Ord, K., & Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographic Analysis*, 24, 286–306.

- Pearce, J., Barnett, J. R., & Jones, I. (2007). Have urban/rural inequalities in suicide in New Zealand grown during the period 1980–2001? *Social Science and Medicine*, 65, 1807–1819.
- Pearce, J., Barnett, J. R., & Kingham, S. (2006). Slip! slap! slop!: Cutaneous malignant melanoma incidence and social status in New Zealand, 1995–2000. *Health and Place*, 12, 239–252.
- Ricketts, T. C., Savitz, L. A., Gesler, W. M., & Osborne, D. N. (1994). *Geographic methods for health services research*. Lanham, MD: University Press of America.
- Riva, M., Apparicio, P., Gauvin, L., & Brodeur, J.-M. (2008). Establishing the soundness of administrative spatial units for operationalising the active living potential of residential environments: An exemplar for designing optimal zones. *International Journal of Health Geographics*, 7, 43.
- Riva, M., Curtis, S., Gauvin, L., & Fagg, J. (2009). Health variation among categories of urban and rural areas: Evidence from a national sample in England. *Social Science and Medicine*, 68, 654–663.
- Riva, M., Gauvin, L., & Barnett, T. A. (2007). Toward the next generation of research into small area effects on health: A synthesis of multilevel investigations published since July 1998. *Journal of Epidemiology and Community Health*, 61, 853–861.
- Robinson, V. B. (1978). Modeling spatial variations in heart disease mortality: Implications of the variable subset selection process. *Social Science and Medicine*, 12D, 165–172.
- Walter, S. D. (1992). The analysis of regional patterns in health data: II. The power to detect environmental effects. *American Journal of Epidemiology*, 136, 742–759.

VIGNETTE 12.1. Chi-Square Statistical Test

A common problem faced in the health sciences involves comparing two sets of frequencies to ascertain how closely their distributions match. This is what is known as a *goodness-of-fit test*. It is applied to problems in which sample data are analyzed with statistics, such as chi-square (χ^2) or Kolmogorov–Smirnov, and the resultant statistic is tested for significance. What does it mean to say that an obtained result is “statistically significant,” or that it departs “significantly” from chance-level expectation?

As a simple example, consider the following situation: The environmental health literature tells us that many health problems associated with living in industrialized nations may be attributable to the presence of heavy metals such as lead or mercury in the human environment (Misch, 1994). Lead is a by-product of the manufacture of batteries and the burning of fossil fuels, and its presence has been associated with certain illnesses, particularly among children under 7 years of age (Bailey, Sargent, Goodman, Freeman, & Brown, 1994). Based on a large number of pediatric blood lead tests, we expect that in any sample of blood tests from that population, the number of individuals in each of seven categories of blood lead concentration should not deviate significantly from the curve of Vignette 12.1, Figure 1. If the results do deviate substantially, then the sample individuals are either at much higher or much lower risk than is typical. As the figure shows, just over half the children should be expected to have fewer than 10 micrograms of lead per deciliter of blood, just under a quarter of them should have between 10 and 14 micrograms per deciliter, and so forth. If too many children in a sample have higher than expected levels of lead in their blood, the sample distribution will



VIGNETTE 12.1, FIGURE 1. Distribution of blood lead levels.

deviate significantly from that shown in the figure. Some deviation might be expected by chance because no two samples are likely to produce exactly the same results. So how much deviation in a sample distribution is considered significant?

To answer this question, a statistical test known as one-sample χ^2 is performed. A table is set up to obtain the necessary terms for the computation of χ^2 (Vignette 12.1, Table 1). The term f_o means *frequency observed*, and f_e means *frequency expected*. Remember that the function of this statistical test is to compare observed results with those to be expected on the basis of an empirically derived distribution. Thus we compare the frequencies in seven categories of a sample of 15,077 Baltimore, Maryland, children's 1991 blood lead levels with expected frequencies. The expected frequencies are derived by using the proportions for each category shown in the figure (e.g., 53% of 15,077 is 7,991). The differences between observed and expected frequencies vary more in some categories than in others. The biggest difference occurs in the 25- to 34-microgram category, and the smallest difference is found in the 50+-microgram category. Could either

VIGNETTE 12.1, TABLE 1. Blood Lead Frequencies for Baltimore City and Maryland Rural County Children, by Lead Content Category, 1991

Blood lead (µg/dl)	Baltimore City	Rural counties	Totals
	f_o	f_o	
50+	63	7	70
35-49	275	9	284
25-34	792	34	826
20-24	951	23	974
15-19	1,768	64	1,832
10-14	3,414	202	3,616
<10	7,814	1,099	8,913
Totals	15,077	1,438	16,515

or all of these differences have happened just by an accident of sampling (chance variation)? And do the row differences matter, or is it just the total amount of variation that matters? The χ^2 test is computed with the equation

$$\chi^2 = \left[\frac{(f_o - f_e)^2}{f_e} \right]$$

which simply says: Subtract each expected frequency from the comparable observed frequency, square this difference, divide the difference squared by the expected frequency, and then add up these quotients. This is done in Vignette 12.1, Table 2.

What does $\chi^2 = 283.3$ mean? The larger χ^2 is, the greater the observed frequencies deviate from the expected frequencies. Chi-square ranges from 0, which indicates no departure of observed from expected frequencies, through a large number of increasing values. A large value of χ^2 indicates a large variation between one or more of the categories in the observed and the expected frequencies. This would suggest that the variation was due to more than just an accident of sampling. Statisticians have determined empirically and tabulated the probabilities that any computed χ^2 value with a given number of categories could have occurred by chance. It is now necessary to consult this χ^2 table, which may be found in almost any statistics book, together with instructions on how to use it. For this particular problem the table reveals that the value of $\chi^2 = 283.3$ could have occurred by chance alone less than 1 time in 10,000 (written as $p < .0001$). Suppose we had specified beforehand that in order for the computed value of χ^2 to be considered insignificant, it must not have occurred by chance more than 1 time in 100 samples of this size ($p < .01$). According to the χ^2 table, the computed value would have to have been less than 16.8. As we can see in the last column of Vignette 12.1, Table 2, the difference in the 25- to 34-microgram category alone would have been enough to lead us to conclude that the two distributions varied much more than could have been accounted for by

VIGNETTE 12.1, TABLE 2. Computation of χ^2 for the One-Sample Blood-Lead Problem

Blood lead ($\mu\text{g}/\text{dl}$)	f_o	f_e	$f_o - f_e$	$[f_o - f_e]^2$	$\frac{[f_o - f_e]^2}{f_e}$
50+	63	75	-12	144	1.9
35-49	275	226	49	2,401	10.6
25-34	792	452	340	115,600	255.8
20-24	951	920	31	961	1.0
15-19	1,768	1,824	-56	3,136	1.7
10-14	3,414	3,589	-174	30,276	8.4
<10	7,814	7,991	-176	30,976	3.9
Totals	15,077	15,077			$\chi^2 = 283.3$

Note. Observed frequencies are from the Maryland Childhood Lead Registry, Maryland Office of the Environment, 1991.

chance. Apparently the city of Baltimore has an unusually large number of children with blood lead levels in that range, and a somewhat excessive number in the 35- to 49-microgram category as well. Thus we may state that the differences in categorical frequencies between Baltimore children and a set of expected frequencies are too great to attribute to chance.

VIGNETTE 12.2. Regression Analysis

Statistics are used to analyze numerical distributions of data. Regression analysis is so basic, widely used, and generally useful that we have elected to explain this method by using a simple example, so that students may actually calculate it with the help of a hand calculator. The principle of regression is that the relationship between characteristics may be one of functional dependence of one on the others. Here we focus on the simplest of these forms, using only two variables. In this bivariate linear model, the magnitude of one characteristic (the *dependent variable*) is assumed to be determined by (i.e., is a function of) the magnitude of a second characteristic (the *independent variable*). If one has access to a computer with appropriate statistical software, it is advisable to use it for any regression or correlation problem involving more than a handful of observations.

As an example of a simple two-variable relationship, consider the maps in Vignette 12.2, Figure 1, showing the distributions of families living in poverty and the rate of low birthweights (below 2,500 grams) in the U.S. capital city, Washington, D.C. Low birthweight is an established surrogate for teen pregnancy and inadequate prenatal care. In the relationship between poverty and low-birthweight births, such births may be considered the dependent variable and poverty rate the independent variable. We may reasonably assume that although low birthweights may be a function of poverty, poverty is not determined by poor prenatal conditions among women. Clearly poverty is not the only determinant of poor prenatal care, but it is considered to be one explanatory factor. The maps show that the area on the east side of the city appears to have higher proportions of both the population below the poverty level and low-birthweight births. The fact that high values of the independent variable are associated with high values of the dependent variable, and similarly for low values, indicates that this is a direct, or positive, correlation.

Regression statistics usually are calculated from metric or ratio data. The resulting predictive equation, $Y = a + bX$, produces coefficients a and b , based on the assumption that both variables may have fractional parts. To return to the poverty and birthweight problem, the U.S. Census Bureau routinely reports the proportion of individuals and families whose income places them below the federal dollar value specified as the poverty level. Thus the independent variable is a ratio measure—percentage. Low-birthweight births are usually reported per 1,000 live births, another ratio measure.

One important statistic produced by regression analysis is the bivariate coefficient of correlation, which uses the symbol r . Potential values of this coefficient range from -1.0 to $+1.0$. A value of $r = 0$ means that there is effectively no measurable correlation



VIGNETTE 12.2, FIGURE 1. Maps of low-birthweight rate (a) and percentage of individuals below poverty level (b).

between the variables. The closer r is to 1.0 (in either direction), the greater is the association between the variables. If the coefficient for two variables was 1.0, we would be able to predict the exact value of Y for every known value of X within a given geographic area. Another way of putting it is to say that there would be no error term in the model.

We geographers are interested in place differences. Depending on the scale at which we are working, our places (observation points) may range from city blocks to nations. In U.S. cities the units of observation are usually census tracts (districts). The objective is to attempt to explain some of the spatial variation in low-birthweight births with the spatial variation in the proportion of families below the poverty line. Working deductively, we might posit that the simple correlation between low-birthweight birth rates (Y) and percentage of population below the poverty line (X) in a sample of census tracts in any U.S. East Coast city in a given year was positive and significantly greater than zero. Any such sample is assumed to have been drawn at random from the population of tracts in the city, and these populations are assumed to be normally distributed.

To test this hypothesis, it is first necessary to conduct the following calculations:

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{n}}{\sqrt{(\Sigma X^2 - \frac{(\Sigma X)^2}{n})(\Sigma Y^2 - \frac{(\Sigma Y)^2}{n})}}$$

where r is the coefficient of correlation, n is the sample size, X is the poverty variable, Y is the birth variable, and Σ is the symbol for summation; XY refers to the cross-product of X and Y , and X^2 or Y^2 refers to the square of the values of X or Y . If a table is initially established listing X , Y , X^2 , Y^2 , and XY for the sample tracts, then it is a relatively simple matter to sum these columns.

The correlation coefficient that is calculated from a sample is an estimate of a population parameter—in this case, the correlation coefficient in the population from which the sample was drawn. Recall that our hypothesis is that there is a significant positive correlation between Y and X , or, symbolically, $r > 0$. There are ways of testing the significance of the value of r . It is, however, necessary to have a basic understanding of inferential statistics to determine the significance of values computed in statistical tests.

To gain more realistic insights into the nature and degree of human disease and health relationships, multivariate statistics and/or mathematical models are far superior to simple univariate or bivariate statistics. One would be severely challenged to identify any measurable population characteristics or behaviors that could be satisfactorily explained by just one or two independent variables. Since the 1960s, the increasing power and availability of computers, coupled with the increasing sophistication of statistical software packages, have put complex statistical analysis capabilities within the reach of spatial analysts in many fields. A few useful multivariate techniques include regression, canonical correlation, factor analysis, discriminant analysis, analysis of variance, and cluster analysis. A good reference for multivariate statistical analysis is the text by Griffith and Amrhein (1997). Recently a useful technique called *multilevel modeling* has enjoyed success in spatial studies because it allows for simultaneous evaluation

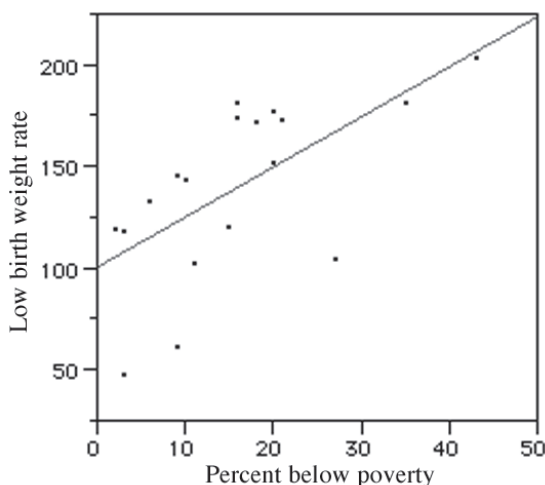
of phenomena at different scales (e.g., local, regional, and national). When we add the capabilities of GIS software to our arsenal, geography indeed has powerful analytical tools for the identification of and solutions for problems involving disease and health services.

The following explains the procedure for computing correlation and regression statistics, using a pocket calculator, for the Washington low-birthweight problem. From the data of Vignette 12.2, Table 1, it is convenient to graph the relationship between the two variables (Vignette 12.2, Figure 2). The line of best fit, or the *least-squares regression line*, may be fitted to the scatter in Vignette 12.2, Figure 2. To compute the correlation coefficient (*r*), we substitute data from the row labeled “Totals” in Vignette 12.2, Table 1, into the equation shown earlier:

$$r = \frac{44,946 - \frac{284(2,528)}{18}}{\sqrt{(6,546 - \frac{80,656}{18})(385,912 - \frac{6,390,784}{18})}}$$
$$r = \frac{5,059.8}{7,984.1}$$
$$r = .63$$

VIGNETTE 12.2, TABLE 1. Data for Computing a Correlation Coefficient for Low-Birthweight Rate per 1,000 Live Births (Y), and Percentage of Families below Poverty Level (X)

Census tract	Y	X	Y ²	X ²	XY
5.1	62	9	3,844	81	558
16	120	2	14,400	4	240
28	106	27	11,236	729	2,862
45	178	20	31,684	400	3,560
61	119	3	14,161	9	357
74.5	182	35	33,124	1,225	6,370
79.1	174	21	30,276	441	3,654
89.3	173	18	29,929	324	3,114
96.1	145	10	21,025	100	1,450
8.1	49	3	2,401	9	147
19.2	134	6	17,956	36	804
33.2	182	16	33,124	256	2,912
52.1	122	15	14,884	225	1,830
64.1	204	43	41,616	1,849	8,772
76.3	103	11	10,609	121	1,133
83.1	147	9	21,609	81	1,323
92.3	153	20	23,409	400	3,060
98.4	175	16	30,625	256	2,800
Totals	2,528	284	385,912	6,546	44,946



VIGNETTE 12.2, FIGURE 2. Regression line for low-birthweight births and percentage of families living below poverty line. The greater the proportion of families living below the poverty line, the greater the rate of low-birthweight births is expected.

As confirmed by the scatter diagram, the coefficient is positive. But, given the size of the sample, is r significantly greater than zero? One test of significance for this is known as the Student t test:

$$t_r = r \sqrt{\frac{n-2}{1r^2}}$$

Substituting for calculated values of r , r^2 , and n , we obtain a value (3.26) that may be compared with a standard table of t values for any size sample and any level of significance.

$$t_r = .63 \sqrt{\frac{16}{.598}} = 3.26$$

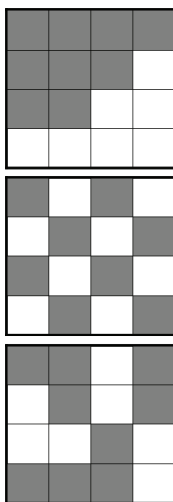
This value has a probability ($p = .005$) of about 5 chances in 1,000 of being due to sampling variation alone. Thus we can be reasonably certain that in Washington, D.C., the more birthweights below 2,500 grams in a neighborhood, the more likely we will be to find families with incomes below the poverty level.

VIGNETTE 12.3. Spatial Autocorrelation

Spatial autocorrelation is often discussed among geographers because it can be a serious hindrance to their work, but it can also be used as an analytical tool. *Spatial autocorrelation* means that observations from places next to each other are influenced by each other, in the same way that the real estate value of one piece of property affects that of the properties around it. If one city block is poor and black, the adjacent one is probably similar; if one county is affluent and has low unemployment, the adjacent county is probably better off than average as well. One important problem with spatial autocorrelation is that the assumption of independent observations, required for most statistical procedures such as linear regression and correlation, may be wrong.

It is perhaps easiest to explain spatial autocorrelation by mapping a dichotomous (nominal data) variable—one that is either present or absent in each spatial unit of a study area. Consider Vignette 12.3, Figure 1, which diagrams three situations in which a certain disease is either present (black) or absent (white) in each of 16 square units. In the first situation, there is positive spatial autocorrelation; black and white units are grouped together. Perhaps the illness is quite contagious in the black part of the study area, but has come up against some type of physical or human barrier in the white area. The second situation illustrates negative autocorrelation; adjacent units are dissimilar. In the third diagram, a random pattern of black and white units indicates no autocorrelation, either positive or negative.

There are two important things to consider in measuring autocorrelation: whether units are adjoining (“have a join”), and what the value of a variable or phenomenon is in each unit. One can say that there is a join if units have a common nonzero boundary (rook’s case, to borrow some terminology from chess), a common vertex or point



VIGNETTE 12.3, FIGURE 1. Spatial autocorrelation. Three types of spatial autocorrelation with a dichotomous variable are illustrated.

(bishop's case), or either of these (queen's case). In the figure, the unit values are simple presence or absence of a phenomenon.

How can one tell, statistically, whether there is autocorrelation in a particular situation? Basically, one counts the number of black–white (BW), black–black (BB), and white–white (WW) joins, and compares these with the number of joins that would be expected if the black and white units were distributed randomly. In the third diagram of Vignette 12.3, Figure 1, there are 6 BB joins, 4 WW joins, and 14 BW joins (rook's case). The appropriate formulas using this joint count can be found in texts that deal with autocorrelation. If there are significantly more BB or WW joins than expected, there is positive autocorrelation; if there are significantly more BW joins than expected, there is negative autocorrelation.

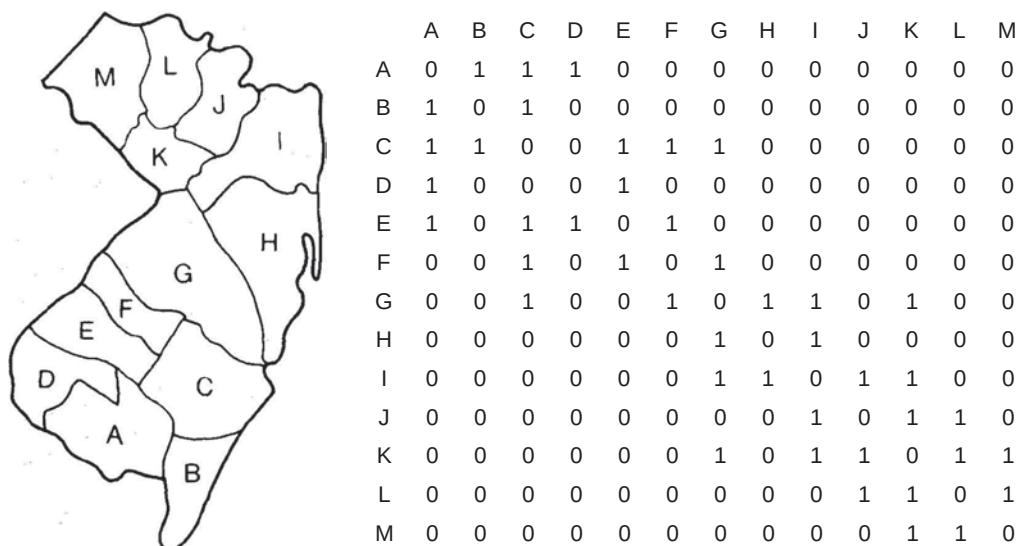
Unit values need not only represent absence or presence of a disease; they can also represent high and low disease rates. In addition, more than two nominal data categories can be considered and the analysis taken from the “two-color” to the “ k -color” case. Furthermore, definitions of a join can be altered in innovative ways. For example, Haggett (1976), in a study of measles diffusion in England (discussed in Chapter 8), defined joins in seven ways based on different types of paths along which the disease might be diffusing (e.g., along journey-to-work routes). If a certain path type indicated positive spatial autocorrelation, that particular path type could be important in the spread of measles—and, indeed, Haggett identified the importance of different pathways at different stages of the epidemic.

Spatial autocorrelation techniques have been used in a constructive manner to determine links between spatial patterns and causal processes. In particular, the techniques can help identify connections between disease rates and environmental and socioeconomic factors. Disease rates, which are interval data, have been most commonly used.

Thirteen counties from central and southern New Jersey have been selected to illustrate how to use spatial autocorrelations. The first step is to construct a 13-by-13 join matrix (rook's case) that consists of 0's if two counties do not have a nonzero boundary and 1's if they do (Vignette 12.3, Figure 2). Note that a county does not join itself.

To examine the heart disease rates for spatial autocorrelation, Moran's I statistic, which is used for interval data, can be calculated. If units with similar rates (high or low) tend to be next to each other, the I statistic will be relatively large (positive autocorrelation); if the opposite is the case, the I statistic will be relatively small (negative autocorrelation). The I statistic can be tested for a significance as a standard normal deviate (Z -score) after determining the mean and variance of its distribution. (Appropriate formulas can be found in books dealing with this subject.) If the calculated Z -score for a particular value of Moran's I is significantly positive, then one can say that a particular heart disease rate shows clustering at the level of adjoining counties.

The original join matrix can be modified or refined to reflect two factors simultaneously: the proportion of the boundary of one county that is common to another county, and the distance between county centers if there is a nonzero boundary. These modifications can be thought of as adding weights to the simple binary scheme of the original join matrix. The original join matrix can be modified to reflect a particular potential risk factor. For example, one can assign a join matrix value of 1 if two counties are urban to

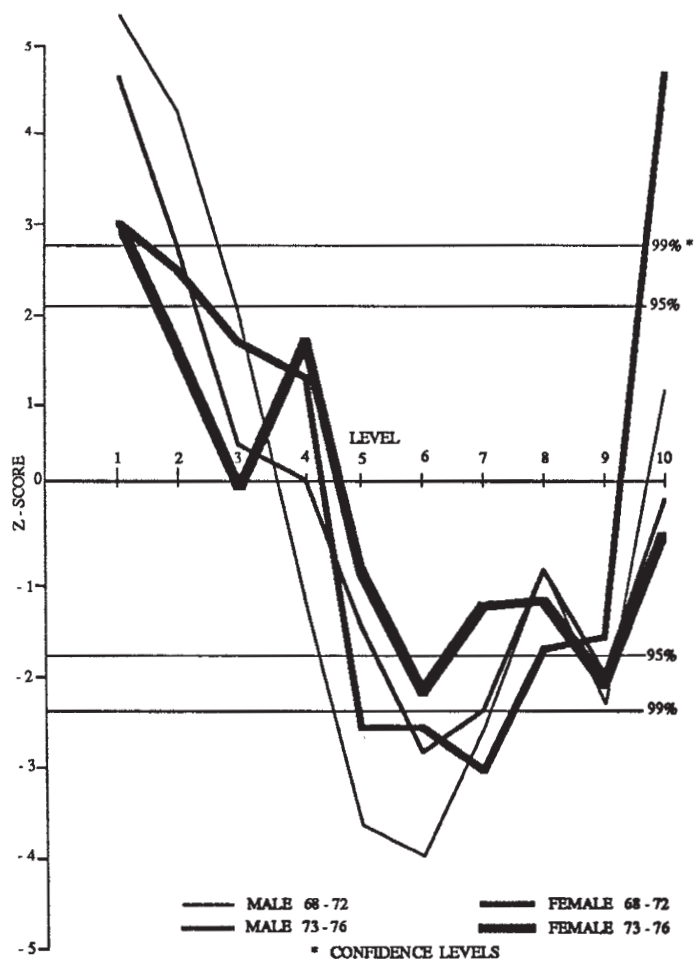


VIGNETTE 12.3, FIGURE 2. The join matrix for lag 1, neighbors, for a 13-county area in southern New Jersey.

the same degree and within a certain maximum distance of each other. If the value of Moran's I increases following this modification, the risk factor may be of importance.

The next stage in the analysis is to produce a spatial correlogram that is a rough indication of the areal extent of disease clustering. The original join matrix is modified to include 1's only if two counties are neighbors of neighbors; for example, county A 's neighbors of neighbors are county F and county G . The new matrix can be tested for spatial autocorrelation. If it is significantly positive, one can say that clustering is manifest at a larger scale than the original adjoining county scale. This process can be continued to neighbors of neighbors of neighbors, and so on (the general rule is to carry out approximately $n/4$ steps, where n is the number of units of observation). The result is a series of Z -scores for I statistics, which can be plotted as a spatial correlogram. The Z -scores from the original binary matrix are called the first lag; succeeding Z -scores are second, third, and fourth lags, and so on.

Vignette 12.3, Figure 3 shows spatial correlograms for white male and female acute ischemic heart disease for two time periods for a 49-county area surrounding New York City and Philadelphia. For all four sets of rates, there is significant positive spatial autocorrelation at the 99% level for the first lag and significant positive spatial autocorrelation at the 95% level for the second lag; on succeeding lags significant positive autocorrelation is no longer evident. At some lags there is significant negative autocorrelation, indicating that at these scales, adjoining groups of counties have dissimilar rates. It seems that for whites, acute ischemic rates cluster up to the neighbor-of-neighbor scale—a fairly large area within the 49-county area. Still there are groups of counties within the area that have dissimilar rates, as maps of acute ischemic heart disease rates attest (Ziegenfus & Gesler, 1984).



VIGNETTE 12.3, FIGURE 3. Spatial correlograms for whites with acute ischemic heart disease in the New York/Philadelphia metropolitan area, 1968–1972 and 1973–1976.

CHAPTER 13

Concluding Words

This textbook describes the geographic approach to studying disease and promoting health. Health-related phenomena have a certain distribution, move in certain directions at varying speeds, and affect people's perceptions of their communities and surrounding environments. Medical geographers who examine these spatial processes draw on the concepts and techniques of all the subdisciplines of geography. New diseases emerge from the changed landscape and patterns of settlement of the earth, and from disease agents' adaptations to our biomedical solutions that have neglected ecological considerations; human population fertility seems to be falling around the world to below replacement levels, and populations are aging and urbanizing; globalization brings democratization of technology and information, but also the devolution of power and new insecurities about the loss of regional culture and status. As these things and more happen over the next several decades, geographers need to bring their focus on scale, space, place, and a holistic approach to understanding health to the public discourse and to the promotion of health.

We have taken you, our readers, on a journey through culture, landscape, and evolution. We have described the power of new ideas, concepts, theories, and methodology to advance the understanding of health and disease. We have examined why new diseases are emerging, and why old ones have gone, as processes of adaptability. Processes of domestication and ecological simplification have resulted in nutritional diets, but also in many epidemics; we have explained why. Urbanization and older age structures have become worldwide phenomena and will profoundly affect your health in the future; we have explained why. Movement and variation in space simply *are*—all around us, and about everything. We have looked at increasingly sophisticated ways to map, analyze, model, integrate, and understand. It is possible, but extremely difficult, to separate the influences of genetics, environment, and group or individual behavior; again, we have explained why.

The medical geography of health care freely incorporates concerns and findings from other disciplines about the social, economic, political, and cultural behavior of individuals and systems. In turn, it contributes its geographic perspective to the emergence of a social science of health care. Similarly, epidemiological design and methodology, parasitology, entomology, microbiology, the anthropology of medical beliefs, and the sociology of groups, along with many other disciplinary insights, are synthesized and used in explaining the spatial distribution of disease occurrence. Medical geographers over the past 30 years have become better trained in the cognate fields relevant to their specializations, although some of this training is being lost today in the embrace of only things social. Most medical geographers recognize their medical limitations, and are eager to collaborate with other health professionals and scientists.

We have discussed problems that arise in available medical statistics and in obtaining micro-level data. These issues are changing rapidly as the importance of digitized data and of spatial analysis becomes ever more widely accepted. More important than data limitations to the beginning medical geographer is awareness that ignorance of basic and relevant biology can cause research hypotheses even about spatial form to be deficient, and that ignorance of basic sociocultural processes can result in overly simplistic genetic or environmental explanations. Spatial perspective, knowledge about differences of scale and spatial autocorrelation, and general familiarity with physical and social sciences can result in fresh geographic insights, new hypotheses, and sounder planning and policy. Spatial modeling without understanding of the disease system is sterile and perhaps irrelevant.

As we hope this text illustrates, the literature of medical geography is complex and exciting. There is a rich diversity to the research—a hybrid vigor and vibrancy derived from the mixing of many new ideas. Individual medical geographers may feel most comfortable with some segment of the subdiscipline. Some wish to be only theoretically informed; others to explore spatial analysis of diffusion in a positivist manner; and still others to focus on micro-level integration and qualitative understanding. Most still pursue the ancient geographic concerns: integrating all the phenomena at a point in space, in order to understand the nature of place and explain the distribution of varied phenomena over space.

Employment opportunities for health geographers are abundant, but usually at least a master's degree in the field is needed (although competence in geographic information systems [GIS] can qualify those with a bachelor's degree for some applied geography jobs). Besides holding academic positions, medical geographers with recent graduate degrees from the University of North Carolina at Chapel Hill alone are now directing international field training and research programs for Family Health International (e.g., AIDS intervention efforts in East Africa); working with the Centers for Disease Control and Prevention in Atlanta, the Research Triangle Institute, and the North Carolina State Center for Environmental Health; contributing to the Carter Center's Global 2000 effort to eradicate guinea worm in West Africa; and participating in numerous hospital and regional health care planning organizations. The term *spatial epidemiology* is appearing more often both in the geographic and public health literature. This usually translates as "the applica-

tion of GIS to disease and health problems.” Students who are interested in this field should pursue training in GIS and mathematics/statistics, as well as geography broadly considered.

Medical geography as a well-defined subdiscipline of geography has existed for only a few decades. During that time it has grown from a few individuals producing sometimes excellent work to a corps of researchers producing a substantial and varied body of studies. As it has matured, it has become more introspective and critical. The intellectual ties of the geographic perspective link together different projects, data sources, techniques, scales, hypotheses, and general paradigms to contribute a different voice to understanding. Medical, or health, geography has reached the point as an emerging body of literature that students and other medical geographers, as well as health researchers generally, can expect to draw on it before initiating a research project. It has reached the point where a text such as this, drawing from the work of dozens of individuals and hundreds of publications, can be written. Its perspectives are basic to addressing broader health questions about the role of social institutions, the consequences of environmental management or misuse, and the impact of sociocultural roles and perceptions—questions that are too frequently ignored. The health impacts of globalization and climate change will still be a matter of “airs, waters, and places,” as Hippocrates put it almost 2,500 years ago.

Index

Page numbers followed by *f* indicate figure, *t* indicate table.

- abiotic characteristics, 53
- abortion clinics, 417
- access to health care, 11
 - distance and, 414–418, 427–428
 - equity of, 11, 412–413, 421–422
- acclimatization, 213–215, 216–217
- acupuncture, 437
- acute disease, 19
- adaptability
 - of disease, 45–46
 - of health, 27, 44, 61, 232
- adiabatic lapse rate, 256
- Africa
 - cholera in, 54–55, 363–364
 - demographic transition, 158
 - development impacts, 181–184
 - HIV/AIDS, 382, 384–385, *plate 8*
 - malaria, 122, *plate 6*
 - mortality and cause of death, 283
 - traditional medicine, 433
- African Americans
 - health risk factors, 292–294
 - vitamin D deficiency in, 61
- age-adjusted death rates, 287, 413
- age as confounding variable, 164–165
- agents of disease
 - air-borne, 216
 - biological, 68–69, 71–72*t*
 - definition, 100
 - pathogenic, 353
 - species-specific, 45–46, 230
 - transmission/population mobility, 169–171
- age-specific death rate, 202–203
- age standardization, 164–165, 201–204
- age structure
 - population pyramids, 160, 163
 - stages of life and, 162–164
- agricultural conversion, 15, 174–175, 177, 179, 196
- agricultural frontier and disease ecology, 154*t*–155*t*, 173–175, 180–184
- agriculture
 - crop selection, 166–167
 - development impacts, 180–181
 - effects of climate change, 231
 - insecticides/herbicides, 274
 - patterns in, 62–63
- AIDS. *See* HIV/AIDS
- air-borne life, 218, 229
- air conditioning, effects on health, 208, 216, 217, 223
- air movement, 216–217
- air quality index and standards, 251–253
- air toxics, 257–258
- alcohol use, 39–40, 161, 299, 316–317
- alleles, 37, 38
- alternative medicine. *See* traditional medicine
- Amazon population resettlement, 177–178
- analogue area analysis, 304, 311–312
- animals
 - in disease systems, 100–101
 - domiciliated, 141
 - and industrial food production, 193–194
 - virus disease agents, 44–45, 68–69
- anthropo-zoonosis, 100, 102
- aquifers, 263

- arbovirus, 101, 106, 192
- ArcGIS software, 87–88, 234, 375, 466
- areal data, 80, 313, 460*f*
- arsenic poisoning in wells, 53
- arthropod vectors, 71*t*–72*t*, 101, 105–110, 246
- aspatial interaction, 347
- asthma, 43, 259–260
- Atlas of United States Mortality*, 310
- atmospheric inversions, 255–256
- atmospheric ionization, 219–220
- atmospheric pressure, 218–219
- attribute files (GIS), 84–86
- automobile emissions, 254–256
- avian influenza virus (AIV), 378, 379, *plate 15*
- ayurvedic medicine, 434, 438
- AZT medication, 384, 391, 392
- babesiosis, 135
- bacteria, 68, 69, 70, 132, 218
- Bangladesh, cholera distribution in, 459–463
- barefoot doctors (China), 421, 429
- barriers to disease diffusion, 357–358
- basal metabolic rate, 214–215
- basic reproductive rate, 353
- basin irrigation, 183
- behavior
 - and cancer, 308
 - and cardiovascular disease, 316–319
 - cultural, 40–43, 298–299
 - definition, 30–31
 - and women's health, 298–300
- beriberi, 56, 59
- best-practices model and health care, 431
- bimodal patterns of seasonality, 221–222, 225, 238–242
- biological classification systems, 68–70, 71*t*–72*t*
- biological controls, 131
- biomedicine, 437–438
- biomes, 105, 110, 137, 207, 244–246
- biometeorology
 - acclimatization, 213–215
 - biological rhythms, 211–213
 - climate change and, 229
 - radiation spectrum, 209–211
 - seasonality of death and births, 221–226
- biotypes (cholera), 361–362
- bird flu (AIV), 378, 379, *plate 15*
- birth control/contraception, 151, 243, 296, 297, 401
- birth seasonality
 - climate and, 224–226
 - models and methods of study, 238–241, 243*f*
 - temperature effects, 241–242
- bivariate weighted mean center, 449
- blood types, 37–38
- bomoh*, 438
- bone setters, 433
- breast cancer, 36, 172–173, 308–309, *plate 12*
- bubonic plague, 112–116
- buffer operations, 463–464
- buffers, cultural, 41, 54, 60–61, 108, 116, 134–135
- built environment
 - and cancer, 308–309
 - effects on health, 346–347
 - habitat, 32–33, 49, 53–54, 298
- bunds*, 53
- Calmette-Guérin (BCG) bacillus, 49
- calories, 55, 62
- cancer
 - behavior/habitat bases, 308–309
 - causation and development, 36, 304–305
 - as cause of death, 283–284, 287–288
 - electromagnetic fields (EMF) and radiation, 267–270
 - genetic base of, 306–308
 - patterns of incidence, 309–310
 - risk measurement, 310–311
 - spatial pattern analysis, 311–314
- carbon monoxide (CO), 251–252, 254–255
- carcinogens, 249, 304–305, 306
- cardiovascular disease
 - behavior/habitat/population bases, 316–321
 - as cause of death, 283–285, 287, 314, *plate 13a*
 - gender differences and, 299–300
 - overview, 314–316
 - regionalization in U.S., 321–325
 - temperature effects and, 223*f*, 224
- carriers of disease, 353
- cartograms, 82, 367
- cartography, 2, 14, 73, 454. *See also* GIS (geographic information systems); maps and mapping
- case-control studies, 302–303, 312
- case fatality rate, 353
- catchment areas, 418
- causal reasoning, 300–301
- causative organism, 68–69
- census data sources, 84, 85*f*, 322–324, 341, 448*f*, 449*t*, 457, 458
- Centers for Disease Control and Prevention (CDC), 98, 122, 167, 254
- central places and urban hierarchy, 355–356, 360*f*, 418–422
- central tendency leveling, 454
- centroid, 87, 369
- cercaria, 129
- cerebrovascular disease, 223
- Chagas disease, 110, 136, 177–178
- chain/step migration, 187
- chelation, 249
- Chernobyl nuclear disaster, 62, 265–267
- chickenpox, 44
- children, health of, 175
 - asthma, 43
 - dengue hemorrhagic fever, 126
 - and HIV/AIDS, 381, 383, 384
 - lead poisoning, 254–255, 289–291
 - nutrition and weight, 166–167, 168*f*

- China
 health care delivery, 429
 iodine deficiency in, 61–62
 pollution and health effects, 248, 257, 259
 traditional medicine systems, 435–437, 438–439
- chi-square test, 238–239, 454, 471–474
- chlorofluorocarbons (CFC), 209–210, 273
- chloroquine/quinine, 121, 123, 124, 176
- cholera
 in Bangladesh, 459–463
 cultural ecology, 52–55, 361–364
 mapping, 76, 91, 92*f*
- cholesterol, 281, 317
- choropleth maps, 80, 81*f*–82*f*, 459, 460*f*–463*f*
- chromosomes, 37
- chronic disease, 19
- chronic obstructive upper respiratory disease (COPD)
 in China, 248, 284
- circadian rhythms, 211–213
- clades, 382
- climate and health, 53, 207–208, 226–227, 228*t*, 244–246
- climate change and health
 atmospheric effects, 229–231
 nutrition and diet, 231–232
 overview, 227–228, 232–234
- clinical condition, 19, 353
- clinical front, 365
- clinical trials, 300
- clonorchiasis, 183
- clusters, focused/spatial, 467
- coefficients, correlation/regression, 455, 474, 476, 477*t*
- cohort effects/studies, 164, 302
- Columbian exchange, 46–47, 100, 177
- commensalism, 45, 46, 48, 69, 108
- common identifier (key), 86–87
- communicable disease, 98
- community and community-based intervention, 388
- compositional effects, 339, 345
- confounding variable, 165, 301–302
- congenital disease, 19, 226
- consumption (disease), 49
- contact/contagious diffusion, 19, 353, 355, 360–361, 364, 378
- contextual effects, 339, 345
- continuing property (of health), 28–30
- control points/variables, 303, 449
- cretinism, 61
- Creutzfeldt–Jakob disease, 69, 194
- criteria pollutants, 251–252
- cryptochromes, 211
- cultural ecology
 defined, 17, 26
 HIV/AIDS, 380–383, 390–391
 overview, 63–64, 116–117
 urbanization and, 188–192
- cultural–environmental interactions, 1, 9, 26
- culture realm, 41, 100, 137
- culture-specific health care facilities, 423
- cystic fibrosis, 36
- Dartmouth Atlas of Health Care*, 21, 407, 429–430
- data, representation of, 77–82
- data, types
 digital, 83–90
 longitudinal sets, 347
 microspatial mobility, 204–206
- data sources, 22, 84, 89–91
- data structure (GIS), 83–86
- DDT, 10, 121–122
- death rates, 157, 158, 165, 202*t*. *See also* standardized death rate
- death seasonality, 221–224
- decibels (dB), 247–248
- deforestation, 174, 176–177
- degenerative disease, 19, 161
- demand cones (distance), 418, 419*f*
- demographic base maps, 82–83
- demographic transition
 developing countries, 152*f*, 153, 159
 industrialized countries, 151–153
- dengue hemorrhagic fever (DHF), 125–127, 129–130
- dengue shock syndrome (DSS), 126
- density (of population), 52–53, 251, 272, 346
- desakota*, 188, 190
- deterministic modeling, 358–359
- developing countries
 adoption of biomedicine, health care, 437
 demographic transition, 152*f*, 153, 159
 development and disease, 183–184
 mobility, 169–171
 national health systems, 419–421
 urbanization in, 187–188
 water projects and sanitation, 183–184, 262
- development transitions
 impacts on rural ecologies, 154*t*–155*t*, 180–184
 resettlement programs, 146–148, 175–180
- diabetes, 36, 167
- diarrhea, 51, 52–55, 158
- diet. *See* nutrition
- diffusion maps, 78
- diffusion waves, 403–404
- dioxins, 248
- direction bias, 406
- directly observed treatment, short course (DOTS), 48
- direct transmission of disease, 47–49
- disabilities, mental/physical, 442
- disability-adjusted life years (DALY), 117*t*, 118, 127, 128*t*, 296
- disaster relief planning, 233–234
- disease
 Chinese traditional medicine, 436–437
 genetics and, 36–40, 42–43
 human ecology of, 26–27, 30–44
 and poverty, 292

- seasonality effects and, 221–224, 226
- terminology, 19–21
- of unknown cause, 325–329
- disease, infectious
 - evolution/adaptation/emergence, 44–47
 - infected persons, 353, 356, 357*f*, 381
 - transmission chains, 47–49
 - water-related, 49–55
- disease, transmissible
 - bubonic plague, 112–116
 - terminology, 100–101
 - transmission chains, 101–104
 - vector ecology, 105–110
 - yellow fever, 103–105
 - See also* landscape epidemiology
- disease, water-based
 - dengue hemorrhagic fever, 125–127
 - malaria, 117–121
 - overview, 116–117
 - schistosomiasis, 127–131
- disease agents. *See* agents of disease
- disease and health
 - affected by climate change, 227–234
 - definitions and terminology, 17–20
 - environmental pollutants and, 249–258
 - neighborhood health studies, 339–348
 - and risk of exposure, 204–206
 - water pollution, 260–265
- disease clusters, 76, 87–88
- disease cycles, 111, 114
- disease diffusion
 - cholera, 361–364
 - discussion, 351–352, 358–359
 - HIV/AIDS pandemic, 380–386
 - influenzas, 375–379
 - measles, 370–373
 - networks and barriers to, 356–358
 - poliomyelitis, 373–374
 - population mobility and, 169–171
 - terminology, 352–354
 - types of, 354–356
 - See also* modeling disease diffusion
- disease distribution and summary statistics, 459–463
- disease ecologies
 - agricultural frontier, 173–175
 - cancer, 304–314
 - cardiovascular disease, 314–325
 - HIV/AIDS, 380–383
 - influenza, 376–377, 378
 - overview, 7–8, 162
 - and unknown causes, 126, 325–329, 456
- disease maps
 - history and development, 12, 74–76, 90
 - HIV/AIDS, 367–369, 387
 - purposes/uses, 87–88
- disease transmission
 - infectious water diseases, 47–49, 50–51
 - sexually transmitted, 383, 386, 392, 394, 398
 - transmission chains, 47–49, 101–104, 123–124
- dispersed settlement, 33, 34
- distance and distance decay
 - diffusion and, 355, 360–364
 - distance demand cones, 418–419
 - locational analysis and, 414–418, 427–428
- distance and location
 - measuring, 87, 417–418
 - overview, 1, 5
- distance decay curves, 416, 418
- distribution
 - in disease studies, 11, 459–463
 - of health care resources, 412–413
 - in spatial analysis, 453–454, 457–458
- DNA
 - carcinogens and, 304–305
 - radiation and cell mutation, 267–268
 - RNA and HIV/AIDS, 380
 - structure and function, 35–37
- doctrine of natural nidality (Pavlovsky), 111
- domiciliated creatures, 141
- dot distribution maps, 77–78
- drinking water. *See* fresh water supplies
- dysentery, 69, 183
- Eastern equine encephalitis, 105–106
- ecological fallacy, 7, 312, 346, 454–455
- ecological simplification, 139–141
- ecology, definitions and descriptions, 3, 5–8
- ecosystems, characteristics of, 138–139, 140*f*
- ehrlichiosis, 135
- Eight Americas (Murray), 285*f*–286*f*, 286, 294, 322
- electrolytes, 214, 219
- electromagnetic fields (EMF), 268–270
- electromagnetic wavelengths, 88
- elephantiasis, 136
- elliptical cells of population, 322–324, 403, 426, 449–451
- El Tor* biotype, 361, 363
- emergency medical services/ambulances, 426–427
- emerging disease, 46–47, 174, 195–196, 383, 389. *See also* disease diffusion
- encephalitis, 131
- endemic disease, 20, 101, 169
- endogeneity problem in health studies, 347
- enigma area, 322
- environment, 4, 5–6, 7, 9–10, 11, 14
- environmental determinism, 238
- environmental disasters, 260–261, 265–267
- environmental justice (EJ), 260, 272
- Environmental Protection Agency (EPA), 251–253
- enzootic disease, 101
- epidemics and epidemic disease
 - major occurrences, 103–105, 125–126, 361–364, 387
 - overview, 44, 51, 54, 353
- epidemiological design, 300–304
- epidemiology, terminology of, 352–354
- epidemiology vs. ecology, 5, 7
- epizootic disease, 101

- equity of health care resources, 412–413, 421–422, 424, 429. *See also* access to health care
- eradication campaigns, 121–125, 181
- ESRI (GIS software), 233, 466
- essential hypertension, 314–315
- estivation, 105
- ethnicity, 38
- ethnomedicine, 432
- etiology
 - unknown causes of disease, 126, 325–329, 456
 - usage of term, 2, 9, 10, 26
- Europe, mortality and cause of death, 283
- expansion diffusion, 355
- extensive drug resistance (XDR), 48
- extremely low-frequency (ELF) waves, 215, 220–221
- farang*, 390
- fecal–oral transmission route, 47–48, 51
- Federal Land Development Authority (FDLA), 175–177, 421
- federal legislation and health (U.S.), 264, 410–411, 424
- fertility, 153, 161
- fetal alcohol syndrome, 161
- filarial worm, 70
- fleas and bubonic plague, 112–113
- flies, 50, 108–110
- flow maps, 78
- fomites, 19
- food production and inspection systems, 194
- fresh water supplies, 261–265
- friction of distance, 355, 366, 422
- frontier malaria, 179–180
- fungi/pollen, 218, 229, 248
- Galenic medicine, 434–435
- garden city movement, 443
- gender differences
 - morbidity and mortality rates, 284–246, 299–300
 - studies of, 294–295
 - See also* men; women
- genes and genetics
 - behavior effects on, 42–43
 - clades, 382
 - disease and, 35–40, 42–43, 306–308
 - research, 331–333
- geochemistry and trace elements, 319–321
- geocoding address data, 84, 89–90
- geographic information systems. *See* GIS
- geophagy, 42
- germs and germ theory, 9, 10, 100, 409
- Gini coefficients, 412
- GIS (geographic information systems)
 - data structure and function, 83–86
 - digitizing, 90–91
 - disease applications, 135, 137, 272, 344, 378, 387, 464
 - and GPS, 84, 89–90, 459–463
 - health care applications, 427–428, 455–458, 463–466
 - overlay analysis, 88–89, 344, 463, 464–465
 - overview, 73–74, 93–94, 374–375
 - sources/software, 87–88, 91, 458–459
 - spatial analysis, 453, 455–458
 - spatial data integration, 86–88
 - See also* disease maps; maps and mapping
- globalization
 - culture and behaviors, 42, 149, 150
 - overview, 2, 4–5, 16, 21, 485
- globalization and health hazards, 2, 272–274, 429
- globalization of movements
 - food production and health, 192–194
 - and population transition, 149
 - wage labor, 194–195
- global warming, 227–228
- goiter, 61, 214
- GPS (global positioning satellite), 74, 89–91, 459f
- gradients (slope), 38, 173, 254, 314, 325, 415f
- gravity models
 - diffusion through transformed space, 367–369
 - hierarchical diffusion, 360–364
 - HIV/AIDS mapping, 387
 - spatial autocorrelation, 370–374, 479–483
- green belts/greenways, 192, 444
- greenhouse gases, 229
- ground truth (remote sensing), 145
- groundwater, 261–264
- habitat
 - built environment, 31–34
 - cancer, 308–309
 - cardiovascular disease, 319–321
 - definition, 30
 - mapping, 145–146
 - social environment, 34–35
 - and women's health, 297–298
- habitat modification, 116
- Hagerstrand's lifelines, 171–172, 375
- Hamer–Soper model, 359
- harmonic analysis, 238–239
- hazards. *See* insults/stimuli/hazards
- health
 - Audy's definition, 27–30
 - biological classification systems and, 68–72
 - effect of status on, 35, 40
 - influence of climate and weather on, 207–209, 216–221
 - neighborhood effect and, 339, 345–346
- health and disease
 - affected by climate change, 227–234
 - definitions and terminology, 17–20
 - ecological simplification, 139–141
 - environmental pollutants and, 249–258, 274–275
 - population transition and, 161, 162–165, 180–184
 - water pollution and, 260–265
- Health and Medical Geography Specialty Group (AAG), 3

- health and nutrition
 - food preferences and, 63, 166–167
 - human nutritional requirements, 55–56, 57–58
 - malnutrition, 63, 157, 166, 231
 - See also* nutrition
- health care
 - costs, 407, 411–412, 416, 431
 - facilities, 423, 451–452, *plate 16*
 - promotion, 442–444
 - styles, 357–358, 441–442
 - utilization and provision, 429, 430–431
- health care, delivery
 - applying spatial statistics to, 448–452, 457–458
 - diffusion modeling and HIV/AIDS, 388
 - discussion, 7, 8, 407–409
 - Malaysian system, 419–421
 - Swedish socialized system, 424–425
 - United States system, 409–412, 428
- health care, resources and resource distribution
 - central place hierarchies, 418–422
 - equity of, 412–413, 415, 421–422
 - planning/location and allocation, 422–424
 - regionalization, 424–429, 430–431
 - utilization and distance, 415–418
- health geography, 2–3, 8, 15
- health level and insult stress, 28–29
- health promotion, 442–444
- health risks
 - assessment and prevention of pollution, 271–272
 - gender differences, 294–295
 - mapping/measurements of, 310–311, 321–325
 - race as a factor, 292–294, 318–319, *plate 13*
- heart disease. *See* cardiovascular disease
- heatwaves/heatstroke, 216–217, 229
- helminths, 69, 127–128
- hemophilia, 36, 381
- herbalists, 432–433
- herd immunity, 353
- hierarchical diffusion, 355–356, 360–361
- high-order
 - goods, 356
 - health services, 418, 424
- Hill–Burton Act of 1946, 410
- Hippocrates and Hippocratic tradition, 1, 9, 75, 434–435
- histocompatibility, 37
- HIV/AIDS
 - ecology of, 380–383
 - epidemics, 48, *plate 8*
 - geographic modeling, 367–369, 386–388
 - morbidity and mortality, 384–386, 395
 - in New York City, 395–397
 - in North Carolina, 397–398
 - social stigma of, 389–391
 - in Thailand, 391–395
- H5N1 virus (AIV), 378–379, 464–465, *plate 15*
- H1N1 virus (swine flu), 193–194, 195, 378–379, 465
- holistic medicine (U.S.), 439–440
- hospice care, 442
- hospitals
 - hierarchical location of, 418–419, *plate 16*
 - privatization and closure of, 428–429
- hosts, 45, 100–101, 102, 353
- housing and poverty, 178, 289–292
- human ecology
 - behavior, 30, 40–43
 - defined, 26–27
 - habitat, 30, 31–35
 - population, 30, 35–40
 - triangle of, 30–31, 52*f*
- human genome, 36
- human leukocyte antigen (HLA), 37
- humidity, 216–217, 218
- humors (body fluids), 434–435
- Huntington's chorea, 36, 292
- hydatidosis, 41
- hypertension, 314, 319, 322
- hypothalamus/hypothalamic mediation, 212–213, 214, 215
- iatrogenic disease, 195
- Iceland, disease diffusion in, 371–373
- imaging, satellite, 84, 88–90, 137, 179, 230, 345*f*
- immunity, 353
- incidence (of disease), 20
- incubation and incubation period, 20, 353, 365
- index case, 371
- index of segregation, 412
- industrialized countries
 - cancer, 38, 310
 - cardiovascular disease in, 314–325
 - pollution in, 247, 259, 268, 271
 - women's health in, 296–297
- Industrial Revolution, 61, 158, 247, 321
- infantile paralysis, 373–374
- infant mortality rates, 151, 157–158, 161, 181, 284
- infection/infected persons, 353, 356, 357*f*, 381
- infectious disease
 - evolution/adaptation/emergence, 44–47
 - infected persons, 353, 356, 357*f*, 381
 - transmission chain, 47–49
 - water related, 49–55
- infectious hepatitis (Australia), 364–366
- influenza
 - diffusion of, 369, 375–379
 - seasonality of, 224
 - viral shift, 376–377, 378–379
- infrared/longer-wave radiation, 210
- infrasound, 220–221
- innovation diffusion, 403–404
- insults/stimuli/hazards, 27–28, 37, 208
- international labor migrants, 161
- international medical graduates (IMG), 417
- interpolation of data, 79
- intervals and distribution patterns, 80–83, 480–481
- intervention and intervention points, 123–124, 130–131, 134–135
- intravenous drug use (IDU), 381, 391, 394, 396

- inverse-distance weighted interpolation, 462–463
- inversion layers, 256–257
- iodine deficiency, 61–62, 214
- ionizing radiation, 209
- irrigation, 182–184
- isolines/isopleths, 78–79, 452*f*
- Japan, disease in, 324–325, 326
- Jarvis's law, 416
- joins, 370, 371–372, 479–480, 481*f*
- jungle cycles, 104, 125
- kuru*, 194
- kwashiorkor*, 55–56, 64
- labor migrants/workers, 161, 194–195
- labor-sheds, 124, 170
- lactose intolerance, 39
- lalang*, 148
- land development and resettlements, 147–148, 177–178
- landscape epidemiology
 - field mapping, 145–148
 - overview, 98–99, 110–111, 136–138
 - tick-borne disease, 135
 - transmissible disease systems, 100–105
 - vector ecology, 105–109
 - See also* disease, transmissible; disease, water-based
- land use and patterns of disease, 322–325
- latency/latency period, 20, 353
- latitude/latitudinal, 210, 227, 246, 256
- latitudinal gradient and disease occurrence, 173, 325–326
- lead and lead poisoning, 249, 254–255, 289–291
- least-squares regression line, 477–478
- leishmaniasis, 110
- leprosy, 33
- lice and disease transmission, 105, 113*t*, 135
- life cycles of vectors, 32, 100–101, 111, 118, 127, 130*f*
- linear data, 78–79
- linear settlement, 33, 34
- linear transmission channels, 120
- location-allocation planning, 422–424
- location of health care facilities/personnel, 426, 430, 451
- locational analysis, 408, 414–418
- locational quotients, 412
- logistic curve, 353, 404
- longitudinal data sets, 347
- Lorenz curves, 412
- Love Canal (New York), 260–261
- low-frequency waves and infrasound, 220–221
- low-order
 - goods, 356
 - health services, 418
- Lyme disease, 41, 132–134, 135
- lymphatic filariasis, 136, 183
- magnitude and rate of population growth, 153
- malaria
 - cultural ecology, 119–121, 179–180, *plate* 6
 - disease cycle/life cycle of vector, 117, 118–119
 - eradication and intervention, 121–125
 - population resettlement in Malaysia, 176–177
 - precipitation and spread of, 230–231
- Malaysia
 - health care system, 419–422
 - malaria eradication, 122–124
 - population resettlement, 146–148, 175–177
- males having sex with males (MSM), 386, 389, 395, 398
- malnutrition, 55, 63, 157, 166, 231
- managed care (HMO), 411–412
- management of patient flows, 426
- map algebra, 463, 465–466
- maps and mapping
 - Dartmouth Atlas of Health Care*, 21, 407, 429–430
 - data representation and visualization, 77–80
 - disease maps, 12, 74–76, 90, 367–369
 - field mapping, 145–148
 - health services and, 412–413
 - multiple hazard surface mapping, 171
 - overview, 73, 75
 - patterns in interval distribution, 80–83
 - potential maps, 78, 452
 - suitability mapping, 290
 - suitability of scale, 77
 - See also* GIS (geographic information systems)
- marasmus, 55
- market-shed, 356
- maternity and reproduction, 295–296, 433
- maximum-rate-of-growth, 159
- mean center in central place, 448–449, 450*t*, 452
- mean centers of distribution, 461*f*
- mean information field (MIF), 405–406
- measles, 45, 48, 370–373
- Medicaid and Medicare, 410–411, 423, 429
- medical geography
 - challenges of, 21–22
 - history and development, 9–14, 74–76, *plates* 1–2
 - interdisciplinary developments, 14–17
 - overview, 1–5, 7–8, 65, 483–485
- medical pluralism, 438–441
- medical taxonomy, 71*t*–72*t*
- megacities, 187–188, 189–190
- melanin, 60
- melatonin, 211
- men
 - effect of temperature on, 225, 240*f*, 241
 - gender differences in studies, 294–295
 - mortality and cause of death, 284–286, 294, 299
 - prostate cancer, 293, 307, 314, 315*f*
- mental illness
 - and deinstitutionalization, 13, 288–289
 - distance decay and location of facilities, 415–416, 417*f*

- metabolism and metabolic differences, 39–40
- metastasis, 306
- microbe, 45, 100
- microfilaria, 70
- microspatial exposure analysis, 204–206
- migration, chain/step, 187
- migration/circulation, 153, 156–157, 159, 160–161, 168–169
- migration streams, 173
- migration to foreign countries. *See* globalization of movements
- Millennium Goals (U.N.), 164, 181, 230
- minimum spanning tree, 371, 372f
- miracidium, 129
- mobility and mobility patterns, 42, 116, 124, 131, 141
- mobility transition, 153–157, 168–173, 204–206
- modeling disease diffusion
 - discussion, 354, 359–360, 377–378
 - distance and hierarchy, 360–364
 - HIV/AIDS modeling, 367–369, 386–389
 - probability simulation, 405–406
 - spatial autocorrelation, 370–374, 479–483
 - spatial data managing, 375
 - spatial interaction models, 366–367
 - stochastic simulation, 364–366, 403–404
 - transformed space, 367–369
- modifiable areal unit, 454, 456–457
- morbidity and mortality rates, 165, 221, 299–300, 383–384
- mortality
 - cancer rates, 309–310
 - cardiovascular disease, 314, 316
 - declining rates, 159
 - dengue hemorrhagic fever, 126, 127
 - health care utilization and provision and, 429–431
 - malaria, 120
 - Parkinson's disease, 327
 - spatial scales of, 283–288
 - and temperature, 221–224
- mortality and cause of death, 283–288
- Mortality and Morbidity Weekly Report* (U.S.), 373
- mosquitoes
 - dengue hemorrhagic fever, 125, 126, 127
 - and disease transmission, 106, 107t, 108, 136
 - yellow fever, 103–105
- mosquitoes (*Anopheles*)
 - eradication and intervention, 121–125, 178
 - malaria vectors, 101–102, 118–120, 176, 230–231
- multifactor zones (May), 111
- multifactorial disease, 282
- multilevel modeling, 476–477
- multiple sclerosis, 173, 325–326
- National Health Planning and Resources Development Act (U.S.), 411
- National Research Council (U.S.), 4–5
- Native Americans, 46–47, 59–60, 286–287, 433
- natural increase of population, 151, 185
- natural nidus region, 100, 111, 127–128, 132–133
- nearness/joins, 370, 371–372
- neighborhood health studies
 - analyzing and implementing, 345–346, 347–348
 - definition and units of analysis, 341–345
 - spatial dimension studies, 339–340
- network nodes, 427
- networks in time and space, 375, 403 *plate 14*
- newly industrialized countries (NIC), 247–248
- new urbanism, 191–192, 444
- New York City
 - HIV/AIDS in, 395–397
 - poliomyelitis in, 373–374
- niacin, 59
- nitrogen dioxide (NO₂), 251
- noise levels, 247–248
- noncommunicable disease
 - mortality levels, 283–288
 - political ecology, 281–283
- North Carolina
 - HIV/AIDS in, 397–398
 - mortality and cause of death, 286–288
 - spatial scale in neighborhood studies, 341–343
- nuclear power and weapons plants, 268f, 270–271
- nuclear settlement, 33–34
- nucleotides, 36
- nutrition
 - affected by climate change, 231–232
 - deficiencies and disease, 56, 59–62, 165–167
 - food preferences and, 63, 166–167
 - human requirements, 55–56, 57t–58t
 - malnutrition, 63, 157, 166, 231
 - and poverty, 292
 - regionalization of diet, 62–65
- obesity, 36, 167, 168f, 347
- odds ratio, 310–311
- Ohio, mapping HIV/AIDS diffusion, 367–369
- onchocerciasis, 91, 109–110, 136, 181
- oncogenes, 305, 307
- organically grown food, 65, 443
- osteoporosis, 61
- overlay analysis, 88–89, 344, 463, 464–465
- ozone (O₃), 209–210, 251, 255, 273–274
- pandemics, 47, 103, 115, 193, 377
- parasite, 100
- Parkinson's disease, 36, 327–329
- particulate matter pollutants, 251, 255, 257
- passage of fronts, 221
- pathogens, 27
- patterns in maps, 80–82
- Pattison, four traditions of geography, 4–5
- pellagra, 59–60
- permeable barriers, 357, 406
- pet therapy, 440
- photochemical smog, 255–257

- photons, 209
- physicians
 - international medical graduates (IMG), 417
 - personnel and facility location, 412–413
 - physician-to-population ratios, 457–458
- pineal gland, 212
- pixels, 88–89
- place, 5, 9, 441–442
- plasmids, 69
- pneumonic plague, 112
- point data, 77–78, 87, 313, 459–460*f*
- Poisson distribution, 364–365
- poliomyelitis, 50–51, 373–374
- political ecology, 15–16, 196, 281–283, 329–331, 333
- pollutants, primary and secondary, 255
- pollution, indoor air, 259–260
- pollution, outdoor air
 - air toxics, 257–258
 - atmospheric, 249–254
 - automobile emissions, 254–256
 - inversions (atmospheric), 255–257
- pollution, radioactive
 - electromagnetic fields (EMF), 268–270
 - medical/mining hazards, 270–271
 - naturally occurring hazards, 265–267
 - radiation and cell mutation, 267–268
- pollution, water
 - environmental disasters and, 260–261
 - fresh water supplies and, 261–265
- pollution syndrome
 - globalization and, 272–274, 276
 - noise levels, 247–248
 - risk assessment and prevention, 271–272
 - spatial dimensions to health hazards, 274–275
 - See also* toxic hazards
- polychlorinated biphenyls (PCB), 249, 261, 298, 309
- polymorphism, 37
- population
 - definition, 30–31
 - genetics and genetic variability, 35–38
 - metabolic differences in, 39–40
 - vulnerable/susceptible, 36–37, 217, 233, 353
 - women's health, 295–297
- population growth, magnitude and rate of, 151–153
- population potential, 78–79, 366, 369, 451–452
- population pyramids of age structure, 160, 163
- population resettlements
 - Amazon, 177–178
 - Malaysia, 148–149, 175–177
- population-sheds, 418
- population transitions
 - demographic, 151–153, 159
 - impacts of, 162–165
 - mobility, 153–157, 168–173, 204–206
 - nutritional, 157, 165–167
 - overview, 149–151, 154*t*–156*t*
 - stages in transitions, 157–162
- potential maps, 78, 452
- poverty syndrome, 288–292
- precautionary principle, 329–331, 333
- precipitation, 229–231, 232
- preference-sensitive care, 440
- prevalence (of disease), 20, *plates* 6–7
- primacy in urban systems, 187
- primary care facilities in Malaysia, 419–422
- prions, 69
- privatization of health care, 428–429
- promoters of cancer, 304
- prospective study, 302
- prostate cancer, 293, 307, 314, 315*f*
- protein-energy malnutrition (PEM), 55–56
- protozoa, 51, 68, 69
- Public Health Service Act. Title XVI, 410
- quarantine, 47, 351
- quintiles, 80–81
- rabies, 191–192
- race
 - and genetic variability, 37–38
 - health risks and, 292–294, 318–319, *plate* 13
- radiation hazards
 - medical/mining sources, 270–271
 - radon, 260, 298
- radiation spectrum, 209–211, 326
- randomization (statistics), 303, 365
- rapid epidemiological mapping of onchocerciasis (REMO), 91
- raster data and display, 83, 88, 89, 461–462, 463, 465
- rationalizing health care delivery, 426–429
- realms of evolution, 99–100, 136
- reduviid bug (kissing), 109*t*, 110, 136, 178
- re-emerging disease, 191–192, 195
- regionalization and diet, 62–65
- regionalization of health care
 - defining a region, 424–427
 - GIS analysis and simulation, 427–429
- regional medical programs (RMPS), 410–411
- regions, epidemiology
 - multifactor zones, 111
 - types of, 99–100, 136–137
- regression analysis, 466–467, 474–478
- relapsing fever, 135
- relative risk and odds ratio, 310–311
- relocation diffusion, 114, 355
- replication of infectious agents, 69
- reportable/unreportable disease, 361, 364
- reservoirs (disease), 100, 104, 128, 134
- reservoirs (water), 184
- retrospective study, 302
- retrovirus, 380–382
- reversing type, seasonal shift, 222
- rheumatic diseases, 326
- rice and beri beri, 59
- rickettsia, 68, 69, 105, 132–133
- ricketts, 61, 297
- RNA (ribonucleic acid), 380
- Rocky Flats nuclear weapon plant (Colorado), 268

- Rocky Mountain spotted fever (RMSF), 132–133, 135
- rodents and bubonic plague, 112–116
- rural ecologies, health impacts, 180–184, 294
- rural health care
- rationalizing and privatization of, 427–428
 - resource distribution, 421, 424
- rural–urban migration, 185, 187
- health impacts of, 189–190, 191
 - mobility transition, 159–160, 161–162
- See also* urbanization
- Safe Drinking Water Act, 264
- salt, iodized, 61–62
- samples and sampling, statistical, 303
- sanitation and water supply, 32, 261–265
- SaTScan software, 88, 467
- Savannah (Georgia), 322–324
- scalar (proportional lines), 78
- scale
- and distribution/variance, 453–455, 456*f*, 457–458, 468
 - of maps, 7, 77, 454
- schistosomiasis, 127–131, 182–183
- schizophrenia, 36, 226
- scrub typhus, 147*f*–148, 297
- seasonal affect disorder (SAD), 213
- seasonality
- effects on health, 53, 218, 221–226
 - monthly indexes, 236–237
 - shifts, 222–223
- secondary attack rate, 353
- secondary transmission, 54
- Sennetsu fever, 135
- serology and serotypes, 19, 125, 126
- serotonin, 212, 220, 226
- settlement patterns/systems, 33–34, 44–45
- severe acute respiratory syndrome (SARS), 379, 389
- sex tourism/trade, 390, 393, 394–395
- sexually transmitted disease (STD), 383, 386, 392, 394, 398
- shingles, 44
- sickle cell anemia, 36, 117, 293
- siddha* system (India), 434, 438
- significance tests, 206, 471–474, 480–481
- silent zone of disease, 111, 135
- simian immunodeficiency virus (SIV), 382
- simultaneous infection, 381
- Sinemet, 328
- skin cancer, 313
- small-area analysis, 418
- smallpox, 48
- smog, 255–257
- snails and schistosomiasis, 127, 129, 131, 183
- social distance, 414, 423
- social ecology, 7
- social environment, 34–35, 53–54
- socialized health care systems, 424–425
- social norms, 42
- social sciences, 14–16, 17, 38
- socioeconomic status (SES)
- cardiovascular disease and, 318
 - health effects of, 35, 40
 - mortality and, 293–294
 - as statistical value, 462–463
- sociopathological complex, 288–289, 292
- solar radiation, 53, 224, 244–246, 308, 326
- spatial analysis
- analogue area analysis, 311–312
 - overview, 13–14, 87, 453–454, 468
 - scale and distribution/variance, 453–454, 457–458
 - types of, 463–464
- spatial–analytical tradition of geography, 4–5, 354
- spatial autocorrelation, 312–314, 370–374, 457–458, 466, 479–483
- spatial basic reproduction number, 374–375
- spatial correlograms, 313–314, 481–482
- spatial data integration, 86–87, 88
- spatial data managing, 375
- spatial epidemiology, 484–485
- spatial interaction models, 366–367, 427
- spatial interpolation, 462–463
- spatial multilevel models, 347
- spatial statistics, 448–452, 466–468
- specialized care facilities in Malaysia, 419–422
- species-specific disease agents, 46, 230
- spectral resolution, 89
- spiritual healers, 432
- sporadics, 306
- squatter settlements/shantytowns, 185
- stages in population transitions, 156*t*, 157–162
- stages of life, 163–164
- standard deviation ellipse, 452, 461
- standardized death rate, 201–204
- standardized mortality ratio (SMR), 203–204, 413
- standortraum*, 132
- starvation, 55, 63, *plate* 2
- statistics
- Poisson distribution, 364–365
 - preparing data for comparison, 236–237
 - significance tests, 206, 471–474, 480–481
- stigma of HIV/AIDS, 389–391, 397
- stimuli. *See* insults/stimuli/hazards
- stochastic diffusion simulation, 364–366, 403–405
- stratification (statistics), 303
- stress and disease, 29, 317–318
- strokes, 322, 324, *plate* 13
- structural loci on chromosomes, 37
- structurization, 30
- subclinical condition, 19, 352
- suitability mapping, 290–291
- sulfur dioxide (SO₂), 255, 257
- summary statistics, 461–463
- supervector, 125, 126–127
- suprachiasmatic nucleus (SCN), 211
- Sweden, health care in, 424–425
- swine flu (H1N1 virus), 193–194, 195, 378–379
- syphilis, 46

- t'ai chi chuan* (Qidong), 436
- temperature
- inversions, 256
 - relation to health, 216–217, 223–224, 225, 256
- temporal resolution, 89
- Thailand, HIV/AIDS in, 73, 391–395
- therapeutic landscapes, 15, 441–442
- thermoregulation, 214
- thiamine, 56, 59
- Third World. *See* developing countries
- threshold of population, 356
- thyroid, 61–62
- ticks and tick-borne disease, 106*t*, 132–133, 134–136
- TIGER GIS files, 77, 84, 458
- time–space geography
- analyzing methods, 168–171, 374–375, *plate 14*
 - measuring, 171–173
- tobacco and tobacco smoke, 259, 260, 306, 316–317
- total fertility rate, 151, 153, 157, 163, 177
- toxic hazards
- origins, 248–249, 250*t*
 - risk assessment and prevention, 271–272
- See also* pollution
- trace elements, 319–321
- trace minerals, 58*t*
- trachoma, 50
- traditional medicine
- cultural alternatives and perceptions, 431–433
 - medical pluralism and, 438–441
 - medical systems (professional), 432, 433–437
 - practitioners, 432–433
- transformed space modeling, 367–369
- transforming landscapes, 443–444
- transitory type of shift, 222
- transmissible disease systems
- cultural ecology, 131–135
 - terminology used, 100–101
- transmission, cross-species, 192
- transmission chains, 47–49, 101–104, 114*f*, 123–124
- transovarial transmission, 105
- triangle of human ecology
- behavior, 40–43, 298–300
 - habitat, 31–35, 43, 297–298
 - population, 35–40, 43, 295–297
- triangulation/trilateralization, 89
- trophozoite, 69
- tropical ecology, 122, 123–124
- trypanosomiasis, 108–109, 136, 178
- tuberculosis, 33, 48–49, *plate 7*
- tube wells, 53
- tularemia, 135
- tumor suppressor genes, 307
- typhoid, 47
- typhus, 105, 353
- ultraviolet radiation
- effects and disease, 315*f*, 326
 - hazards, 209–210, 273, 315*f*
 - vitamin D deficiency and, 60–61
- unani* system, 435
- United Nations
- Joint Programme on AIDS, 383
 - Millennium Goals, 164, 181, 230
- United States
- Eight Americas, 286
 - health care system, 409–412, 428
 - HIV/AIDS in, 384, 385–386
 - mortality and cause of death, 284–288
 - new urbanism, 191–192
 - regional pattern of cardiovascular disease, 321–324
- units of observation, 455–457
- urbanization
- central place hierarchies, 355–356, 360*f*, 418–422, 427–428
 - cultural ecologies and disease, 188–191
 - dimensions of, 184–185, 186*t*
 - methods of growth, 185, 187
- See also* rural–urban migration
- urban/village cycle in yellow fever, 104
- utilization models, 11
- vaccinations, 49, 69, 158, 363, 373, 374
- variables/variability
- genetic, 37–40
 - in regression analysis, 474–478
- vector data (GIS), 83, 84, 85*f*, 463
- vector ecology, 105–110, 122–123
- vectors and vector disease systems
- arthropods, 11, 105, 108, 246
 - bubonic plague, 112–116
 - classification of, 71–72*f*
 - overview, 45–46
 - transmission chains, 101–102, 103
 - urbanization and, 190–192
 - water-based, 230–231
 - yellow fever, 103–104
- viremia, 102, 193, 353
- virus
- cross-species, 192, 375–376
 - as disease agents, 44–45, 68, 69
- visible light and biological rhythms, 210, 211–213
- vitamin D
- deficiencies, 60–61
 - ultraviolet radiation and, 210
- vitamin deficiencies, 56, 59–61, 295
- waste disposal
- hazardous, 260–261, 270
 - solid, 263–264, 275
- water-based transmittable disease
- dengue hemorrhagic fever, 125–127
 - malaria, 117–121
 - overview, 116–117
 - schistosomiasis, 127–131
- water-borne diseases, 50–51
- water factor (hardness), 320–321, 322, 324
- water impoundment, 182–184

- water-related infectious disease, 49–51
- weather, influences on health
 - air-borne life, 218, 229
 - atmospheric ionization, 219–220
 - atmospheric pressure, 218–219
 - definition and discussion, 208–209, 210, 215
 - extremely low-frequency (ELF) waves and infrasound, 220–221
 - passage of fronts, 221
 - precipitation, 229–231
 - temperature, humidity, and air movement, 216–217, 319
 - winds, 219
 - See also* biometeorology
- Web resources, 91, 181, 233, 379, 383, 430, 467
- wells, 32, 53
- West Nile Virus (WNV), 91, 93*f*, 106, 108
- Whitehall Studies, 34–35
- winds, 219
- Wingspread Statement, 330–331
- women
 - behavior factors, 298–300
 - gender differences in studies, 294–295
 - habitat factors, 297–298
 - health and status, 151, 162, 166, 175, 297
 - HIV/AIDS and, 384, 385, 393–394
 - maternity and reproduction, 295–296
 - mortality and cause of death, 284–286, 299
 - population factors, 295–297
 - vitamin D deficiency in, 61
 - See also* gender differences
- World Health Organization Statistical Information Sytem (WHOSIS), 181
- World Health Organization (WHO), 18, 48, 99, 393
- worms, 68, 69, 70
- xerophthalmia, 60, 295
- years of life lost (YLL), 161
- yellow fever, 75, 103–105
- Zelinsky's three destinations and mobility, 156–157, 168
- zip code (postal) areas, 87, 341–342, 415, 425–426*f*, 428
- zonation of climates and biomes, 244–246
- zoonosis, 100, 105–106, 111, 195–196

About the Authors

Melinda S. Meade, a medical geographer, is Professor of Geography at the University of North Carolina at Chapel Hill. She is also Adjunct Professor of Epidemiology, a member of the Institute of Environment, and a fellow of the Carolina Population Center. In addition to the first and second editions of *Medical Geography*, Dr. Meade has published journal articles on the disease ecology of tropical Asia and the United States. She is a member of the Association of American Geographers and other geographic, Asian studies, and population associations.

Michael Emch, a medical geographer, is Associate Professor of Geography at the University of North Carolina at Chapel Hill, where he directs the Spatial Health Research Group. He is also Adjunct Associate Professor of Epidemiology and a fellow of the Carolina Population Center. Dr. Emch has published journal articles on disease ecology, primarily of infectious diseases of the tropical world. He is a member of the Association of American Geographers and is also an advisory editor for the international journal *Social Science and Medicine*.

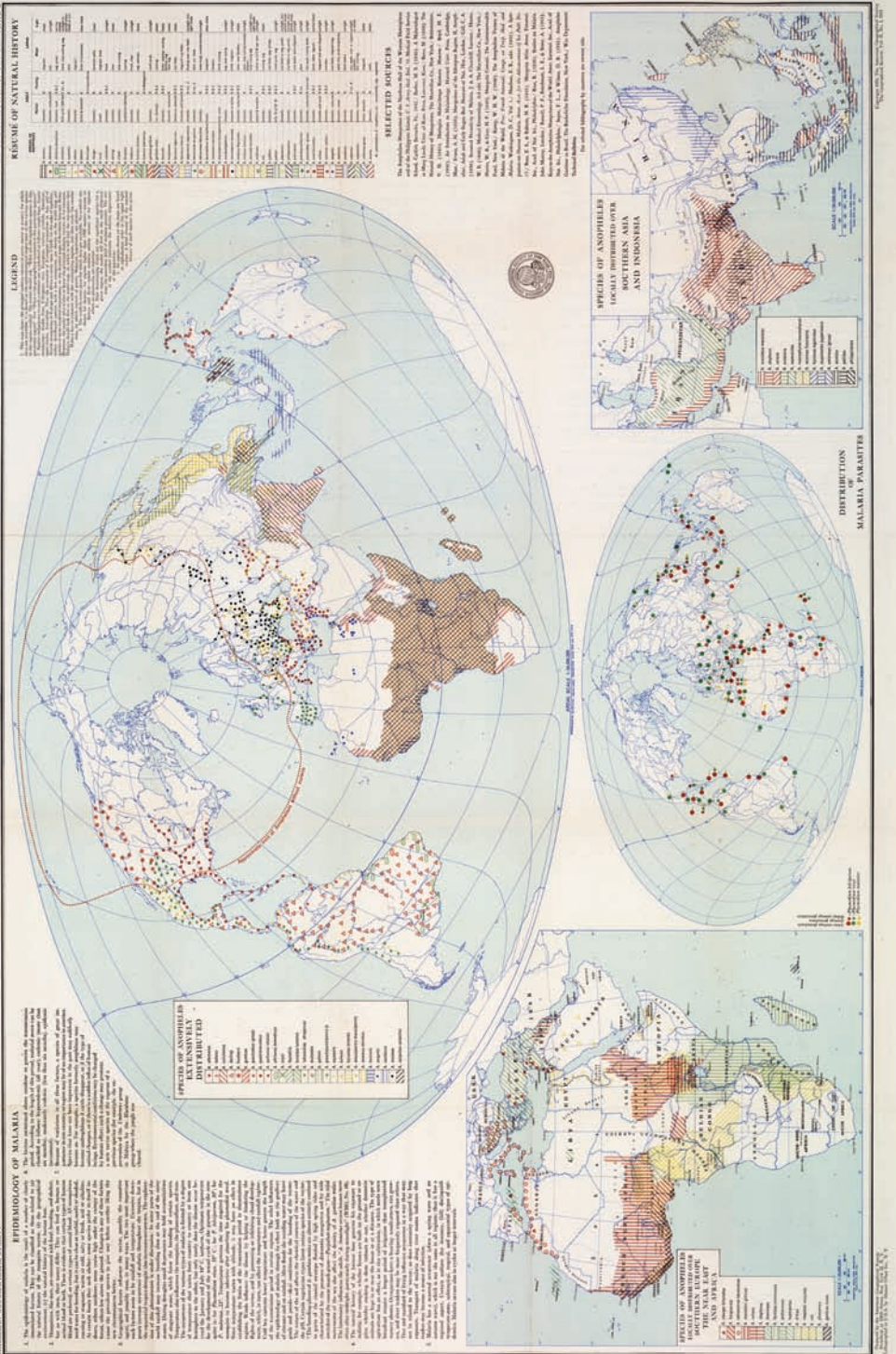


PLATE 1. May's map of the distribution of species of *Anopheles* mosquito vectors of malaria, inserted into the *Geographical Review* in 1951. Copyright 1951 by the American Geographical Society.

AMERICAN GEOGRAPHICAL SOCIETY



PLATE 2. May's map of starvation, diet, and deficiencies, inserted into the *Geographical Review* in 1953. Copyright 1953 by the American Geographical Society.



PLATE 3. Household type and vulnerability. *Upper left:* Rural Bangladesh (2008). Houseboats where people live permanently. The river is a reservoir for cholera. Cooking is done on small stoves fueled with wood or leaves; these stoves give off a lot of smoke, which is inhaled. *Upper right:* Northwest Frontier Province, Pakistan, near Afghanistan border (1988). Mud and stone house where cracked walls could be habitat for sand flies that transmit visceral leishmaniasis (kala-azar). *Middle left:* Northeast Thailand (1968). Local (*lao*) houses with typical architecture—open porch/house side, on stilts (to shade animals and weaving, etc.), now with a corrugated iron roof; burning garbage pit; water jar on porch. *Middle right:* Southern Belize (2005). A virtually identical house to the northeast Thailand house on stilts, but almost 40 years later and halfway across the world. *Lower left:* Southern Belize (2005). Mayan house with thatched roof, pigs in courtyard. *Lower right:* Chapel Hill, North Carolina (2008). House with screens, central heating, central air conditioning.



PLATE 4. Resources for drinking water. *Upper left:* Guanxi, China (2006). Dug well with hand pump. Cement cover to protect the water. *Upper right:* Rural Malaysia (1974). High school lesson for making a safe well: Common dipper, so individual pails don't contaminate; protected by fence, so chickens, other animals, and small children don't fall in. But really useless, with water table so high and contaminated water so near. *Lower left:* Rural Punjab, India (1984). Residential tube well in Punjab as Green Revolution prosperity transformed water safety. *Lower right:* Rural Bangladesh (2007). Shallow tube well painted red because of high arsenic concentrations. Arsenic occurs naturally in the shallow aquifer in much of rural Bangladesh and causes cancer when water from these wells is drunk over long periods.

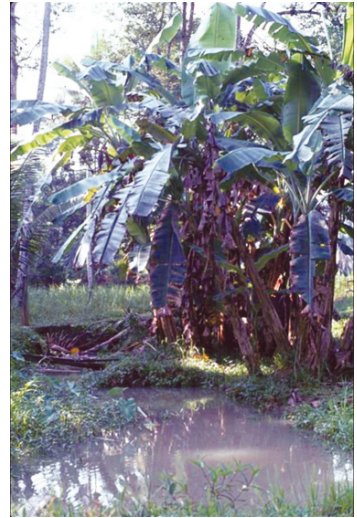


PLATE 5. Sanitation systems. *Upper left:* Guanxi, China (2006). Latrines at a school where a typhoid fever outbreak occurred. *Upper right:* Rural Malaysia (1974). This banana-shaded private place was designated in the national census as a latrine shared by three households. It drains right into the water! *Middle left:* Rural Bangladesh (2008). Latrine draining into pond that is often used for bathing or washing dishes. *Middle right:* Thailand (1974). Thai "squat toilet," a common type of Asian floor toilet. It's water-sealed, but jar must be filled with water from well to fill bowl for flushing; jar of water breeds *Aedes aegypti*. *Lower left:* Chapel Hill, North Carolina (2008). Water-sealed, water-saving toilet. *Lower right:* Jaipur, India (1984). Alley designated locally for defecation.

Malaria Cases in 2000–2003

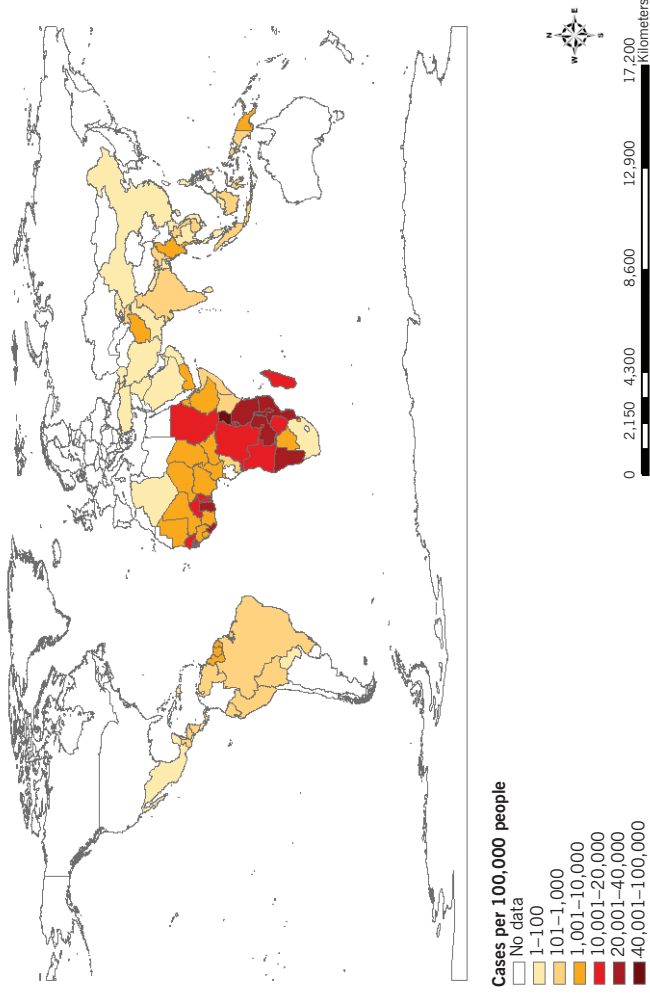


PLATE 6. Sub-Saharan Africa has the highest malaria prevalence.

Tuberculosis Cases in 2006

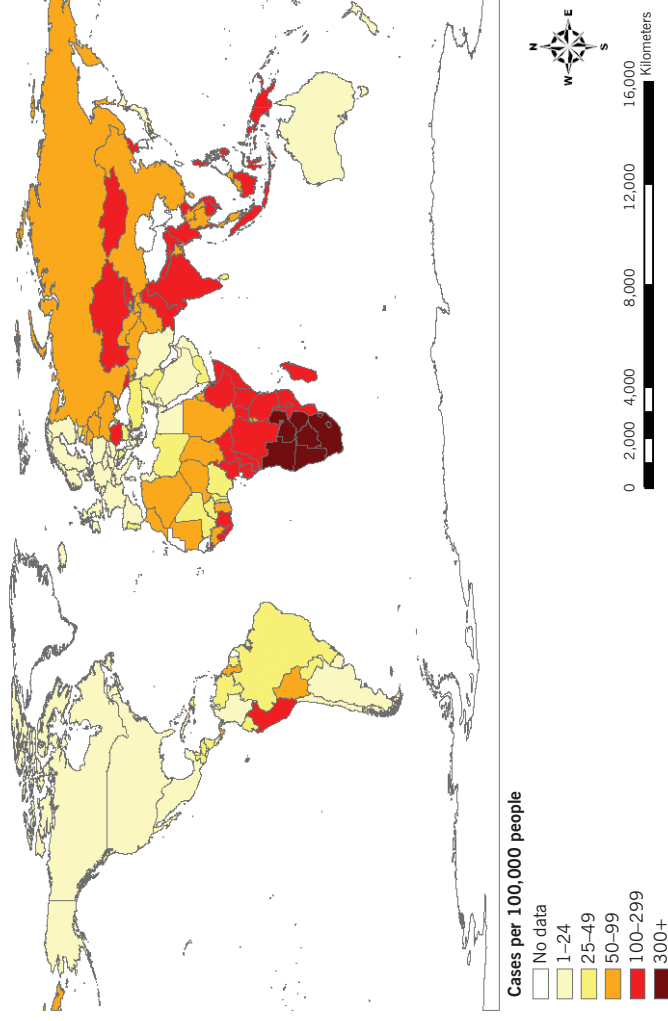


PLATE 7. Sub-Saharan Africa also has the highest tuberculosis prevalence. Southern Africa is especially hard hit. Asia also has quite high tuberculosis prevalence levels.

Estimated Number of HIV Cases in 2007

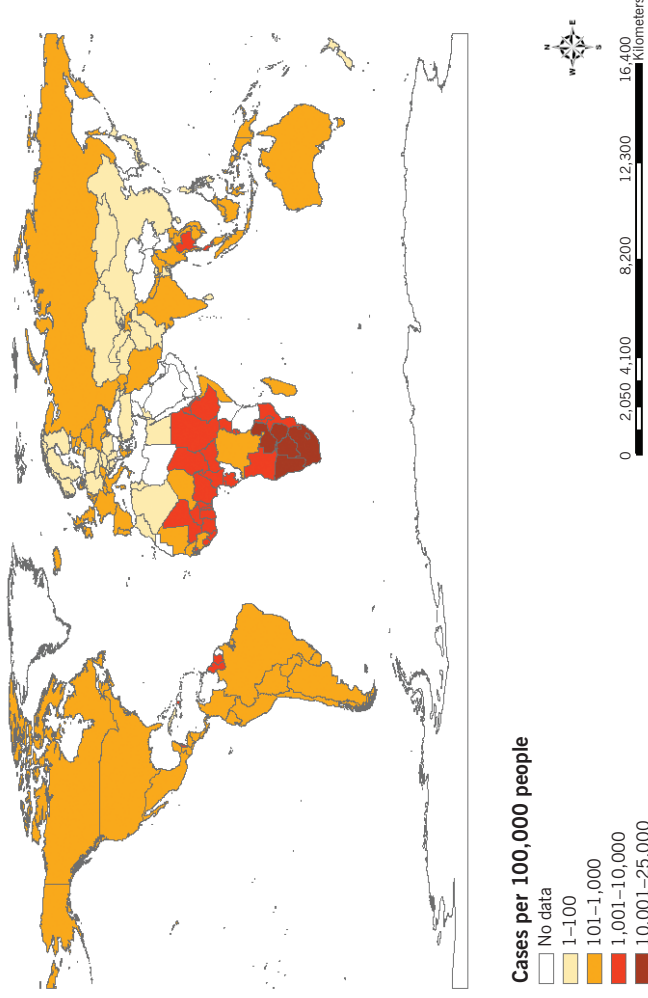


PLATE 8. HIV/AIDS is a global epidemic; however, sub-Saharan Africa has the highest HIV prevalence. Southern Africa is especially hard hit.



PLATE 9. Conditions of the agricultural frontier in Malaysia. *Clockwise from top left:* primary rain forest in west Malaysia; the same site months later, clear-felled; cloned rubber saplings, African nitrogen-fixing ground cover, and houses being erected by Malaysian Chinese contractors; land scheme occupied with rubber-farming settlers on home lots growing tapioca for “vegetables” amidst invasive *Imperator cylindrica* grass.

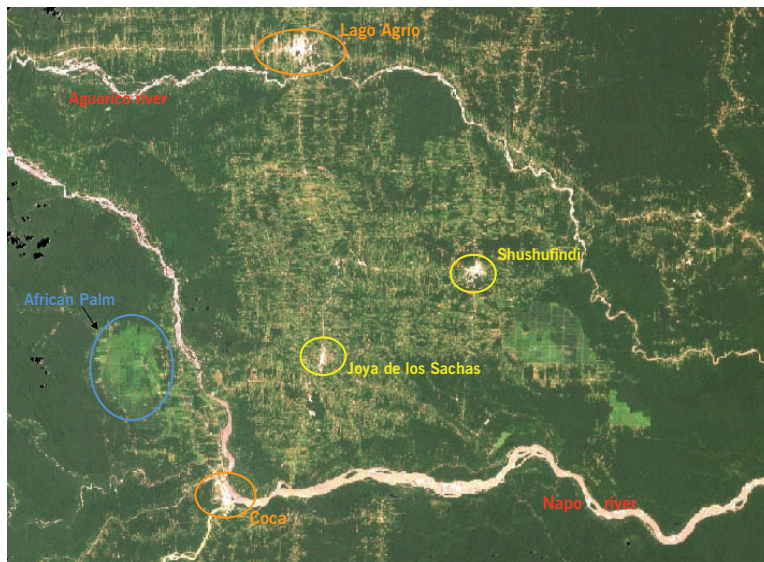


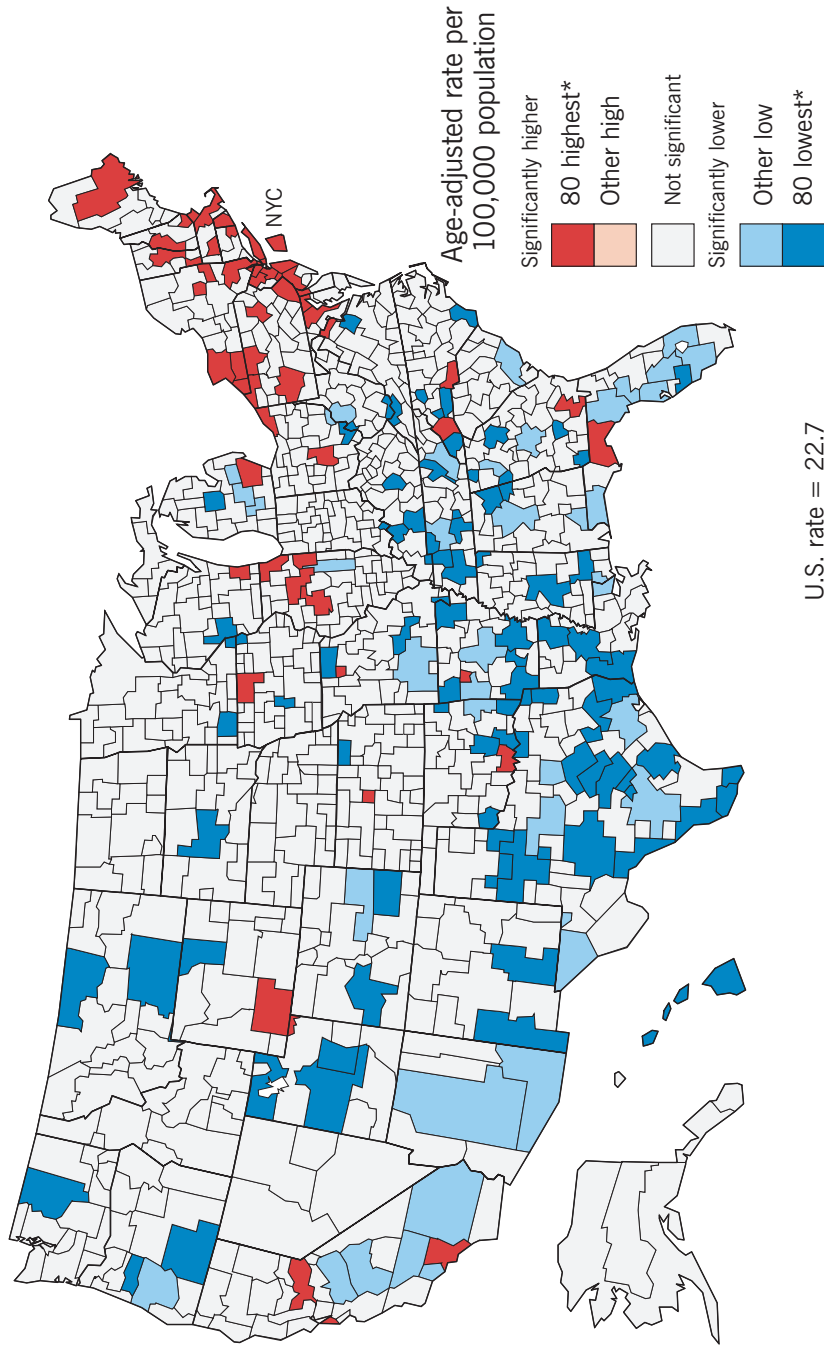
PLATE 10. Ecuadorian Amazon agricultural frontier. *Upper left and right:* House types in the agricultural frontier of the northern Ecuadorian Amazon. Left picture is a cement house of the *colonos* (settlers); right picture, a wooden house. *Center:* Land use in the northern Ecuadorian Amazon (Landsat image, September 2002). Lago Agrio and Coca (Francisco de Orellana) are urban areas; Joya de los Sachas and Shushufindi are rural settlements. Ecuadorian Amazon Project, R. E. Bilsborrow and S. J. Walsh, Co-Principal Investigators. Spatial Analysis Unit, UNC, Chapel Hill. *Lower:* Raw settlements along road in the northern Ecuadorian Amazon. Fieldwork by Patricia E. Polo, July 2007.



PLATE 11. Conditions of developing cities. *Left:* Bangkok, Thailand, in 1966 (*upper*), 1974 (*middle*), and 1991 (*lower*). *Top right:* A street in Bombay (now Mumbai), India, 1986. *Middle right:* A settlement in Dhaka, Bangladesh, 2006. *Lower right:* An open-ditch habitat in Lilongwe, capital of Malawi, 2008.

DEATH RATES OF EACH HSA COMPARED WITH U.S. RATE

BREAST CANCER WHITE FEMALE

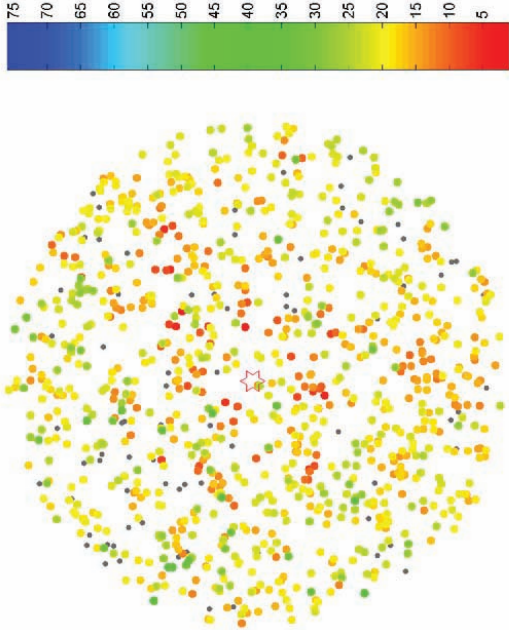


SOURCE: CDC/NCHS

* See text

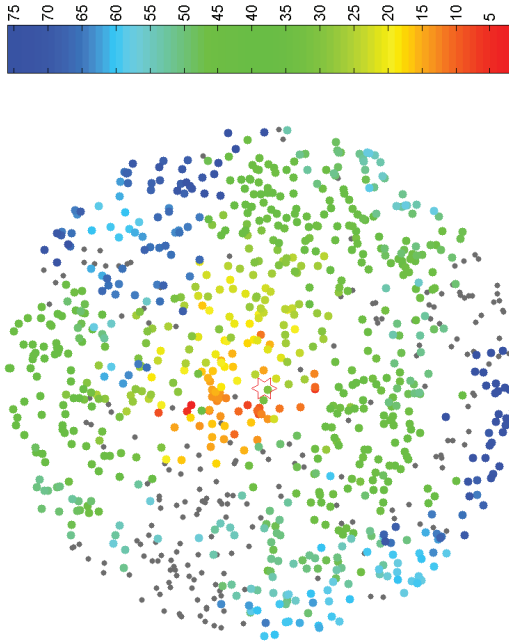
PLATE 12. White female breast cancer mortality age-adjusted rates per health service area. Map shows health service areas with significantly higher and lower rates than U.S. rate, emphasizing the 80 highest and lowest. From the National Center for Health Statistics, *Atlas of United States Mortality* (1997).

Spatial and Temporal Pattern



Dispersed social network

Family and coworkers are separate

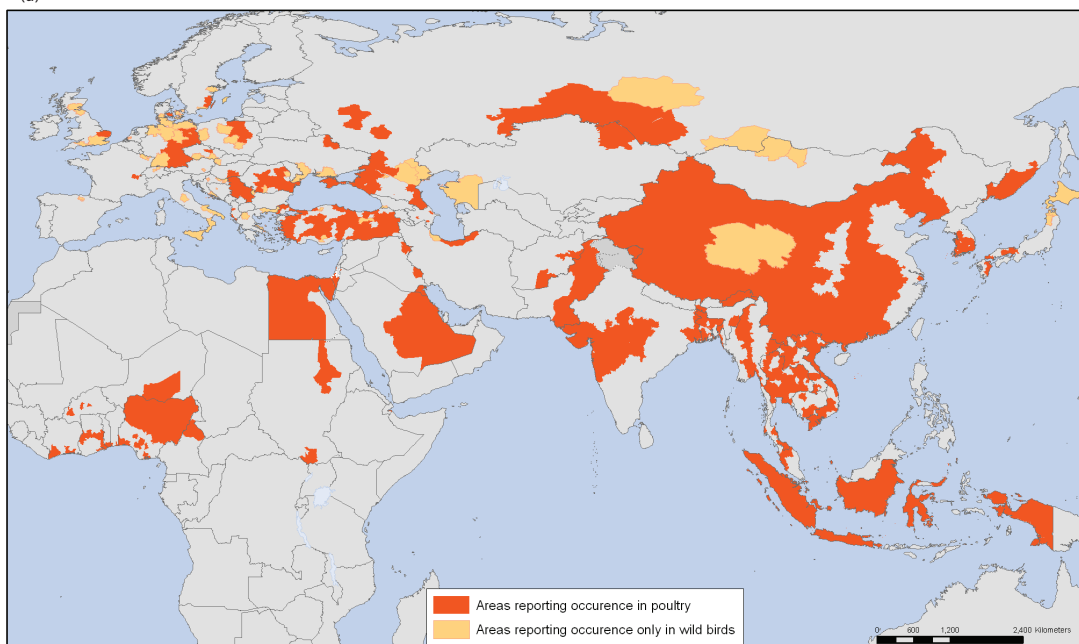


Clustered social network

Family and coworkers are together

PLATE 14. Dispersed and clustered social networks. With the first case introduced into a community (at the center of the graph), the figures show the location and date of infection. The legend bar indicates the number of days for any individual to become infected. Each dot is an individual; the red ones are infected early, and the blue ones late. The two dispersion patterns are results of different social networks of human contact. Used by permission of Ling Bian.

(a)



(b)

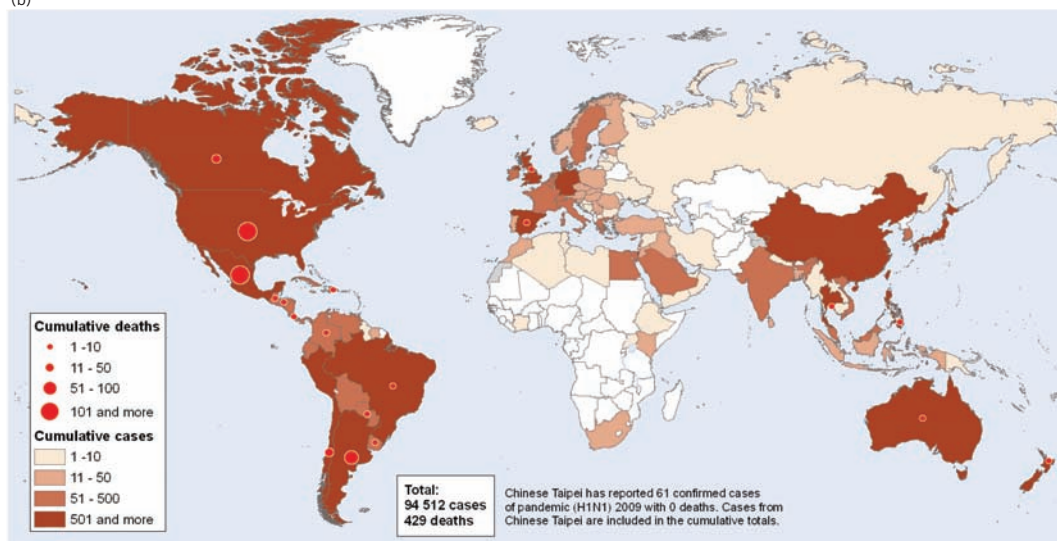
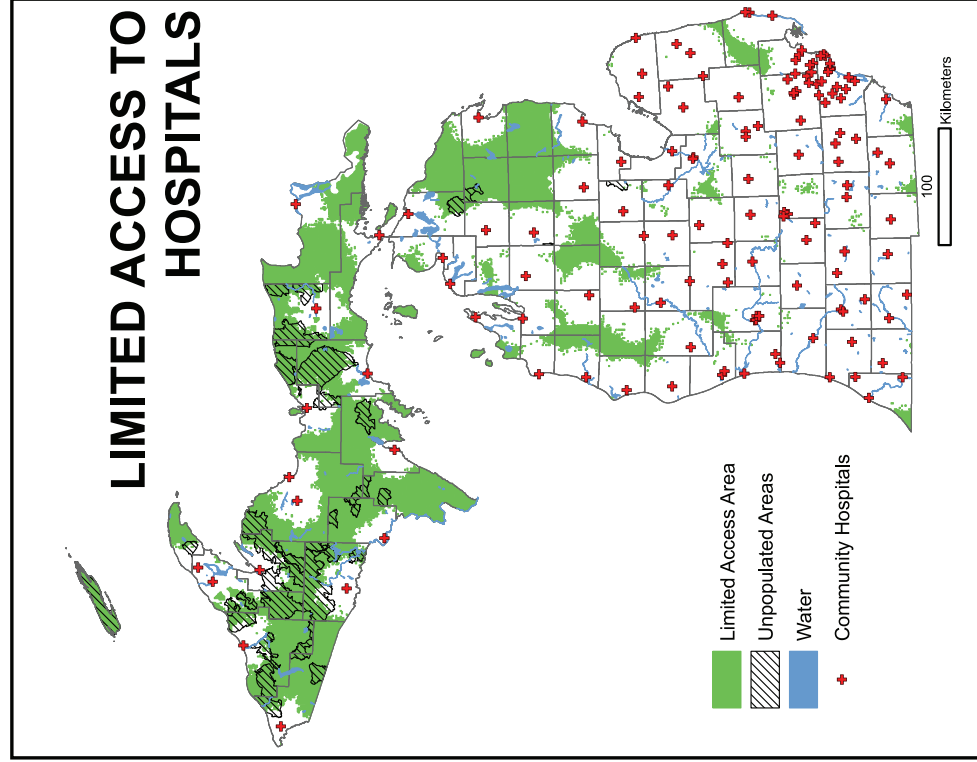


PLATE 15. (a) Areas reporting H5N1 avian influenza cases from 2003 to November 2008. From WHO (gamapservr.who.int/mapLibrary/app/searchResults.aspx). (b) Cases of H1N1 influenza in July 2009. From WHO (www.who.int/csr/don/h1n1_20090706a_1100.png).

(a)



(b)

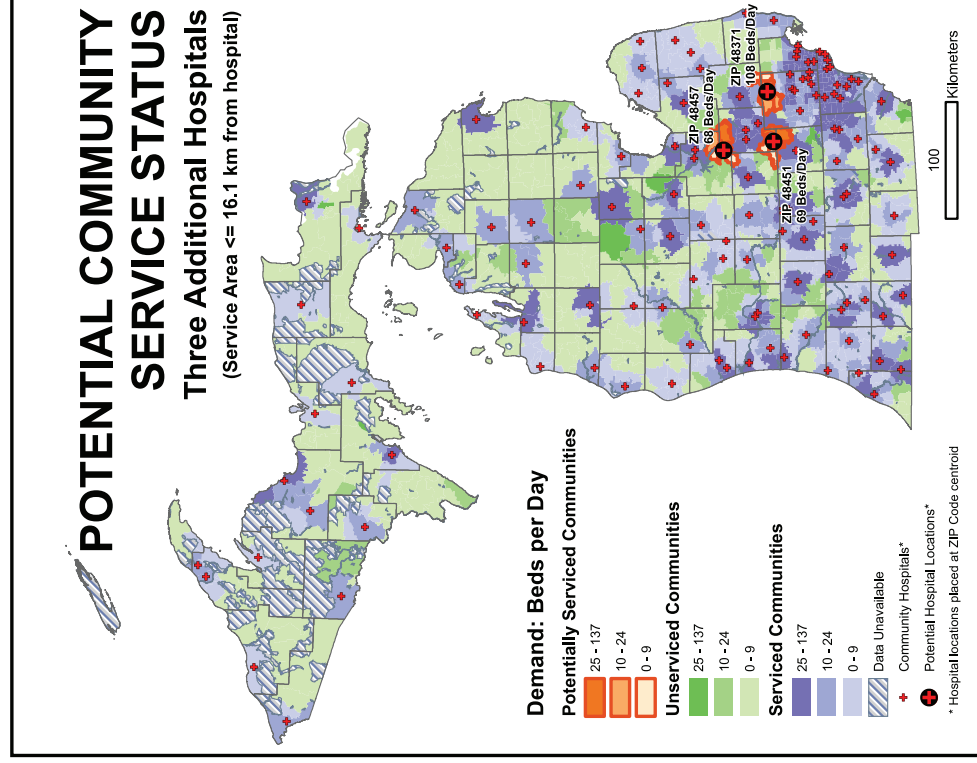


PLATE 16. (a) Areas with limited access to hospitals in Michigan. (b) Potential hospital locations in Michigan. From Messina, Shortridge, Groop, Varnakovida, and Finn (2006). Copyright 2006 by BioMed Central Ltd., part of Springer + Business Media. Reprinted by permission.