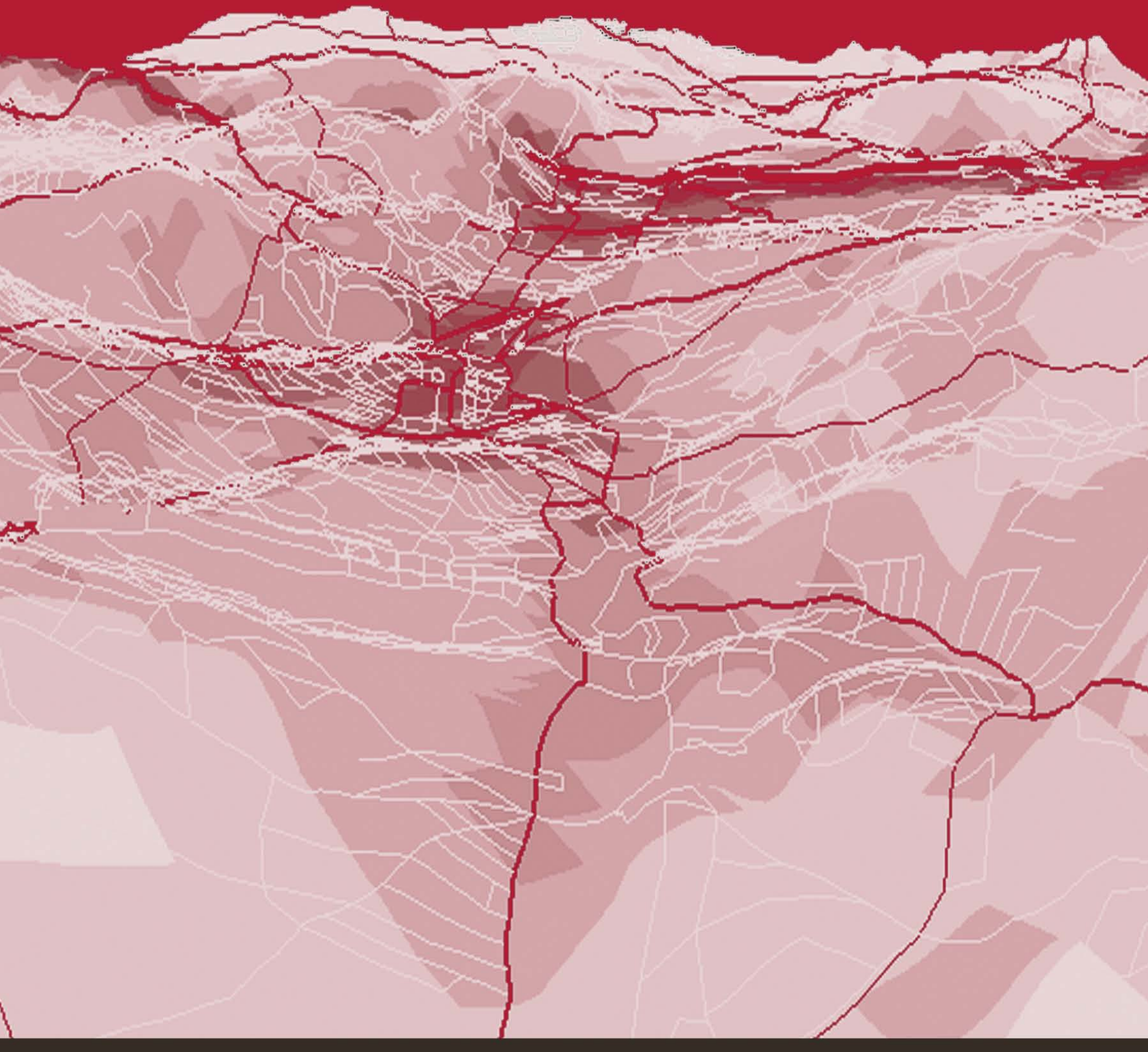
GIS in Public Health Practice

Edited by

Ravi Maheswaran

Massimo Craglia



CRC PRESS

GIS in Public Health Practice

Edited by
Ravi Maheswaran
Massimo Craglia

CRC PRESS

Boca Raton London New York Washington, D.C.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20141006

International Standard Book Number-13: 978-0-203-72034-9 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Preface

Public health has been defined as the "science and art of prolonging life, preventing disease, and promoting health through organized efforts of society." The geographical perspective is a key aspect of public health. Populations and communities are geographically distributed and communities tend to have their own defining characteristics. Factors influencing health are commonly classified under four groups:

1. Inherited conditions
2. Environment, which includes both physical (i.e., air quality, water quality, soil characteristics, radiation) and socioeconomic aspects
3. Lifestyle
4. Healthcare

Each of these factors may have marked geographical variation. The practice of key elements of public health, including communicable disease control, environmental health protection, health needs assessment, planning and policy, surveillance, monitoring and evaluation, and operational public health management, is often explicitly geographical in nature. In addition, resource allocation at the macro and micro levels has a strong geographical component based on demography, health needs, existing provisions, and other factors. GIS, the definition of which has evolved from geographic information systems to geographic information science, involves a scientific problem-solving approach, encompassing the development and application of scientific methods to solve societal problems. It, therefore, has become an integral and essential part of public health research and practice.

Significant advances in scientific approaches to evaluating and using geographic information are taking place. Health information at a fine spatial resolution has become widely available; the same can be said for mapping technology. These developments enable public health practitioners to link and analyze data in new ways at the international, regional, and even street levels. As part of the drive to promote the use of GIS within public health, the European Commission supported the First European Conference on Geographic Information Sciences in Public Health held in Sheffield, United Kingdom, in September 2001. The scientific program drew upon many of the leading public health researchers and practitioners in this area. The breadth of knowledge and expertise at the conference and the clear interest in the field from practitioners from a wide variety of specialisms led us to believe that a book that recognized the breadth of the field would be useful. For this book, specifically selected expanded contributions were invited from participants to illustrate particular areas of application or address issues pertinent to the field. Further chapters were sought from other specialists to cover specific aspects. Many of the chapters have a United Kingdom or European focus, but the principles, issues, and methods discussed should be equally relevant beyond Europe. Although this is not the first book on GIS and public health, we believe it is the first to treat GIS as more than a technology in relation to public health practice. We hope it will be of benefit to practitioners, researchers, and students with an interest in public health.

Dr. Ravi Maheswaran, MD, MRCP, MFPHM
Clinical Senior Lecturer in Public Health Medicine

Dr. Massimo Craglia, PhD, MRTPI
Senior Lecturer in Town and Regional Planning

Acknowledgments

We would like to thank Jane Shields for providing administrative support and preparing the manuscript. We would also like to thank all the contributors for their efforts and support. Acknowledgments of permission to reproduce published material are provided in the corresponding chapters. The Public Health GIS Unit at the University of Sheffield was established with core funding from Trent Regional Health Authority. The unit received a grant from the European Commission's Fifth Framework Programme for promoting collaborative work in the public health GIS field, which included hosting the First European Conference on Geographic Information Sciences in Public Health in 2001.

Contributors

Alessandro Annoni
Institute for Environment and Sustainability
Joint Research Centre of the European Commission
Ispra (VA), Italy

Paul Aylin
Department of Epidemiology and Public Health
Imperial College School of Medicine
London, United Kingdom

Deryck Beyleveld
Department of Law
University of Sheffield
Sheffield, United Kingdom

Alex Bottle
Department of Epidemiology and Public Health
Imperial College School of Medicine
London, United Kingdom

David Briggs
Department of Epidemiology and Public Health
Imperial College School of Medicine
London, United Kingdom

Paul Brindley
Sheffield Centre for Geographic Information and Spatial Analysis
University of Sheffield
Sheffield, United Kingdom

Samantha Cockings
Department of Geography
University of Southampton
Southampton, United Kingdom

Massimo Craglia
Department of Town and Regional Planning
University of Sheffield
Sheffield, United Kingdom

Kees de Hoogh
Department of Epidemiology and Public Health
Imperial College School of Medicine
London, United Kingdom

Gregory A. Elmes
Department of Geology and Geography
West Virginia University
Morgantown, West Virginia

Robert P. Haining
Department of Geography
University of Cambridge
Cambridge, United Kingdom

Robin Haynes
School of Environmental Sciences
University of East Anglia
Norwich, United Kingdom

John Holmes
Public Health South
Dunedin, New Zealand

Edmund Jessop
Department of Public Health
West Surrey Health Authority
Surrey, United Kingdom

Thomas Kistemann
Institute for Hygiene and Public Health
University of Bonn
Bonn, Germany

Andrew B. Lawson
Department of Epidemiology and Biostatistics
School of Public Health
University of South Carolina
Columbia, South Carolina

Andrew Lovett
School of Environmental Sciences
University of East Anglia
Norwich, United Kingdom

Markku Löytönen
Department of Geography
University of Helsinki
Helsinki, Finland

Ravi Maheswaran
Public Health GIS Unit
Section of Public Health
School of Health and Related Research
University of Sheffield
Sheffield, United Kingdom

Tim Pearson
Public Health GIS Unit
Section of Public Health
School of Health and Related Research
University of Sheffield
Sheffield, United Kingdom

Angela Queste
WHO Collaborating Centre for Health Promoting Water Management and Risk Communication
Institute for Hygiene and Public Health
University of Bonn
Bonn, Germany

Clive E. Sabel
Department of Geography
University of Canterbury
Christchurch, New Zealand

Darren Shickle
Section of Public Health
School of Health and Related Research
University of Sheffield
Sheffield, United Kingdom

Ralph Smith
West Midlands Cancer Intelligence Unit
University of Birmingham
Birmingham, United Kingdom

Rupert Suckling
Section of Public Health
School of Health and Related Research
University of Sheffield
Sheffield, United Kingdom

Gilla Sünnenberg
School of Environmental Sciences
University of East Anglia
Norwich, United Kingdom

David Townend
Department of Law
University of Sheffield
Sheffield, United Kingdom

Susan Wallace
Section of Public Health
School of Health and Related Research
University of Sheffield
Sheffield, United Kingdom

Stephen Wise
Sheffield Centre for Geographic
Information and Spatial Analysis
University of Sheffield
Sheffield, United Kingdom

Contents

SECTION 4 Data Protection and E-Governance Issues in Public Health

1 Introduction and Overview

Ravi Maheswaran and Massimo Craglia

CONTENTS

1.1 INTRODUCTION

Public health has been defined as "the science and art of preventing disease, prolonging life, and promoting health through the organized efforts of society." This definition was arrived at in the inquiry established to consider the future development of the public health function including the control of communicable disease in England (Acheson, 1988). The inquiry was set up following failures in the system to protect the health of the public from two major outbreaks of communicable disease caused by salmonella and Legionnaires' disease. Since then, a number of health scares have highlighted the need for continuing improvements in public health protection systems. Recent high profile examples include the outbreak of severe acute respiratory syndrome (SARS) and variant Creutzfeldt-Jakob disease, the human form of bovine spongiform encephalopathy commonly known as mad cow disease. The description of disease epidemiology typically has three elements: time, place, and person. Describing the outbreak and spread of a communicable disease therefore explicitly includes a spatial component. Although this has long been recognized (e.g., the investigation of cholera outbreaks in London by John Snow), an important barrier to examining the spatial element of disease outbreaks has been the lack of both digitized spatial data and the computer tools for mapping and spatial analysis.

Environmental health issues have been gaining importance both scientifically and in terms of public concern. These issues range from the global scale — for example, in terms of the health effects of climate change — to the local level where

0-415-30655-8/04/$0.00+$1.50

there may be concern about an increase in cancer around a chemical factory. The interaction between environment and health outcomes is being increasingly recognized at the international level by major organizations including the World Health Organization (WHO) and the European Commission. For example, a recent report for the European Commission estimated that up to 20 percent of ill health might be due to environmental factors, with air pollution being the worst culprit (Commission of the European Communities, 2003). It was estimated that 10 percent of children in Europe have asthma, with a higher incidence in Western Europe than in accession countries, indicating a complex mix of environmental, social, and lifestyle factors. This is not just an environmental and social problem, but also an economic one. In the United Kingdom alone, the total annual cost of asthma is estimated at over EUR 3.9 billion (van Tongelen, 2003).

Understanding the complex relationships between combined exposure over time to a cocktail of chemicals and emissions and health outcomes is becoming a top priority in Europe, requiring integrated research programs with greater availability of shared information and geographically referenced information systems. In addition, there are increased concerns to the public, organizations, and governments of the risks to health posed by the potential of bioterrorism following the events of September 11, 2001. Health protection establishments have been strengthened in numerous countries throughout the world and part of the requirement for active surveillance is the temporal-spatial component for early warning systems.

Although the potential benefits of spatial information and analysis are immediately apparent in the communicable disease and environmental health fields, there are major and not fully recognized benefits to be gained in relation to policy and planning in the health and healthcare fields. The basic elements of work in this branch of public health work include health needs assessment, planning and implementation, and monitoring and evaluation. Related and overlapping aspects include resource allocation, surveillance, and health impact assessment.

There are a number of approaches to health needs assessment (Stevens and Raftery, 1997). Epidemiologically based needs assessment combines epidemiological approaches, including health status assessments, with assessment of the effectiveness and cost effectiveness of interventions. Comparative needs assessment involves comparing levels of service utilization between different populations while corporate needs assessment incorporates obtaining the views of professionals, patients, and other interested and relevant parties. There is a clear spatial element to population-based health needs assessment and the use of Geograhic Information Systems (GIS) in this field will bring many potential benefits. These include resource allocation based on need, and current practice in resource allocation in most countries already includes a geographical element based on regions or smaller areas.

Following a population-based needs assessment, the next step in public healthcare management is to plan and implement strategies and interventions to meet these needs to improve health and well-being. If the health needs assessment has a detailed spatial element, then appropriate geographical targeting of interventions or tailoring interventions to meet varying geographical factors could substantially improve the effectiveness of interventions and efficiency in the use of scarce resources. The interventions may be at the structural level, such as building new healthcare facilities, or at the human resource level, such as increasing the number of specialists per head of population. Interventions to increase geographical access may include the provision of community transport services. Geographical techniques, such as spatial decision support systems, are currently already widely deployed in the commercial and retail sectors. Tailoring interventions may be improved by detailed spatial analysis of existing referral patterns and care pathways, though the latter may pose challenges in the representation of relevant spatial and temporal information within a Geographic Information System (GIS). The need for targeted assessment and policy action is also particularly important given the widely acknowledged relationship between poor health and socioeconomic conditions (see, for example, Acheson, 1998).

The logical step that follows intervention is monitoring and evaluation to ensure that the intervention is achieving the desired goal. Again, the spatial element may dictate that goals should be set at appropriate levels for different geographical areas. Monitoring and evaluation is often underresourced as there is a tendency to move on quickly to a "new" initiative to satisfy political requirements, rather than undertaking careful assessment of the effectiveness of existing policies and interventions. Although cognizant of the political environment in which health policy is couched, we cannot but stress the importance of this part of the policy cycle and the contribution that a spatial perspective brings, particularly when looking at the cumulative effects of cross-sectoral policies.

Health impact assessment has been defined as the estimation of the overall effects of a specified action on the health of a defined population (Scott-Samuel, 1998). These actions may range from something specific such as a new municipal incinerator or landfill site to much larger and complex projects, programs and policies such as housing development projects, urban regeneration programs, and integrated transport policies. GIS has a useful role to play in health impact assessment, both in terms of descriptive and analytical work and in terms of visual presentation of information and evidence to a wider audience to promote public participation. The importance of health impact assessment is being increasingly recognized internationally, including at the European level with the launch by the European Commission of the Integrated Impact Assessment Tool in 2003 (see Chapter 14).

Public need for information is also on the increase and can be supported by the greater availability of information and communication technologies and the Internet in particular. Health-related information has been identified as one of the main application areas for E-government strategies across the world. Although some online services providing general health information (e.g., information on symptoms and potential treatments) are not location specific, there is a wide range of services underpinned by geography, such as the location of the nearest pharmacy, health center or dental practice.

Current policy pressures and requirements indicate that the geographical dimension of public health research and practice is being increasingly recognized. This converges with another set of developments from the technological and geographical information communities. We can distinguish here three main developments. From a technological perspective, there is increasing diffusion of the Internet and related technologies and these have started to become embedded into organizational processes and the daily lives of millions of individuals worldwide. There are therefore increased opportunities for information sharing and dissemination in every field including health. In parallel, we are seeing increased availability of geographically referenced information from public and private sectors, facilitating data acquisition, visualization, and analysis. Again, this is a generic trend supported by policies such as E-government and more generally the development of the Information Society, across Europe, for example (see Chapter 14 for a discussion of this phenomenon). Internet services that deliver maps are enjoying significant increases in popularity across Europe and are set to increase. A report in April 2003 in Italy, for example, highlighted several Web sites with maps that had double-figure increases in the number of hits over a three-month period.[1]

Specific technologies to integrate, analyze, and display geographic information i.e., GIS) have become much more widespread. Systems that until 15 years ago were still very expensive and within the domain of a few experts have now become much cheaper and widely available, spawning an industry estimated to turn over approximately US$1 billion per annum worldwide (www.daratech.com) in software sales alone. Thus, the basic techniques of geographical query and analysis have become available to many and in some instances directly available from basic spreadsheet software packages. This is reflected in the increased use of GIS in public health practice, a response to the requirements and needs identified above.

Against these opportunities, a number of important constraints need to be recognized. First, in spite of the increased technological prowess of current systems, there are still major inequalities in the opportunities to access information (i.e., physical access) and make effective use of it (i.e., social access requiring knowledge and skills). The digital divide is in many instances increasing and cannot be underestimated. Second, although information in general and geographic information in particular is becoming more widely available, it has also been constrained by lack of interoperability (different geographical scales, formats, definitions, and so on). This is particularly acute at the transnational level, with policy frameworks defining conditions of access that are often opaque and unduly restrictive. Furthermore, most organizations in both public and private sectors have yet to come to grips with the full impact of having digital information, as opposed to paper records, and have yet to revise their policies and operations to move into the digital age. For example, most government organizations have yet to define consistent and transparent access and pricing policies and continue to live in a culture of "my" information that undermines information sharing needs. Although this applies to all types of information, geographic and health information pose particular challenges. Geographic information is an expensive commodity to collect and maintain and therefore raises potential conflicts between the pursuit of social objectives, which call for wide dissemination, and economic objectives that call for profit maximization. Health information may be of a very sensitive nature and its use and misuse could lead to substantial problems if appropriate safeguards are not in place to protect confidentiality.

[1] Nielsen Netratings, press release for Italy 20/05/2003, http://www.nielsen-netratings.com/pr/pr_030520_italy.pdf.

The wide range of issues highlighted above and the convergence between public health requirements and GIS opportunities are at the origin of this book. It takes a broad view of GIS, not confined to the technology perspective (the S in GIS), but couched in recent views of GIS as a science: a body of knowledge and theory regarding geographic information (Goodchild, 1992; Longley et al., 2001). GIS addresses important issues and questions (e.g., issues of representation, relationships, data models and structures, visualization) in an application-orientated setting and the technology is only part of the approach toward finding solutions to the problems at hand. A wide range of disciplines including epidemiology, statistics, biomedical science, sociology, management science, and health economics underpins public health itself. Other subjects relevant to public health research and practice include ethics, law, and public policy. Hence the intersection between GIS (as a science) and public health requires, in our view, a broad treatment to explore a variety of issues and not be confined to a purely application view of the problem (as seen from a GIS perspective) or a spatial perspective of the subject (as seen from the public health angle). With these considerations in mind, this book is structured into four sections addressing the following sets of issues:

1. Disease mapping and spatial analysis
2. GIS applications in communicable disease control and environmental health protection
3. GIS applications in healthcare planning and policy
4. Data protection and E-governance issues in public health

This book focuses on some of the issues and topics particularly pertinent to current research and practice in the public health GIS field and presents examples that illuminate useful applications in this area of work. It is aimed at practitioners and researchers in the public health field and all sections have strong input from practitioners or academics working in practice. In addition, this book should appeal to a wider audience of people with multidisciplinary interests in health outcomes, policy, and practice. This book will also be relevant to students in masters and doctorate programs in public health and epidemiology. Public health practitioners will be especially interested in the methods used in the different applications, but also in getting a critical appraisal of some of the limitations, including issues of data availability, data protection, and impacts of current policy initiatives to provide an increasing number of services via the Internet.

Throughout this book, the acronym GIS is used interchangeably, referring to geographic information system (or systems) or to geographic information science (or sciences), the latter acknowledging the range of sciences underpinning public health that have a geographical perspective. This book complements others in the GIS and public health field (Cromley and McLafferty, 2002; de Lepper et al., 1995; Melnick, 2002). Existing books focus on technical aspects of GIS, while this book views the subject from a broader perspective as described above. In particular, the sections on disease mapping and spatial analysis and data protection and E-governance issues, described below, cover aspects that a focus on GIS as a technology would traditionally exclude.

1.1.1 Disease Mapping and Spatial Analysis

This first section of this book has a strong methodological flavor and contains three chapters. Chapter 2 by Ravi Maheswaran and Robert P. Haining describes basic elements of geographical analysis, including an introduction to data types and models and issues of data quality. It then provides an overview of cartographic operations relevant to public health and defines the key elements and techniques of exploratory spatial data analysis and ecological studies. The importance of understanding data before using it comes out as a strong message from this chapter.

Chapter 3 by Andrew B. Lawson is a detailed description of disease mapping from a statistical perspective. The chapter moves logically from basic representations of disease distribution to the calculation of rates and standardized rates, interpolation methods and models. The latter section includes a useful discussion of basic likelihood models, random effects, Bayesian models and other more recent developments in disease mapping. An important message from this chapter is the need to be aware of the potential for misrepresentation inherent in some of the most widely available and commonly used "simple" methods of data analysis.

In Chapter 4, Clive E. Sabel and Markku Löytönen focus on disease clusters and methods for detecting clustering, making a useful distinction between the two. This is a good review of the strengths and weaknesses of alternative methods, some of which are illustrated with examples from Finland. The challenges identified by the authors include the need to integrate space and time into public health GIS research and to find appropriate ways to integrate comprehensive individual-level data, such as that available in some of the Nordic countries, into spatial analytical frameworks.

1.1.2 GIS Applications in Communicable Disease Control and Environmental Health Protection

This second section introduces an important application area with contributions from both researchers and practitioners. Chapter 5 by Thomas Kistemann and Angela Queste is an overview of GIS applications in communicable disease control, made all the more topical since the SARS outbreak in early 2003. It traces the historical routes to the spatial analysis of communicable disease and describes the opportunities for using GIS for the prevention, surveillance, and control of communicable diseases. Some of the pitfalls highlighted included the lack of robust statistical methods embedded within widely available standard GIS software and the lack of trained staff.

Chapter 6 by John Holmes provides a good example of the added value of a spatial perspective in outbreak investigation. It describes the analysis of recurrent outbreaks of *Salmonella* Brandenburg in sheep and humans over a five-year period in New Zealand. In this practical application, we see the opportunity afforded by GIS not only in terms of spatial analysis and visualization but also as a platform to enable the integration of data from different sources. As a result of the analysis, educational and other preventative measures have been put in place that have reduced considerably the extent and effects of such outbreaks, demonstrating the value of this type of analysis in a practical situation.

Chapter 7 is contributed by Kees de Hoogh, David Briggs, Samantha Cockings, and Alex Bottle of the Small Area Statistics Unit at Imperial College, London. The chapter draws on the experiences of this unit to illustrate a range of methods, packed with examples, to assess environmental exposure. The benefits of GIS are clear from these examples, as well as new challenges lying ahead with respect to the need to model individuals rather than areas and to develop dynamic rather than static models.

Chapter 8 by Paul Brindley, Ravi Maheswaran, Tim Pearson, Stephen Wise, and Robert P. Haining provides a good example of using modeled outdoor air pollution data for health surveillance based on a project they undertook in Sheffield. The chapter contains a detailed discussion of data interpolation methods for linking modeled air pollution estimates to areas and their populations to enable the assessment of the health impact of air pollution. It also addresses the issue of data quality of the modeled pollution estimates.

Chapter 9 by Paul Aylin and Samantha Cockings describes the value of establishing a rapid inquiry facility for the rapid initial assessment of apparent disease clusters and of the health impacts of point sources of environmental pollution in the United Kingdom. The chapter then presents preliminary experiences from the European Health and Environment Information System project, which involves setting up similar facilities in other European countries.

1.1.3 GIS Applications in Healthcare Planning and Policy

The focus on practical applications continues in this section with four chapters on healthcare planning and policy. Ralph Smith in Chapter 10 provides a valuable practitioner's perspective of the importance of regional structures for supporting GIS-based analysis for the National Health Service in the United Kingdom. The regional dimension is important with respect to the stability of institutions, availability of critical skills, data integration, and data analysis across counties and districts. This dimension is well recognized across Europe where regions play a very important role. This chapter adds weight to current debates in the United Kingdom and elsewhere on devolution and regional structures of governance.

Chapter 11 by Edmund Jessop continues the theme but from a smaller territorial entity: the district. The use of GIS in practice at this level is described well. What is important here is not technical sophistication of the methods deployed but the real issues that need addressing, from boundary maps and changes to travel time zones, as well as the practical use of some of the techniques described earlier in the book. Practitioners at the local level will find this chapter valuable for describing what GIS could do to their work.

Andrew Lovett, Gilla Sünnenberg, and Robin Haynes in Chapter 12 focus on one of the key aspects of public health research and practice: access to services. Using work undertaken by the authors in East Anglia, they present a clear view of the potential of GIS to evaluate accessibility to services using both public and private transport. The chapter discusses data issues and devotes most of its contents to the discussion of methods for data analysis. The value of calculating accessibility across the network rather than in straight lines is confirmed. The chapter discusses challenges in evaluating the complex mix of options and modes of travel currently available at different times of the day and days of the week.

Although the chapters thus far in this section have provided a United Kingdom perspective, Chapter 13 by Gregory A. Elmes gives an incisive overview of recent trends and developments in the use of GIS for public health practice in the United States. The chapter describes some of the broad characteristics of the U.S. health system and its use of GIS in public health planning with several examples of programs and initiatives at federal and state levels. The many opportunities for wider use of GIS in public health in the United States are reviewed and in many ways are similar to the situation in Europe. Against these opportunities, Gregory Elmes does well to warn the reader about the highly politicized nature of the healthcare system that may well inhibit the smooth integration of GIS into day-to-day practice.

1.1.4 Data Protection and E-Governance Issues in Public Health

Although the first three sections of the book have addressed broad methodological issues and applications, the four chapters in this section provide four complementary perspectives on policy, the legal framework, trust, and technology, respectively.

Chapter 14 by Massimo Craglia and Alessandro Annoni sets GIS and public health within the wider perspective of the development of the Information Society and E-government and the increasing focus on the complex relationships between environment and health across Europe. With particular regard to the initiatives launched by the European Commission, the chapter reviews current environmental and health policy requirements and the development of an infrastructure for spatial information in Europe. The conclusion to this chapter draws attention to the legal frameworks necessary to ensure data sharing and flows as well as confidentiality of the individual. It also emphasizes the importance of maintaining the trust of citizens and patients in their relationships with the healthcare system and more generally with the system of government.

The 1995 European Data Protection Directive provides a key aspect of the legal framework within which all GIS and health practitioners in the European Union have to operate. Deryck Beyleveld and David Townend discuss this directive in detail in Chapter 15. The authors introduce the directive and its objectives, requirements and exemptions, with a particular focus on informed consent. The important issue of anonymous data is discussed together with differences between the European Directive and its transposition into U.K. law.

Obtaining informed consent is crucial not only to respect the law but also to develop and maintain the relationship of trust with the data subjects, such as patients. The risks are otherwise incalculable for both researchers and the citizens themselves. Rupert Suckling, Darren Shickle, and Susan Wallace address this important issue in Chapter 16. Reviewing existing international literature, the authors analyze several aspects of public attitudes to the use of health information. Knowledge of existing rights is explored together with the extent to which health professionals are perceived as requiring access to sensitive information. The findings of this review are potentially constrained by the U.S. bias of available literature but provide a benchmark against which more recent studies (e.g., Spadaro, 2003) can be assessed.

Chapter 17 by Markku Löytönen and Clive E. Sabel reflects on the opportunities and limitations of recent technological advances in a range of fields including mobile telephony, remote sensing, and positioning systems. Their combination into a new generation of technologies and services is explored through speculative scenarios that ought to add to the debate about the boundaries between what is technologically possible and what is socially desirable.

Chapter 18 by the editors draws together the many threads of the four sections through an analysis of the strengths, weaknesses, opportunities, and threats facing GIS in public health.

References

Acheson, D. (chair of committee), 1988, Public health in England: The report of the Committee of Inquiry into the future development of the public health function, Cmnd 289, London: HMSO.

Acheson, D., 1998, *Independent Inquiry into Inequalities in Health,* London: HMSO.

Commission of the European Communities, 2003, A European environment and health strategy, COM(2003) 338, Brussels: European Commission.

Cromley, E.K. and McLafferty, S.L., 2002, *GIS and Public Health,* New York: Guilford Press.

de Lepper, M.J.C., Scholten, H.J., and Stern, R.M., Eds., 1995, *The Added Value of Geographical Information Systems in Public and Environmental Health*, Dordrecht, Netherlands: Kluwer Academic Publishers.

Goodchild, M.F., 1992, Geographical information science, *International Journal of Geographical Information Systems,* 6, 1, 31–45.

Longley, P. et al., 2001, *Geographic Information Systems and Science*, Chichester, U.K.: Wiley.

Melnick, A.L., 2002, *Introduction to Geographic Information Systems in Public Health,* Gaithersburg, MD: Aspen Publishers.

Scott-Samuel, A., 1998, Health impact assessment: Theory into practice, *Journal of Epidemiology and Community Health,* 52, 74–75.

Spadaro, R., 2003, European Union citizens and sources of information about health, Eurobarometer 58, March 2003. Survey of the European Opinion Research Group for DG SANCO, Brussels: EORG.

Stevens, A. and Raftery, J., 1997, *Health Care Needs Assessment,* 2nd series, Oxford: Radcliffe Medical Press.

van Tongelen, B., 2003, Personal communication from European Commission's DG Environment, 13 June 2003.

Section 1

Disease Mapping and Spatial Analysis

2 Basic Issues in Geographical Analysis

Ravi Maheswaran and Robert P. Haining

CONTENTS

2.1 INTRODUCTION

The purpose of this chapter is to consider basic issues in relation to the geographical analysis of public health data. The chapter is divided into two parts. The first part examines basic aspects of geographical data. It addresses spatially referenced data issues, including classification of data and assessment of data quality. The second part of the chapter is more technical in content. It describes the approaches to geographical analysis and focuses on cartographic operations, exploratory spatial data analysis, and ecological studies.

Cartographic operations are described briefly followed by a more detailed description of exploratory spatial data analysis. The section on ecological studies focuses on epidemiological and conceptual issues. Several of the later chapters in this book illustrate and expand on the data issues and some of the methods described in this chapter.

2.2 DATA ISSUES

The term data in the geographical context refers to both the attribute (e.g., a health outcome or a socioeconomic attribute) and the location identifier for the attribute (e.g., an address, a postcode or a census enumeration district). Since every

0-415-30655-8/04/$0.00+$1.50

measurement occurs at some point or during an interval of time, then any item of spatial data consists of the triple attributes: value (what), location (where) and time (when). The distribution of an attribute in geographical space is conceptualized as either a continuous surface or field or as a collection of discrete point, line and area objects in the study area. Air pollution across a region would be conceptualized as a field, the set of houses, roads or urban places as objects. An attribute like population is an exception to the suggestion of a unique conceptualization because it can be treated either as a field (e.g., a population density map) or as a collection of objects (e.g., a dot map of addresses).

For the practical purpose of storing either type of spatial data in a (finite) data matrix, which is a prerequisite for many areas of spatial data analysis, the analyst must choose a particular finite, discrete representation for the data. Spatial data are represented as points, lines, or areas. Points can be used to represent a field variable (e.g., a sample of points from the field) or point objects (e.g., an incinerator chimney). Lines can be used to represent boundaries on a surface (e.g., contours) or linear objects. Areas can be used to represent aggregations of a field variable, aggregations of point objects (e.g., census tracts) or areal objects. Often areas are artificial constructions that are unrelated to the underlying spatial variability in the attribute that is recorded (e.g., a ward). Areas can be in the form of a regular grid or in the form of irregular areas, such as enumeration districts or wards. In GIS, the terms raster and vector are used to distinguish these different forms of area representation.

Raster data are essentially gridded data (equivalent to a bitmap) with an attribute value attached to each cell of the grid. This may be the average level of pollution in an area. A polygrid is a similar concept where a range of attribute values is attached to each cell. Vector data are essentially data related to coordinates. These may be coordinates (x and y coordinates) of a point location or the longitude and latitude of that location. The point location may be that of a house or the postcode centroid. Vector data also commonly represent lines, where each line is represented by a series of point coordinates. Coordinates for the ends of the line can represent a straight line, while a curved line is approximated by a number of short straight segments and represented by a series of point locations. The vector database can also hold information on relationships (e.g., the left and right of a line). A series of lines joined together in a closed loop will form a polygon (e.g., the boundary of an electoral ward). These aspects of spatial data are described in detail elsewhere (Longley et al., 2001; Wise, 2002).

The term attribute data is sometimes used to refer to attribute information that may be attached to a spatial location or object. For example, at a point location such as a postcode centroid, we may have information on the number of households, people, age distribution, and other factors associated with that postcode. Similarly, there will be information on linear structures such as roads (e.g., type of road, traffic flow along the road) and polygons (e.g., census-based characteristics of an electoral ward).

A limited range of locational spatial data is often sufficient for routine public health work. The common ones are point locations (e.g., addresses, postcode centroids) and polygons (e.g., health authority boundaries, census ward boundaries).

Linear spatial information is often related to potential sources of exposure (e.g., roads, overhead power lines) or used to calculate road travel distances and times. Once georeferenced, the major interest for public health purposes relates to the attribute data. The classification of attribute data can be done in a number of ways. A formal way is to classify attribute data in terms of the level of measurement (nominal, ordinal, interval, or ratio), which is important since it specifies what mathematical operations are permissible on the data and hence what types of statistical tests can be performed. A substantive approach that is important with respect to public health applications is to classify it as data related to compositional attributes such as age and sex, exposures or risk factors, socioeconomic attributes (including confounders such as deprivation in a dose–response model) and data related to health outcomes. The exposure data may be classified conceptually as point, line, or area exposures. Examples of point and line sources of exposure include incinerators (points) and power lines (lines). Although area implies a continuous exposure, this may in practice be represented by a raster dataset with pollution values for each cell or as a vector dataset with polygons (e.g., water quality indicators for water supply zones). A range of census variables is commonly used to measure compositional and confounder attributes (e.g., percentage of owner-occupied houses, unemployment rate).

Health outcome data may be obtained from routine data sources or from specifically collected datasets. Routine data tends to refer to data routinely collected for mainly administrative and management purposes, but which may be used for public health analysis. The main examples are mortality data, hospital admission data, cancer registration data, and congenital malformation data.

Often, analysis in public health is based on areas rather than point locations because data confidentiality means point location (individual level) analysis is not possible. Analysis may be based on geographical units of varying sizes. A key requirement for this type of analysis is an appropriate denominator population for the calculation of rates. This commonly comes from census-derived information. Censuses are carried out periodically, not infrequently at ten-year intervals; they provide detailed information at the small-area level regarding the age-sex composition of the resident population. However, the denominator becomes progressively less reliable the further it is from the last census. Often, midyear estimates take into account births, deaths, and migration. These estimates are produced for larger administrative areas, such as local government or health authority areas. National or government bodies may carry out such estimates but in some areas, local authorities also produce their own estimates using additional local information (e.g., the development of new housing estates).

In the United Kingdom, an alternative source of denominator population information increasingly used is patient registration data from general practices. This information may be aggregated to the postcode level and subsequently to higher levels of aggregation. This information potentially provides an up-to-date denominator database for calculation of rates. In some other countries, administrative records are maintained at very detailed levels such as by household, so the denominator can be aggregated and used for calculation in very flexible ways (e.g., street-level rates).

Factors that influence health may be classified in a number of ways. A simple and useful classification uses four headings: genetic and inherited, environment, lifestyle and healthcare. The environment may be further classified into physical (air, soil, water, and radiation), socioeconomic and psychosocial, with the built environment as an additional category incorporating characteristics such as type of housing and degree of open space. The ways in which these factors are used in public health analyses are varied. There may be a simple spatial description of the distribution of these factors. On the other hand, the association between these factors and health outcomes may be examined. In some cases, one of these factors may be the main exposure of interest (e.g., effect of air quality on respiratory disease) and other factors (e.g., smoking) may be treated as confounders or covariables in the analysis. Another approach would be to see if the effect of the main factor of interest on the outcome varies with the level of another factor (e.g., socioeconomic deprivation). In this latter analysis, the technical term is effect modification or interaction. There are a number of issues related to both the conceptual and statistical aspects of interaction and these are addressed in detail elsewhere (Rothman and Greenland, 1998).

A common problem regarding factors that influence health is the availability of data for geographical analysis. Data on the physical environment are becoming increasingly available often through remote sensing technology, although there are still considerable barriers to obtaining exposure data at a fine spatial resolution. Socioeconomic data may be available from censuses and other sources (e.g., claims for unemployment benefit, free school meals). Often the data may be available in some parts of the system (e.g., within social services departments) but there are strong barriers to data sharing due to real or perceived confidentiality and data protection issues. Another issue is the level at which these data are georeferenced and the accuracy of this georeferencing. Data on the psychosocial environment are not normally available through any routine sources and are usually obtained through special surveys. Similarly, data on lifestyle factors (e.g., smoking, alcohol consumption, physical exercise) are usually obtained through surveys. Such surveys are normally conducted for other purposes but could be used for geographical analysis providing there is appropriate spatial stratification. An important limitation often is the sparsity of data at a small area level. An alternative approach to analysis would be to examine these data at the individual level using cohort or case-control methodologies rather than ecological analysis.

Data on healthcare factors may be available through routine health information systems that collect data for administrative purposes. Such data may include numbers of beds available, numbers of consultants and other specialist staff, bed occupancy rates or provision of specialist services, for example. Information that is more detailed will usually require additional efforts to gather such as policies for treatment and referral or care pathways. These data will then usually need to be converted to rates (e.g., specialists per 1000 population) for comparison across areas and for use in ecological analyses. To do this, the population being served by the staff or facilities will need to be determined. If service provision areas have been clearly agreed upon, then calculation of rates is straightforward, based on the resident population in these areas. Otherwise, catchment populations will need to be estimated and this is not always easy. In some cases, the local geography will mean that virtually all the local population will attend the main hospital for a town. However, in larger urban conurbations, with a number of provider hospitals, determining the catchment population could be more difficult. A common approach is to allocate a geographical unit to the healthcare facility that the majority of referrals or admissions from that area attend.

Spatial data quality may be considered in terms of accuracy, resolution, completeness, and consistency. These four dimensions can be related to the three components of a single item of spatial data: attribute, location, and time (Haining, 2003; Veregin and Hargitai, 1995). For example, if a study is being carried out to examine the association between overhead power lines and cancer, high levels of spatial resolution of locations is needed as the electromagnetic fields decrease very rapidly with increasing distance from power lines. This high level of resolution is important for both the point location of cases and controls (or the denominator population) and for the power lines. If, however, postcode centroids are used to link cases and controls to exposure to different levels of drinking water constituents and the latter data are for water supply zones with several thousand people per zone, then lower levels of location resolution for postcode centroid georeferencing will be acceptable.

The quality of health data has received much attention in epidemiological and public health circles. With regard to routinely collected health data, such as mortality and hospital admission data, coding of cause of death or admission may or may not be accurate. Mortality coding may become progressively less specific with increasing age. A further issue is the contribution of multiple pathologies to the cause of death in very elderly people. There are agreed systems of coding such as the International Classification of Disease but there may still be variations in interpretation of the guidelines for coding and classification.

Several factors may affect the interpretation of hospital admissions statistics (Anderson, 1978). The etiology of the disease determines the underlying levels of morbidity; however, underlying morbidity is not all observed or known about. This will depend on a range of factors including medical practice, illness behavior and the organization of medical care. These factors also influence admission to hospital and admission criteria. Diagnostic coding and even fashion influence the recording of hospital statistics. A further issue is identifying readmissions, which require a unique patient identifier. Errors and biases at several stages in the process generating hospital admissions data would lead to variations in data quality that need to be taken into account when interpreting the data.

Although death registration is mandatory and recording of hospital admissions is firmly governed by administrative systems, cancer registration often relies on a voluntary system. Thus, in addition to the issues related to diagnosis and coding, there may be major discrepancies in completeness of data capture across geographical areas. It should also be noted that even when recording is mandatory there might still be significant underreporting and a good example is in relation to notifiable diseases.

The quality of denominator data also needs careful consideration. Censuses may not reliably enumerate the entire population for various reasons, including misconceptions about the use of census data, concerns about privacy and negative publicity.

In the U.K. 1991 census, there was substantial underenumeration partly related to the perception that census data would not be confidential and could be used for surveillance, including identifying individuals for taxation purposes. Some sections of the population might be more likely to be underenumerated, such as young men, refugees, and ethnic minority groups. Alternative sources for the denominator population, such as general practice registered populations, are also potentially problematic. List inflation is a recognized problem, especially in areas where there is a highly mobile population and people who have left the area may not be removed from general practice lists. Underregistration may also be a problem, especially of young adults, refugees, and homeless people. The place of registration of students is also potentially problematic as they can be registered either with their general practice near their home address or with student medical services at their higher education institution. This is also a problem with censuses and needs careful consideration in geographical studies with regard to the denominator population at risk (e.g., geographical studies of the incidence of sexually transmitted disease, leukemias, and lymphomas).

2.3 GEOGRAPHICAL ANALYSIS

There are a number of ways of classifying geographical analysis. One approach is to distinguish between cartographic operations, spatial data analysis, and mathematical modeling. Cartographic operations include buffering, overlay, geometric (as opposed to statistical) forms of spatial interpolation and regionalization. These are described in the following section on cartographic operations.

Spatial data analysis is usually divided into two sections: exploratory spatial data analysis and confirmatory spatial data analysis. Exploratory spatial data analysis may be purely descriptive or may involve an element of hypothesis testing where no model is proposed for the alternative hypothesis, such as testing for significant clusters of cases of a disease (as described in Chapter 4). Exploratory spatial data analysis has numerical and visualization aspects. It is described in detail in the section on exploratory spatial data analysis.

Confirmatory spatial data analysis is concerned with inference and hypothesis testing. A model is proposed to represent variation in the data, which is then used for the purposes of hypothesis testing. A typical example would be the investigation of an association between an exposure and a health outcome, with the term exposure being used in a broad sense to include any factor that may affect a health outcome. The term health outcome may also be viewed more generally and could be substituted with proxies for health outcomes or intermediate process measures. The main type of epidemiological study used to test hypotheses in a geographical setting is the ecological study. It is described in detail in the section on ecological studies.

Mathematical modeling includes optimization and models for describing the spread of communicable disease. Optimization involves the process of obtaining the best solution to a problem. One use is identifying locations that are most suitable based on defined criteria. Examples include siting of primary care facilities to maximize geographical access and population coverage and identifying the shortest or quickest routes for travel for emergency ambulances or patient transport vehicles from a variety of locations to healthcare facilities. This chapter does not cover mathematical modeling of spatial data in detail.

2.3.1 Cartographic Operations

Longley et al. (2001) provide a clear and detailed description of cartographic operations and argue that these operations are in many ways the crux of GIS. Cartographic operations often involve map-based or spatial querying of a database to identify properties. They are an important component of what it means to add value to geographic data and we summarize the key features in this section.

One group of cartographic operations has to do with measurement and there are four elements:

1. Distance and length
2. Area
3. Shape
4. Slope and aspect

Measuring distance relies on a metric or rule for determining distance. The simplest and easiest to implement is calculating straight-line distance. When calculating greater distances, the earth's curvature needs to be taken into account. More complicated distance measurements include distance along a road network, often referred to as a stick network because segments of roads are represented as straight lines.

The area of a polygon is measured using the trapezium rule implemented within a GIS. The area measurement in public health applications is often used as part of a calculation, for example, as a denominator for calculating population density. Sometimes it is used to describe the geographical size of a patch served by a healthcare unit such as a psychiatric community outreach team.

The shape of an area is often measured in relation to the compactness of the geographical area. The measurement of shape may be used when creating geographical areas for administrative purposes or for targeting interventions and may be part of the optimization process in constructing regions.

Slope and aspect have clear applications in relation to digital elevation (terrain) models. In the public health field, these measures may be useful in modeling the ecology and climate in relation to communicable disease, such as climate change and changes in the local environment that may enhance the spread of vectors for disease.

Transformations are arguably the key element within a GIS for public health use. Longley et al. (2001) classify transformations under five headings:

1. Buffering
2. Point-in-polygon
3. Polygon overlay
4. Spatial interpolation
5. Density estimation

Buffering is the process by which an area is created around an object such as a 200-meter circle around a point source of pollution. Buffers may be created around line and area sources of pollution. Other examples of use include defining potential catchment areas around healthcare centers.

Point-in-polygon methods are essentially a means of linking points to areas. For instance, all postcodes falling within a geographically defined urban regeneration area will be linked to that area using this GIS process.

Polygon overlay is a method of linking information from two or more sources with boundaries that do not correspond. Reassigning values from one dataset to another may then be carried out (e.g., using an allocation based on proportional area).

Cartographically based methods of spatial interpolation include cell declustering, triangulation, and inverse distance weighting, which are described in detail elsewhere (Isaaks and Srivastava, 1989).

The fifth heading under transformations is density estimation, e.g., estimating population density. There is an important distinction between the last two categories: density estimation creates a field from discrete objects, while for spatial interpolation a field already exists (e.g., outdoor air pollution levels) but we only have samples from this field to work with.

Other elements sometimes included within the list of cartographic operations include the identification of centers and measures of dispersion, spatial dependence and fragmentation and fractional dimension.

Measures of centers include centroids and the point of minimum aggregate travel, both of which are described in detail by Longley et al. (2001). The centroid is the weighted average of x and y coordinates of points (e.g., individuals' addresses) within a given area and is thus a measure of the center of the distribution of these points. It is the point that minimizes the sum of squared distances. Centroids are frequently used in public health GIS work. There are two types: geometric centroids and population weighted centroids. The latter tends to be used more as it reflects the location of the bulk of the population within a geographical area. Population weighted centroids are typically used to attribute an exposure to a population in small geographical areas (e.g., census enumeration districts) or to calculate distances to healthcare locations. Population weighted centroids tend to be used because point locations (of addresses or postcode centroids) are not available for cases and the population at risk within small areas.

Measures of dispersion, spatial dependence, and fragmentation and fractional dimension are described in more detail elsewhere (Longley et al., 2001). They tend not to be used largely within public health GIS analysis although spatial dependence measures, for example, are potentially of use in exploring questions about the infectiousness of some diseases. These are as-yet-unexplored areas of GIS capability in relation to public health questions. They may also overlap with other categories of geographical analysis.

Region building, or zone designing, is the process of partitioning an area into (quasi) homogenous subareas or regions using small areal units as the basic building blocks (e.g., deprivation regions using enumeration districts as the building blocks). One of the purposes of region building is obtaining spatial units in which calculated rates of disease, for example, are of equivalent statistical precision. These regions