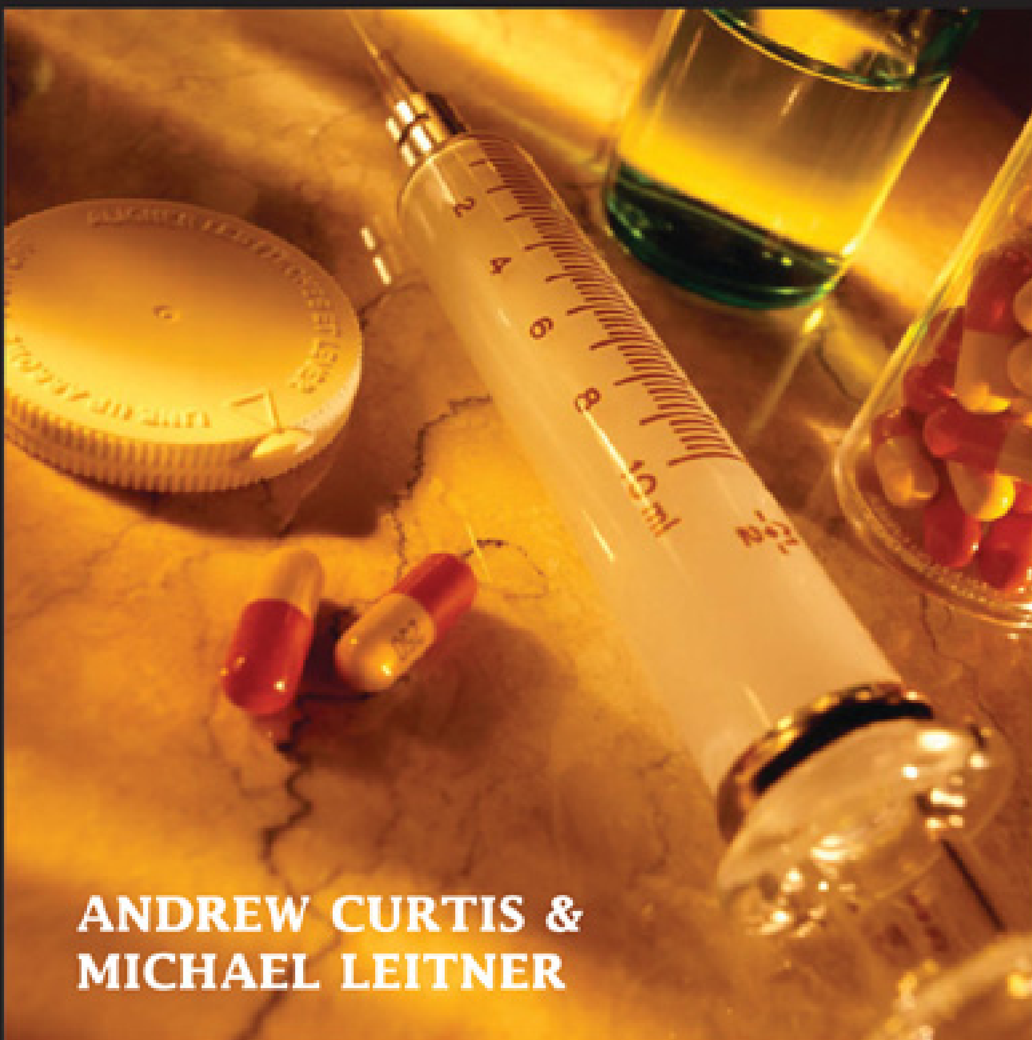
Geographic Information Systems and Public Health

Eliminating Perinatal Disparity



ANDREW CURTIS &
MICHAEL LEITNER

Geographic Information Systems and Public Health:
Eliminating Perinatal Disparity

Andrew Curtis
Louisiana State University, USA

Michael Leitner
Louisiana State University, USA

IRM Press
Publisher of innovative scholarly and professional information technology titles in the cyberage
Hershey • London • Melbourne • Singapore

Acquisitions Editor: Michelle Potter
Development Editor: Kristin Roth
Senior Managing Editor: Amanda Appicello
Managing Editor: Jennifer Neidig
Copy Editor: Killian Piraro
Typesetter: Amanda Kirlin
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
IRM Press (an imprint of Idea Group Inc.)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033-1240
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@idea-group.com
Web site: http://www.irm-press.com

and in the United Kingdom by
IRM Press (an imprint of Idea Group Inc.)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanonline.com

Library of Congress Cataloging-in-Publication Data

Geographic information systems and public health : eliminating perinatal disparity / Andrew Curtis and Michael Leitner, editors.
p. ; cm.
Includes bibliographical references and index.
Summary: "This book provides an overview of why geography is important in the investigation of health, the importance of the main components of a GIS, how important neighborhood context is when using a GIS, and the general differences found between urban and rural health environments"--Provided by publisher.
ISBN 1-59140-756-7 (hardcover) -- ISBN 1-59140-609-9 (softcover)
1. Geographic information systems. 2. Public health--Data processing.
[DNLM: 1. Geographic Information Systems. 2. Infant Mortality. 3. Community Health Planning. 4. Pregnancy, High-Risk. WS 16 G345 2006] I. Curtis, Andrew, 1967- II. Leitner, Michael, 1965-
RA566.G462 2006
362.1'0285--dc22
2005023873
eISBN 1-59140-610-2

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

This book is dedicated to everyone who has been involved with the Baton Rouge Healthy Start program.

Geographic Information Systems and Public Health:

Eliminating Perinatal Disparity

Table of Contents

Preface .... **ix**

Acknowledgments .... **xxi**

Chapter I
Explaining the Geography of Infant Health **1**
Geographic Variations in Infant Health 2
Smoking is Bad 9
What Does It Mean to Be Poor? 10
Stress 11
The Geography of Health 12
References 18

Chapter II
An Introduction to GIS (All Things Data) **21**
Data Input 24
Health Data 24
Confidentiality Issues 27
Address Matching/Geocoding 28
Other Useful Data 1: Socioeconomic Data 31
Other Useful Data 2: Boundary and Background Data 32
Data Manipulation 37
Aggregating into Spatial Units 37
Data Reduction 41

Creating New Data 42
Calculating Deprivation Indexes 42
Improving Health Outcome Information 44
Perinatal Periods of Risk (PPOR) 45
References 47

Chapter III
An Introduction to GIS (All Things Spatial) 52
Visualizing the Data 52
Choropleth Map 57
Common Dot Map 60
Isarithmic (Isoline) Map 61
Proportional (Graduated) Point Symbol Map 61
Spatial Analysis 62
CrimeStat® 64
GeoDa™ 65
Geographically Weighted Regression (GWR) 65
SaTScan™ 66
GIS as a Management Information System 68
What is a Neighborhood? 69
Including Geography in the Analysis 70
Holistic Neighborhood Investigations 72
Spatially Synthesizing Previous Research 73
References 73

Chapter IV
The Geography of Health Risks 79
Infant Deaths, Low Birth Weight, and Short Gestation Deliveries 83
Medical Risks 85
Behavioral Risks 87
So What Can We Do With GIS? 91
Cohort or Social Risks 95
Social Risks: Disparities in African American Neighborhoods 96
Spatial Cohort 98
Neighborhood Risks 99
Suffer the Children 100
Environmental Risks 103
GIS Analyses of Environmental Risks 107

GIS, Cancer, and Low Birth Weight Research in Louisiana 110
Cancer and Birth Outcome Co-Investigation Template ... 111
Summarizing It All: The Relationship Between Risk and Stress 113
So What Can Be Done? 114
References 119

Chapter V
GIS and Spatial Analysis: Keeping It Simple **146**
Exploratory Analysis vs. Hypothesis Testing 146
Spatial Design 148
Spatial Sampling 149
Aggregation Effects 153
Three Simple Techniques: Overlay, Density, and a Difference of Proportions Test 154
Overlay as Analysis 154
A Cautionary Tale 157
Density Analysis 157
Difference of Proportions Test 160
Results for Year One (Table 1) 165
Results for Year Two (Table 2) 165
Results for Year Three (Table 3) 167
Under-18 Pregnancies (Table 4) 169
References 172

Chapter VI
Advanced Spatial Analysis **174**
Spatial Autocorrelation 174
Global Spatial Autocorrelation 175
Local Spatial Autocorrelation 178
Cluster Analysis 179
Cluster Techniques 182
Spatial Filtering (DMAP) 182
Nearest Neighbor Hierarchical Clustering (NNHC) 183
Kernel Density Estimation 184
Infant Mortality and Prenatal Risks: The Case of East Baton Rouge 188
Regressing Selected Prenatal Risk Factors on the Infant Mortality Rate 192
Geographically Weighted Regression 194
References 199

Chapter VII
Spatial/Temporal Stability in Neighborhoods of Risk: The Mobility of Mothers 203
How Far Do the Mothers Move? 204
Temporal Stability and Implications for Outreach 208
Developing a Neighborhood Categorization Scheme Based on Temporal Stability 208
Constructing Neighborhoods Around Mortality Locations 210
Temporal Stability in Risks Around Infant Deaths 211
Temporal Stability in a Global Risk Investigation 216
Temporal Stability in the Four Neighborhoods 218
Results from the Difference of Proportions t-test 219
Conclusions on Temporal Stability 221
References 222

Chapter VIII
Patient Confidentiality 224
Confidentiality in Maps 226
Statistical (Attribute) Confidentiality 226
Spatial (Locational) Confidentiality 227
Preserving Confidentiality in Governmental Agencies 227
U.S. Department of Health and Human Services 228
U.S. Census 229
U.S. Department of Justice 229
Geographically Masking the Location of Confidential Point Data 230
Experimental Testing 230
Results for Global Geographic Masking 231
Results for Local Geographic Masking 232
Preserving Spatial Confidentiality of Two Locally Masked Point Patterns 237
Manipulating Both Area Boundaries and the Location of Confidential Point Data 240
References 243

Chapter IX
Creating the Baton Rouge Healthy Start GIS 245
Beginnings 246
Determining the Program Area 258

Identifying Areas With No Prenatal Care 259
Neighborhood Profiling 262
Creating the Database 262
Data Input 264
Reaching Out 265
What Next? 266
Post Script 266
References 266

Chapter X
Bioterrorism, Pregnancy, and Old White Men 268
Vulnerability in the U.S. 268
Bioterrorism and Pregnancy Risk 269
GIS and Vulnerability Mapping 271
Identifying the Vulnerable 272
So How Do We Bring Healthy Start into This? 274
Are Pregnant Women Really Vulnerable? 275
Criticisms of Syndromic Surveillance 279
References 282

Chapter XI
Rural Health Issues and Their Investigation in a GIS Environment 287
Introduction 287
The Complexity of Rurality 288
Rural Places and Health 289
An Overview of Some Rural Health Issues 290
Rural Geography and Dealing With Rural Data 295
Conclusion 300
References 301

About the Authors 305

Index 307

Preface

Why Write this Book?

This book has been written for anyone involved in public or community health with an interest in learning about a geographic information system (GIS). Although the focus is on infant health, many of the topics covered are transferable to other community health issues. It is also written for the upper division undergraduate or graduate student interested in GIS. Having worked with both of these audiences for several years, I understood the need for a breezy, readable text that complements what is already known and what is being taught. I wanted a book that was readable but at the same time included a lot of detail. The hope is that the overall message comes through, but that there is a point in keeping the book around as a reference when it is needed.

This book is written to fill a gap, as many other community health groups, especially local Healthy Start organizations at regional and national meetings, have asked me questions about how to incorporate a GIS into their program. So far, little has been written on this subject. Equally so, my GIS students have a plethora of "texts" that adequately describe the right buttons to push and how metadata should be written, but again, relatively little exists in terms of a detailed, contextual work that helps one understand why one would want to use a GIS. I hope these two audiences will start to meet up at some point, maybe even as a result of this book. A few of my GIS students have already been "turned onto" public health and medical geography. The demand for their services will certainly grow over the next few years.

The language of the book is meant to be more readable than that of a typical GIS text; indeed, one reviewer commented on the "subjective…informal approach to GIS…" and this is exactly the spirit I have hoped to capture: to demystify a subject for those who are not entirely comfortable in the world of computing.

But first, let me introduce you to our group at Louisiana State University...

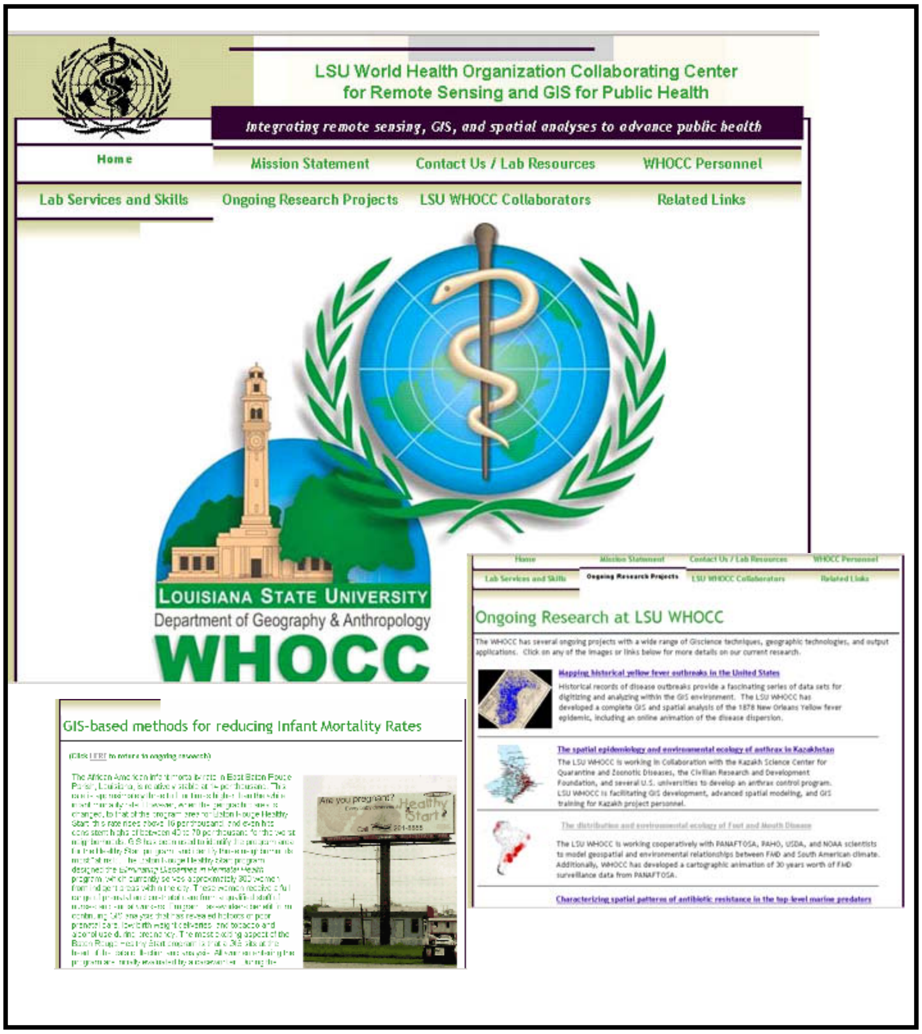
The World Health Organization's Collaborating Center (WHOCC) for Remote Sensing and GIS for Public Health is housed in the Department of Geography and Anthropology at LSU. This center is involved in several health initiatives, ranging from understanding the ecological and anthropological dimensions of Chagas disease in Mexico, antibiotic resistance found in sharks in the Gulf of Mexico, and anthrax in Kazakhstan, to the spatial modeling of diseases in history. The common factor in all these projects is the use of geographic information science (GISc) in solving health-related issues. The use of spatial technologies is also reflected in the mission statement of the center, and can be found on its Web site (www.whocc.lsu.edu):

- *To apply geographic technologies and techniques to public health problems, whether they be local, regional, national, or international in nature.*
- *To provide geographic information science (GISc) training and guidance to those in health related fields.*
- *To promote the use of GISc in public health.*
- *To advance GISc in association with public health through research initiatives with scientists and public health workers at other institutions.*
- *To promote and advance quantitative, geospatial modeling, and mapping techniques with special applications to public health questions.*

The Web site includes synopses about our research projects, as well as a few more "fun" items, like a cartographic animation of the 1878 yellow fever epidemic. One of those projects is the ongoing relationship between the WHOCC and the Healthy Start program of Baton Rouge (see Figure 1). This book has been 4 years in the writing, and it attempts to capture the essence of this relationship, from the initial understanding of the problem to the development of a Healthy Start initiative to eliminate racial disparities in infant mortality. GISc has been used on all stages of this journey. This book attempts to provide a working guide, first introduction, and insight into research possibilities for others about to embark on a similar journey. Chapter IX details this journey, drawing on the approaches described in other chapters to show how GIS was used in the initial grant writing, and how it is implemented today.

This book meets every one of the WHOCC mission statements, one of which is training and guidance, though this book should not be viewed as a training manual. Although techniques are discussed in depth, these are framed within the context of infant health. The complexity of working on health issues in some of our poorest neighborhoods is not skirted. It is not enough to know how to press the

Figure 1. The WHOCC Web site



right buttons, it is more important to know how to ask the right questions. The user should make the GIS work for him or her and the local community, and not the other way round.

As previously mentioned, this book has been written for the non-skilled GIS user. The first few chapters cover some of the GIS basics, including data sources and types, spatial analysis, and how to make maps. These chapters are not written to guide the user through a step-by-step learning process; rather, they are meant to provide an overview, a flavor, to those taking their first steps down the GIS path. It is more important to understand why and how we use GIS rather than knowing one particular software package. Once these initial building blocks have been gained, subsequent chapters expand and illustrate how more involved questions relevant to a community health unit can be addressed. Although these chapters may appear too advanced for a first-time user, it is hoped that once interest is generated, the reader will want to progress on to the next level of implementation and inquiry. Several detailed research examples are provided, and it may be advisable for the first-time reader to skim through these until he or she feels more comfortable with GIS and geographical analysis. These more involved topics range from a selection of advanced spatial analysis methods to the preservation of confidential data within the GIS environment.

This book is also a suitable companion piece for the first-time GIS student taking a university course, whether specializing in the public health field or not. Many of the chapters follow the traditional course outline used in the introductory GIS classes taught at Louisiana State University. Scattered throughout the book are also examples of our current research initiatives, again written in an informal way, but giving the topic substance beyond just another GIS manual.

Overview of the Chapters

Many of the topics initially presented in the first three chapters are revisited in subsequent sections of the book. Indeed, spatial analysis is revisited twice more, such is its importance. In this way the book is somewhat hierarchical, allowing the reader to choose the depth of material required.

Chapter I

The goal of this chapter is to present an initial understanding of the role space plays in health analysis. Ask yourself, what disease or health outcome does not create a spatial pattern? The answer: none.

This chapter provides an introduction to many of the themes that run through the book, and should be seen as a springboard, a means to jump off into the subsequent sections depending on where the interest lies. The topic of negative infant health is first discussed here, along with an initial appreciation of the geography of health. Summary statistics for infant health outcomes are provided at a variety of geographic scales, descending from the national level to eventually end up at the zip codes served by the Baton Rouge Healthy Start. This geographic area will act as the canvas on which most of the book's examples are drawn. In this way it is hoped the reader will gain a degree of familiarity with the situation in Baton Rouge. This should also help clarify why the GIS approaches and solutions presented were applied.

This chapter also teases the reader with the first sense of how important geography can be in terms of understanding health surfaces. This is achieved by giving an initial insight into the complexities of working with neighborhood-level health data. Three main "risks" usually associated with pregnancy outcomes — smoking, poverty, and stress — are presented to illustrate this point. Understanding risks such as these are presented as simple "overlays" onto a city map in order to identify neighborhood vulnerability. This common GIS technique allows us to combine multiple datasets, often drawn from multiple disciplines, into one holistic impression. The topic of neighborhood vulnerability is again broached in Chapter IV, where the full weight of its complexity is described.

The chapter ends with what is hoped to be an interesting example of how space/geography and epidemiology/public health have long been bedfellows. An example from the 1800s is used to give the reader an insight into how the collection, display, and exploratory analysis of data can be used to understand a public health problem and begin the search for a mitigation solution. Apart from being an interesting case study, it is hoped these descriptions from the 1800s will excite the reader, in the same way they do the authors, about what can be achieved by approaching a health problem spatially.

Chapter II

The goal of this chapter is to provide the non-GIS reader (a community health investigator or even a student taking a GIS class for the first time) with an overview of what a GIS is, what data sources are available, and how these data can be manipulated to answer questions.

This chapter is not meant for those skilled in GIS use. An overview is presented about the capabilities of a GIS, how different spatial layers can be manipulated to answer different questions, and how the power of a simple query can reduce data so that specific neighborhood level investigations can be made. The reader is then given a tour of typical data types and how to bring them into the GIS, with particular emphasis placed on geocoding addresses, as this is probably the most useful data input procedure for community health units. Background "map" data such as 1:24,000 Scanned Topographic Maps (DRGs) and Digital Orthophoto Quarter Quadrangles (DOQQs) are also discussed.

As most forms of neighborhood analysis will involve socioeconomic information, the reader is briefly introduced to census data, in all their aggregations (and where to find them). Woven within these data concepts are other related issues such as the power and pitfalls of spatial analysis and visualization. *Power* includes the ability for local communities to control their own data, analysis, and understanding of local problems, while *pitfalls* range from the variations that occur with different spatial aggregations to the issue of confidentiality and data access. Finally, four important pregnancy-related data manipulations are outlined, three of which can be created within a GIS: the Kessner and Kotelchuck Indices, which judge acceptable levels of prenatal care, and Perinatal Periods of Risk (PPOR), which link key components of the birth outcome to a fetal or infant demise. The fourth type of data are typically aspatial: Pregnancy Risk Assessment Monitoring System (PRAMS). These data can still be used as a template for other neighborhood level survey collections. These four data sources are probably unfamiliar to most GIS users, while most infant health researchers might not have thought about them in a GIS context.

Chapter III

The goal of this chapter is to introduce the reader to arguably the two most powerful uses of a GIS, the ability to extract understanding from spatial data in the form of maps and spatial analysis.

One of the first uses of a GIS for the newly initiated is the production of spatial displays or "maps." This chapter briefly covers the errors that can be faced in this process, from overreliance on spatial identifiers (zip codes are often filled in incorrectly), to how to avoid basic cartographic mistakes. Choropleth (graduated color) maps, dot maps, isoline surfaces (contours), and proportional symbol maps are all discussed. Color schemes, appropriate data manipulation, and the difference between quantitative and qualitative displays are also presented. Although this chapter is not a substitute for a good computer cartography class, or even a dedicated book on the topic, it should provide a reasonable starting point to making legible and acceptable maps.

The reader is then introduced to the power of spatial analysis. Although there has been an upswing in GIS use, especially in government agencies, the ability to ask the right spatial questions and know how to look for the answers is still sadly lacking. Yet having the ability to ask and answer spatial questions is probably the most powerful role a GIS can play for a community health organization. Questions could include where are our risks, what are our risks, how do our risks change over time and space? And one of the most important questions of course is, do our neighborhoods meet the thresholds set for federal funding?

Several available online spatial analysis packages are briefly reviewed. Important geographic concepts are also discussed that are relevant to most spatial analyses, such as how does the spatial aggregation of data affect results, are data spatial autocorrelated (events are not independent of each other), and what is distance decay? Again these concepts are kept simple, the goal being to inform rather than overwhelm.

The chapter concludes with a discussion as to how a neighborhood should be defined. This is important in terms of creating the geographies analyzed in the GIS. It is doubly important for a new form of analysis that combines both individual health outcome data (a low-birth-weight delivery), and neighborhood context (the education level or income base of the mother's neighborhood).

Chapter IV

The goal of this chapter is to provide an insight to a GIS user, whether newly introduced or skilled, to the complexities involved in working with community health at the neighborhood level.

A common criticism of GIS and spatial analysis is that people are "reduced" to numbers, that the analyst does not understand the context of the situation, and by extension the results have little validity in solving real-world problems. This chapter hopes to address this criticism by presenting the complexity faced in working with neighborhood level GIS analyses. It is heavy with citations, offering the reader many examples of further reading assignments, or even template research ideas. The focus of the chapter is to identify some of the risks that can lead to a negative birth outcome. In so doing, risks that may not directly, traditionally, affect a delivery are also covered. This is for two reasons: First, because they are still stress inducing which is related to pregnancy outcome, and second, because a broad spectrum of academics believe that the social context determines many health outcomes. This social context does not begin on the first day of a pregnancy.

Following the disease ecology approach often championed in medical geography, risks are broken into biological, behavioral, social, cultural, economic, and environmental categories. Many of these risks combine to create stress-inducing neighborhoods. Simple "choices" such as smoking during pregnancy, or attending a prenatal visit, soon become more complex, and are as much to do with the living environment as any individual behavior.

Neighborhood risks, such as crime, poor building materials, presence of lead, proximity to busy roads are all presented as being inputs into a complex milieu. It is important to appreciate this complexity. GIS is a tool that can begin to build a holistic vulnerability surface by layering one risk on top of another. However, the complexity may be missed if only GIS analysis is used. Qualitative approaches, such as focus groups, are needed to help explain why hot spots occur where they do. Paraphrasing Stan Openshaw, a famous geographer and one of the original GIS modelers, a cluster is only a data anomaly until it is given context.

Chapter V

The goal of this chapter is to introduce the novice GIS user to a series of spatial analysis approaches that can be applied in a community health environment.

As has been previously mentioned, the ability to ask the right questions and to search for answers is vital for a community health organization. This chapter assumes little or no previous background in spatial analysis and leads the reader through the concepts of designing a spatial question, how and where to collect sample data, and the pitfalls of working with spatial data. One such data issue comes with having an adequate sample, and a background population against which to compare or test the sample, what is known in the game as numerators and denominators. For example, if one house has a birth, and that baby dies, the house and possibly the road will have an infant mortality rate of 1,000/1,000. Similarly, if a car crashes with twins on board, a hot spot of mortality will likely show in the neighborhood. However, if we demand a minimum number of births for any area to counteract this "law of small numbers," what happens when two babies die in the same apartment building to two different mothers when there were only ten births in total? Is something going on there?

The reader is also presented with a simple analysis approach, a version of the difference of proportions t-test that can easily be created in a spreadsheet such as Excel. Simply put, the GIS can reduce a dataset by both queries that are spatial (for one zip code or half a mile around a clinic), and aspatial (by all teen births). These data are exported to Excel, where a comparison is made to the larger population, such as the whole city. The question in this case is, is there a significant difference to teen mothers in this area? Certainly, more powerful techniques are available, and some of these will be presented in Chapter VI, but there is an argument for keeping it simple. The novice GIS user, along with any introductory GIS student, as long as he or she has had at least one basic statistics class, can perform the analysis example provided in this chapter.

Chapter VI

The goal of this chapter is to expose the reader to a series of techniques on the "cutting edge" of spatial analysis. These techniques can be applied to find spatial intensities or "hot spots" as well as causative associations.

The inexperienced GIS user might want to skip this chapter, at least during an initial learning phase. For those who are more used to GIS analysis, or with a reasonable background in statistics, this chapter is designed to discuss approaches to find clusters (including nearest neighbor hierarchical clustering and the spatial filter), fill-in space (kernel density analysis), and spatial associations (spatial regression). Again detailed examples are provided of these approaches.

Several excellent books have been written on GIS and spatial analysis, and this chapter only scratches the surface. Nevertheless, Chapters V and VI should provide a reasonable armory for a GIS user irrespective of where he or she falls on the expertise continuum.

Chapter VII

The goal of this chapter is to present an example of how a data condition can affect GIS analysis. This condition is how mobile are pregnant women? It is hoped the reader will begin to understand how important data are — and not fall for garbage-in, garbage-out.

Spatial variation will always occur between analyses. It is very unlikely that an analysis will reveal a hot spot that can be used to precisely identify subsequent hot spots — unless there is an environmental association, such as proximity to a landfill. In order to counter this problem, the areas of analysis have to be large enough to reduce the effect of this "spatial error." This leads to another problem, in that too large an area does not really help outreach.

This is the problem: What is the point of performing a GIS analysis unless we can extract some good out of it, such as targeting resources? In order to effectively target resources, relatively small areas are needed to be identified — and yet what is the point in sending outreach teams into a neighborhood if the problem by that time has cycled out? This chapter presents different research approaches that have been used to identify how spatially stable results are.

A further issue is raised — although spatial variation will always exist, what happens if the women themselves are mobile? If we create a hot spot map of infant death — what is the important residence? That on the birth certificate, or on the death certificate, or at a certain stage of the pregnancy? Unfortunately, the chapter offers no gold standard in terms of a solution, because it is not really discussed in the geographic analysis literature.

The purpose of this chapter is therefore twofold: to initiate this discussion, and to make the reader aware that most data have limitations.

Chapter VIII

The goal of this chapter is to provide some guidelines as to how to map health while preserving patient confidentiality.

In this chapter, it is argued that national standards (e.g., HIPAA) are good at preserving somebody's statistical information, usually recorded as text or in a spreadsheet format, but lack appropriate rules for visualizing this information on maps. Privacy rules for visualizing confidential data on maps (i.e., spatial confidentiality) become more and more important, as governmental agencies increasingly use GIS as a standard tool for collecting, storing, analyzing, and disseminating spatial information. One possibility to preserve somebody's spatial confidentiality is by manipulating the location of health records just enough to protect the confidentiality of individuals, but not too much to change essential visual characteristics of the true, original spatial distribution of those records. This method is referred to as "geographic masking" and this chapter evaluates the usefulness of different geographic masking methods in preserving spatial confidentiality.

Chapter IX

The goal of this chapter is to show the reader how a GIS might be woven into a community health initiative, from inception (grant getting), through outreach, basic research and the day-to-day running of the program.

This chapter may be the most important for many readers, as it can be seen as a quick reference as how to involve GIS within a grant application. The chapter then goes on to show how GIS has been implemented in all aspects of the program over the following four years, from outreach to a data collection tool. Although many of the actual approaches described in this chapter have previously been mentioned, it is here where they are set in a context.

If one message comes from this chapter, is should be that a stronger working relationship should exist between community health organizations and academia. There are many able geography graduate students who would love to bring their skills to a topic as important as improving infant health. These collaborations could easily end up being theses and dissertations, resulting in even more applicable research for the health unit. So go seek out your local geography department…

Chapter X

The goal of this chapter is to open everyone's eyes to the way pregnant women are missed in disaster response and recovery planning.

Although this chapter might seem somewhat of a departure from the rest of the book, it raises an important point. Pregnant populations, especially those in neighborhoods with high pregnancy risks, are often the most vulnerable whenever disaster strikes. These women often have little family support, live in densely packed inner-city areas, and have no transport. In addition, these women are pregnant. Although vulnerability mapping is not a new concept, very little has been written about the specific risks faced by pregnant women. During a disaster, and often after the event has occurred, the victim suffers from an elevated stress load. Stress has been found to link directly, or indirectly through coping measures such as smoking, to negative birth outcomes.

This chapter describes measures that can be used to reduce this stress load, and by using the type of GIS incorporated at the Baton Rouge Healthy Start — (near) real-time help can be directed to those most at risk. The chapter ends with a suggestion as to how Homeland Security funding could be channeled to develop community health GIS systems in the name of syndromic surveillance.

Chapter XI — Guest Chapter by Jackie Mills

The goal of this chapter is to explain some of the differences a community health unit will face if they come from a more rural location.

The authors are fully aware that many community health centers are located in rural areas. Indeed, the Monroe Healthy Start in Louisiana works with rural counties. Although most GIS books tend to focus on either urban or rural (usually the former), many of the points raised in this text are still relevant to both, though with a little tweaking. We therefore asked Jackie Mills, who has worked on GIS and poverty in the Mississippi Delta for the last six years, to give us her insight into the particular problems likely to be faced by a rural health unit implementing GIS.

This chapter highlights the importance of understanding the geography of an area of analysis in regard to its urban or rural nature. This distinction has ramifications not only for types of health care issues specific to rural places, but in particular to the availability and applicability of data in a GIS environment. This chapter discusses the complexity of rurality, that a firm or best definition for being rural does not exist. Several agencies of the U.S. government provide definitions based on county or census designations, but the scale varies and these classifications must be researched in respect to your particular place and health care interests.

Some of the “risks” faced in rural areas will be distinctly different. For example, exposure to agricultural chemical applications has been noted as a health risk to pregnant and lactating women in rural areas. The agricultural structure (agristructure) of a rural community can also impact the availability of jobs, services, and overall well-being of rural places which impacts the age of pregnancy, health of pregnancy, and infant health in general. In addition, the existence of rural ghettos in the southern Cotton Belt of the U.S. has its own risks, much like those faced in urban ghettos, yet these places are not often considered when examining health issues in the rural south.

Finally, this chapter delves into the types of data that may be important in a GIS analysis of rural health issues. Concerns of confidentiality, availability, and scale are particular areas of note, as are some helpful ideas for making the best use of data.

Summary

I hope you enjoy this book. Working with the Baton Rouge Healthy Start has been a fulfilling experience. This experience has ranged from presenting at a congressional breakfast to working late nights with nurses and social workers as we plan the next step of our program. I have met some wonderful people along the way, and, hopefully, employed GIS to its maximum potential. But there is so much more that can be done. I would love to see a national approach by organizations such as Healthy Start providing the resources to encourage all community health groups to incorporate the technology. But for now, this book will have to do. If we have made any mistakes in the writing, as I am sure we have, I apologize. This was our best effort at covering a large subject in such a way to make it palatable and hopefully, a somewhat enjoyable read.

All writers royalties, as such they are, will be fed directly back into the research described in this project.

Andrew Curtis
Michael Leitner

Acknowledgments

There are many people who deserve to be acknowledged for their role in working with the Healthy Start program of Baton Rouge: all the ladies (and gents) working for Healthy Start over the last four years, all those at Family Road of Greater Baton Rouge, the MCH group in New Orleans, and our initial CityMatch taskforce. Special mention must go to Dena Morrison and Jamie Roques, who spent those long initial hours writing the original Healthy Start grant with me.

At Louisiana State University (LSU), special mention goes to Martin Hugh-Jones who has provided the epidemiology expertise through all stages of this project. In addition, Monika Arthold, Marina Mons, and Misty Richard have either worked with Healthy Start itself or data/Geographic Information System (GIS) aspects — we thank you all. Again, two special mentions must be made: Farrell Jones who designed the Healthy Start database and has spent many hours working on, perfecting, and tweaking it, and Jason Blackburn, who has been my right-hand man through all the WHOCC projects. Even though he is a PhD student, his GIS expertise already outdistances mine.

Andrew Curtis would also like to acknowledge the help and expertise offered over the years by Nina Lam at LSU, Charles Croner at the Centers for Disease Control and Prevention, my PhD advisor Stewart Fotheringham, and John Anderson, John Ernst, and Paul Farnsworth for their continued support.

Finally, thanks must be extended to Louisiana State University, which funded this research to varying degrees through a host of seed and travel grants.

Chapter I

Explaining the Geography of Infant Health

During a seminar presentation to the Geography and Anthropology Faculty at Louisiana State University, a series of summary statistics were presented concerning the racial inequality in infant health outcomes in Baton Rouge. The numbers were quite startling, spurring one colleague to question the accuracy of the findings, as, in their words, "that would put us at developing world levels." This sentiment was echoed by a National Science Foundation reviewer (I suspect from the United Kingdom) who read one of our proposals and expressed disbelief. The United States is a modern developed country, and yet in geographical pockets the infant mortality rate is alarmingly high. For some neighborhoods, between 1996 and 1998, in East Baton Rouge Parish the infant mortality rate was consistently above 40/1,000, reaching its highest rate in 1998 of 70/1,000. This means if 1,000 babies were born in this 0.25-mile neighborhood, 70 would die in the first year of life. Obviously, there is a problem. This book will investigate a host of ways to consider, and hopefully move towards solving, problems such as these using a Geographic Information System (GIS).

In this chapter we will provide an introduction to the main themes of the book, namely negative pregnancy outcomes and the role geography and geographic techniques can play in their improvement.

Geographic Variations in Infant Health

Infant mortality is often presented in introductory geography classes as a litmus test for the level of national development. The argument goes something like this: Most diseases, in fact most health conditions, are in some way reflective of the environment around us, the way we live, the food we eat, our standard of living, and the "insults" we experience (for a more detailed appreciation of the geography of health insults see Meade & Earickson, 2000). Yet the infant mortality rate, which is defined as the number of babies who die in the first year of life, usually expressed per 1,000 live births, is thought to be a reflection of the quality of health care, which in turn can be used as a proxy for the level of societal development.

There have been many successes within the United States. In general, infant mortality rates and prenatal risks continue to decrease. Trends in perinatal mortality (those babies that die in the first month of life), low birth weight, and short gestation births have also displayed downward trends, especially since the 1950s. Today the leading causes of infant mortality include Sudden Infant Death Syndrome (SIDS), congenital malformations, and low birth weight. What this book will show, especially in Chapter IV, is that these causes cannot be taken in isolation, that they are often expressions of a multitude of other risk factors that combine to form a single negative outcome. Many of these risks not only combine in the individual, but also across space in the form of shared neighborhood risks, making GIS and geographic approaches in general appropriate methods of investigation.

These spatial risk combinations occur at different geographic scales, displaying as regional patterns, as variations between cities, and as characteristics of individual neighborhoods (Anderson, 1952; J. C. Kleinman, 1986; Shannon & Pyle, 1993). At the national scale, Table 1 displays infant mortality rates for United States for the periods 1990, 1998, and 2000. If we consider improvements in outcomes, the infant mortality rate has decreased for all states over the three time periods, except for Hawaii (6.7 to 6.9 to 8.0), and North Dakota (8.0 to 8.6 to 9.2). The table also displays the rankings of the states, with five states appearing in the top ten for all of the three time periods: Alabama, Louisiana, Mississippi, North Carolina, and South Carolina.

Another way of looking at these data is geographically (Figure 1). A distinct regional pattern emerges, with the southern states (except Florida, which can be argued is no longer traditionally *southern*) displayed as a block of high infant mortality. The same pattern is not as evident for the 1998 and 2000 surfaces, though this weakening of the pattern is more an artifact of the same map classification breaks being kept. The interpretation of this is that although the southern states still perform worse than the rest of the United States, even this region has generally been improving in terms of the infant mortality rate.

Table 1. Infant mortality changes in the United States (1990, 1998, and 2000) and unlinked rate (per 1,000)

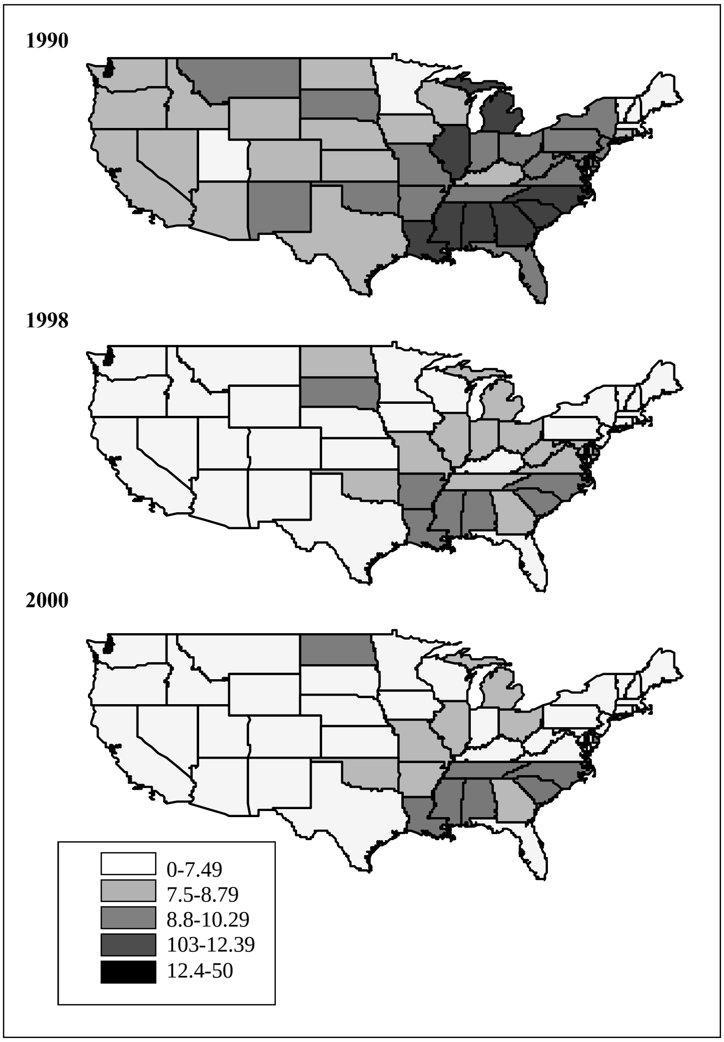
STATE	1990	Rank	1998	Rank	2000	Rank	Unlink
Alabama	10.8	5	10.2	1	9.8	3	0.0
Alaska	10.5	9	5.9	43	6.8	28	0.0
Arizona	8.8	27	7.5	21	6.5	31	8.9
Arkansas	9.2	22	8.9	8	8.0	14	3.4
California	7.9	41	5.8	45	5.5	46	19.2
Colorado	8.8	27	6.7	37	6.2	37	0.0
Connecticut	7.9	41	7.0	31	6.2	36	3.7
Delaware	10.1	12	9.6	3	9.2	5	9.3
District of Columbia	20.7		12.5		10.0	2	6.6
Florida	9.6	16	7.2	27	7.0	24	0.7
Georgia	12.4	1	8.5	11	8.4	11	0.0
Hawaii	6.7	48	6.9	36	8.0	15	28.4
Idaho	8.7	29	7.2	25	6.6	30	0.0
Illinois	10.7	6	8.4	13	8.2	13	8.0
Indiana	9.6	16	7.6	20	7.4	19	15.3
Iowa	8.1	37	6.6	38	6.1	39	0.0
Kansas	8.4	32	7.0	31	6.9	25	18.4
Kentucky	8.5	31	7.5	21	6.9	26	8.0
Louisiana	11.1	4	9.1	6	9.0	7	29.4
Maine	6.2	50	6.3	41	5.1	49	14.7
Maryland	9.5	20	8.6	10	7.4	18	19.3
Massachusetts	7.0	47	5.1	49	4.7	51	7.7
Michigan	10.7	6	8.2	14	8.3	12	0.9
Minnesota	7.3	45	5.9	43	5.6	44	2.7
Mississippi	12.1	2	10.1	2	10.0	1	0.0
Missouri	9.4	21	7.7	18	7.8	17	4.9
Montana	9.0	24	7.4	23	5.9	42	0.0
Nebraska	8.3	34	7.3	24	7.2	21	5.6
Nevada	8.4	32	7.0	31	6.4	33	36.1
New Hampshire	7.1	46	4.4	50	5.4	48	13.3
New Jersey	9.0	24	6.4	39	6.1	38	34.8
New Mexico	9.0	24	7.2	25	6.3	34	64.7
New York	9.6	16	6.3	41	6.3	35	16.5
North Carolina	10.6	8	9.3	5	8.6	9	4.8
North Dakota	8.0	40	8.6	9	9.2	6	0.0
Ohio	9.8	15	8.0	16	7.9	16	40.7
Oklahoma	9.2	22	8.5	11	8.5	10	74.5
Oregon	8.3	34	5.4	48	6.0	40	0.0
Pennsylvania	9.6	16	7.1	30	7.4	20	2.8
Rhode Island	8.1	37	7.0	31	6.8	27	0.0
South Carolina	11.7	3	9.6	3	9.0	8	0.0
South Dakota	10.1	12	9.1	6	5.9	41	0.0
Tennessee	10.3	10	8.2	14	9.4	4	0.0
Texas	8.1	37	6.4	39	5.6	45	29.3
Utah	7.5	44	5.6	47	5.8	43	14.2
Vermont	6.4	49	7.0	31	6.7	29	0.0
Virginia	10.2	11	7.7	18	7.1	22	11.7
Washington	7.8	43	5.7	46	5.4	47	2.3
West Virginia	9.9	14	8.0	16	7.0	23	0.0
Wisconsin	8.2	36	7.2	27	6.5	32	0.0
Wyoming	8.6	30	7.2	27	4.8	50	0.0

Source: U.S. National Center for Health Statistics, Table 125, 2000 Statistical Abstract of the United States, CDC Wonder Table 6 Linked Birth and Death File.

Note: Infant mortality for 2000 means the baby was born in that year, irrespective of whether the death was in 2000 or 2001.

When considering pregnancy outcomes at this largest of geographies, although regional patterns do emerge, no alarm bells immediately sound. It is when we change the focus of investigation either spatially, or racially, that concern mounts. When we zoom into smaller areas of interest, perhaps at the county level, or even at a finer resolution of zip code or neighborhood, negative pregnancy outcome rates can rise dramatically. The reason these problem areas are masked within an aggregate display of data is because extreme values are "smoothed" out. For example, for the United States as a whole, the more affluent northern regions moderate the southern pattern, seen in Figure 1. Within the South, some states will fare better than others. Within the worst states, such as

Figure 1. Infant mortality rates for all states for the years 1990, 1998, and 2000



Mississippi or Louisiana, other geographic areas of affluence will moderate problem counties and parishes. Even within a single county, maybe one with a large urban area, the suburbs will smooth out the negative health outcomes of the inner city neighborhoods. For example, although the low birth weight and infant mortality rate in Illinois during the 1990s was 7.6% and 10.7/1,000 respectively, within some neighborhoods of Chicago these numbers rose to 19.5% and 31.9/ 1,000 (Roberts, 1997). The definition of what constitutes a neighborhood will be discussed in Chapter III, though a 0.25-mile neighborhood in Baton Rouge presented in Chapter IV had a greater than 60/1,000 infant mortality rate for each of the 3 years under investigation.

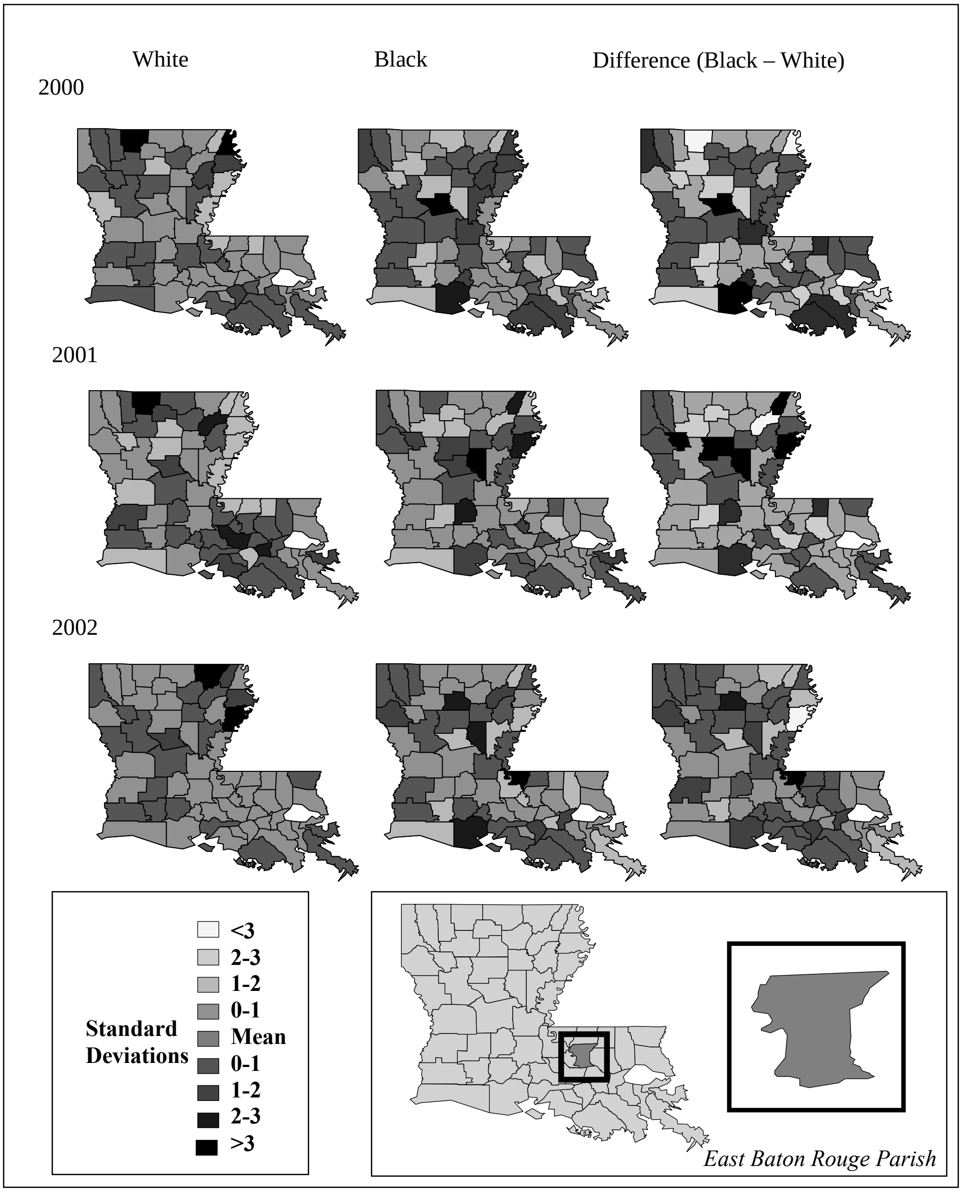
It is important for the reader to appreciate how geographic scale affects analysis results. This topic will be revisited throughout this book, but it is important to recognize that results will vary even within a city space, depending on how that area is divided up.

The inner city environment may traditionally suffer from higher infant mortality rates than the suburbs (Pearl, Braveman, & Abrams, 2001; Rauh, Andrews, & Garfinkel, 2001). Reasons for these inner city "risk areas" will be described in Chapter IV, and usually include unemployment, poverty, high crime, substance use, and poor quality housing (Wright et al., 2004). Unfortunately, the proportion of African Americans residing in these neighborhoods is also far higher than in suburban areas, making urban social problems and negative health outcomes (including birth outcomes) to often be seen as racial problems. This results in a large disparity in negative pregnancy outcomes, such as low-birth-weight deliveries, between whites and minorities in the United States, and especially between whites and African Americans.

However, due to the mixing of genetic traits, it has been positioned that race is not really the association, but a mixture of economic, environmental, and behavioral traits accompanying what are traditionally known as African American populations (Farmer & Ferraro, 2005; Krieger, 1993; Krieger, Rowley, Herman, Avery, & Phillips, 1993; Pappas, 1994; Richardus & Kunst, 2001; Santelli, Lowry, Brener, & Robin, 2000; Whaley, 2003). For example, Roberts (1997) found that even the mix of race within a neighborhood could influence outcomes such as low birth weight, with neighborhoods with higher proportions of African Americans having more established support networks than would be found for African American families in relative isolation. The complex issue of race as a risk variable will be discussed in more detail in Chapter IV. For now we'll just accept that there is a disparity in those populations recording race as African American on the birth certificate, as compared to those recording white (Blackmore et al., 1993; David & Collins, 1991; J. Kleinman & Kessel, 1987; LaVeist, Keith, & Gutierrez, 1995), and that there is a geography to that difference.

Figure 2 displays infant mortality rates between the races in Louisiana for the years 2000-2002. A third map, showing the difference between the two rates (African American infant mortality rate minus white infant mortality rate) is also displayed. Unfortunately, these maps are hard to interpret because of a statistical condition called the *law of small numbers*.

Figure 2. Infant mortality rates in Louisiana 2000 to 2002



Source: Ogada and Tran, 2002

Simply put, the proportion of African Americans residing in the area will influence the infant mortality rate due to their actual number, rather than any association between race and outcome. This can lead to a somewhat misleading relationship when presented graphically, such as in the maps in Figure 2 because although areas with large African American populations will present stable high infant mortality rates, it is actually the areas of small African American populations that can have the highest infant mortality. In other words, if one baby dies, the total birth population is so low that an excessively high rate results.

Consider Table 2, which shows the three parishes with the highest African American infant mortality rate in 2000. The percentage of African Americans in any of the three parishes does not rise above 14%. The highest African American population is just over 10,000 in LaFourche, but is only 2,500 in Grant. The low population in Grant explains how the infant mortality rate fluctuated from 38.5/1,000 in 2000 to 0 in 2002.

And yet not all patterns can be explained away by this low denominator population. Vermilion has a stable high infant mortality rate for all 3 years, and a low African American population of around 7,000. The low number of births this population would generate will always make the infant mortality rate suspect, and yet one would expect more variation than is evident. Geography is important, but the geography is complex. The last chapter of this book will concentrate on the care that must be taken when using a GIS to investigate rural pregnancy health outcomes because of this small numbers problem.

For parishes with sizeable urban areas (and therefore sizeable African American populations) the infant mortality rates and disparities between races can be viewed with a degree of confidence due to stability over time. This is the case with East Baton Rouge Parish (which contains Baton Rouge), Orleans Parish (which contains New Orleans), Lafayette Parish (which contains Lafayette), and Caddo Parish (which contains Shreveport).

The majority of this book will focus on the fourth of these parishes, East Baton Rouge, home of a Healthy Start program designed specifically to eliminate racial disparities. This parish is relatively typical of many other U.S. cities, especially southern cities. There is no definable inner-city poverty core as one might find in a northern city, though the zip codes of concern are largely coalesced. However, within these zip codes, the urban area is fragmented by neighborhoods with poor birth outcomes being interspersed with relative affluence. More

Table 2.

	2000 Black IMR	2001 Black IMR	2002 Black IMR	% Black
Grant	38.5	29.4	0	14.49275
Vermilion	31.1	33.6	43.7	13.19949
LaFourche	28.6	12.2	18.7	11.88179

Source: Ogada and Tran, 2002

Table 3.

	2000 Black IMR	2000 White IMR	2001 Black IMR	2001 White IMR	2002 Black IMR	2002 White IMR	%Black
Orleans	7.7	3.8	12.4	3	14.9	6.8	22
Lafayette	21.4	2.6	15.4	6.3	23.4	3.1	40
Cado	20.7	4.4	19	5.8	17.7	8.3	35

Source: Ogada and Tran, 2002

Table 4.

	EBR1996	EBR1997	EBR1998	EBR2000	EBR2001	EBR2002
Infant Mortality Rate						
All	11	11.7	12	9.2	10.2	10.4
African American	14.6	19.4	18.9	13	13.5	14.9
White	6.7	3.5	4.6	5.2	6.5	4.9
Preterm Delivery (20-36 weeks) %						
All	11.2	11.8	11.1	11.9	11.2	12
African American	15.3	16	16.3	14.8	13.8	14.7
White	6.9	7.4	6.1	8.4	8.2	8.7
No Prenatal Care in First Trimester %						
All	19.8	19.9	19.5	18.7	18.6	18.1
African American	31.1	30.6	30.4	28.1	28	26.1
White	8.7	8	8.1	7.3	7.6	7.9

traditional suburban developments have spread through the southern section of the parish, though one risk neighborhood is found even here disconnected from the urban core. A further area of risk is found to the north of the city following the industrial corridor that borders the Mississippi River. There are approximately 6,000 births in the parish every year, of which almost exactly 50% are African American.

Table 4 clearly displays the disparity between African American and white birth outcomes in the parish. For none of the years displayed did the disparity in infant mortality rates drop below a 2:1 ratio, and at its worst it exceeded 5:1. The disparity in preterm births was not as marked, ranging from slightly less than 2:1 to slightly less than 3:1. However, it is the disparity in those receiving prenatal care after the first trimester that is most consistent, never once dropping below a 3:1 ratio. The argument for this last disparity is partly one of economics, as many African American women in the parish rely on Medicaid payments for prenatal visits. Anecdotal evidence suggests some doctors postpone visits until the source of payment is secured. Unfortunately, the waiting period to apply for and receive Medicaid is often 30 days or more, though measures are currently in place to limit this delay. Presuming the woman is not aware of her pregnancy until the sixth week, this processing time almost puts her into the second trimester before she is finally able to make her first visit. All things being considered, the negative impact Medicaid application might have in prolonging initial prenatal care is relatively easy to understand. Other risk combinations are far more complex.

Of course, not all disparity can be attributed to racial difference and the inner city/suburban divide. Chapter XI discusses the problems associated with working with rural populations, and these "peripheral" areas often suffer similar disparities in birth outcomes compared to the metropolitan "core" (Clarke, Farmer, & Miller, 1994). For now though, let us introduce three tasters showing how complex the pregnancy risk surface can be. Chapter IV will further develop these points into a disease ecology framework of biological, behavioral, economic, social, cultural, and environmental risks (Mayer & Meade, 1994).

Smoking Is Bad

Smoking is bad for everyone — not just for the mother, fetus, and any other child in the household, but for all family members. Considerable research now supports the negative impact smoking can have on health. Chapter IV will detail how smoking can cause low-birth-weight deliveries, which in turn can increase the chance of an infant death. It is therefore remarkable that any mother smokes during pregnancy, or for that matter, that any other member of the family exposes the fetus to second hand smoke (Blackburn et al., 2003; Wakefield, Reid, Roberts, Mullins, & Gillies, 1998). Chapter IV will return to this point by showing how many mothers currently enrolled in the Healthy Start Program reside with smokers.

If it is so bad, why not stop? The situation is far more complex than simply a behavioral one. Indeed, a common criticism of many GIS style approaches to understanding smoking risk is that the underlying causations are dismissed in the search for "hot spots," this over reliance on data being referred to as "structural functionalism" (Litva & Eyles, 1995). There is validity in this argument as the social, economic, and political neighborhood backcloth certainly plays a role in shaping individual behavior beyond a yes or no choice (Eyles & Woods, 1983; Jones & Moon, 1987).

For example, a mother who smokes may do so as part of a coping mechanism. What if she is a victim of abuse — what other pressures, either chronic or acute, affect her? It is a vicious circle, for if the mother lives in a stress-inducing environment, which, either through fears for safety or because of a general lack of activities, keeps her and her other children inside, the effect of smoking is magnified. A further problem of being confined inside in poorer quality housing is that respiratory problems such as asthma can be accentuated, for example through exposure to roach feces. And, of course, smoking compounds these problems. If she is scared to leave the house, if she does not know where the next meal will come from, or if smoking is an alternative to other substance use, how should the education strategy be designed? Chapter IV will briefly mention examples of different programmatic approaches to reduce smoking in these types of neighborhoods. The situation is obviously more than just "smoking is bad."

What Does It Mean to Be Poor?

Social deprivation has frequently been linked to negative health outcomes (Townsend, Simpson, & Tibbs, 1985). Health care options are often more limited. The woman is probably reliant on federal and state assistance programs, which could limit the number of prenatal visits and when these visits begin. A poor family may have limited transport to the clinic. Of course, the clinic may also run on limited hours because of being situated in a dangerous neighborhood. Being poor extends far beyond the economics of the situation, as there is also a social dimension. A mother in poverty, especially if she is single, is more likely to depend on neighborhood and family structures. If she resides in a neighborhood where these resources are either stretched or hampered, possibly because of crime, the mother may find her ability to provide for her family or to attend prenatal visits even harder. Nutrition could also be compromised. It is possible that the mother has not had sufficient education to know what foods she should be eating during the pregnancy; however, the habits of a lifetime imposed by the surrounding neighborhood infrastructure will also play a part (LaVeist, 1990). Obesity is a problem of the current age, although this can be in part blamed on a sedentary lifestyle (again as a result of neighborhood crime or lack of local opportunities). Obesity has also been found to accentuate the effects of asthma. Part of the problem is an actual lack of nutrition opportunity. Stores found in poor neighborhoods are less likely to carry the choices provided by suburban supermarkets, and local restaurants are more likely to be fast food chains. Add to this the impact of large competitors forcing smaller local chains out of business because of their lower prices. Unfortunately, these "super" stores are located by necessity in open areas of land served by large roadways. The family in poverty may not have access to transport.

And then there are the physical problems of the poor neighborhood, the excessive trash, pests and vermin within the house, and broken garden fences allowing young children to wander into roads which are often poorly marked and cared for. The neighborhood may also find itself close to other polluting or unfavorable activities, such as a landfill, manufacturing or industrial plants, and major arterial roadways, partly due to the limited political bargaining power of the neighborhood, and partly because of the lack of financial ability of the family to move.

Stress

Stress has been linked to negative birth outcomes. Again, these pathways will be more fully developed in Chapter IV, though it is important to realize that stress can also be part of the “smoking is bad” and “living in poverty” life situations. A woman (or her partner) may smoke because of the chronic and acute stressors they face in their day-to-day activities, both in the family and from the neighborhood in which they live.

There are multiple reasons why a mother might experience stress. Many of these are also linked to her economic status. The immediate environment may cause her stress — from crime, living in close proximity to a busy road, or living in the shadow of an industrial plant. Crime would cause stress because of the fear of harm to herself or her immediate family. This might result in the mother not leaving the house. This in turn could have two further impacts on the pregnancy: The mother may miss or have fewer than preferred prenatal visits because of the perceived risk of leaving the house, or the mother may lack exercise (as might other children in the house) due to the perceived need to stay within the safety of the house. Unfortunately, mothers living in poverty are more likely to experience crime firsthand. Poor neighborhoods have higher crime rates for a variety of reasons, including lack of opportunity, lack of a deterrent in terms of going to jail, and even police redlining — the lack of an active effective police presence. Crime has also been linked to environmental variables more likely to be found in poorer neighborhoods. For example, lead exposure has been associated with a variety of individual characteristics which result in personality traits leading to crime. Why mention lead? Well, the mother and children who are forced to stay in the house because of fear are more likely to suffer exposure from lead-based products, including lead paint, which are more likely to be found in the older houses of indigent neighborhoods. Therefore, a cycle could be in effect. Children are more likely to be exposed in high crime areas, and the exposure may result in their eventual fulfilling of the crime cycle. For example, associated personality traits include Attention Deficit Disorder-like symptoms, which could result in education problems and a lack of opportunity to escape the cycle.

Lead exposure does not necessarily have to result from older housing; airborne exposure from living close to busy roads or close to heavy industry could also be factors. Yet again, both of these environmental exposures are linked to poverty. The family in poverty has less ability to leave the area, therefore the pregnant woman has to endure the stress of living next to the busy road, and the other possible health exposures this can bring. Similarly, heavy industry and poorer neighborhoods all too often tend to coexist. The poor do not have the luxury of being able to relocate, and other wealth certainly does not want to move in, leaving the neighborhoods to slowly decay.

Of course, stress comes in many other forms: not having adequate finances, not having a partner, not being able to care for the children at home, being the victim of domestic abuse. It is hardly surprising that smoking during pregnancy as a coping mechanism is a more common approach than one would normally have thought (Sheehan, 1998).

All of these points will be discussed again in more detail in Chapter IV. Chapter X will provide a further example of how an external stressor, such as a disaster, can compound these stresses into a negative health footprint that lasts long after the event.

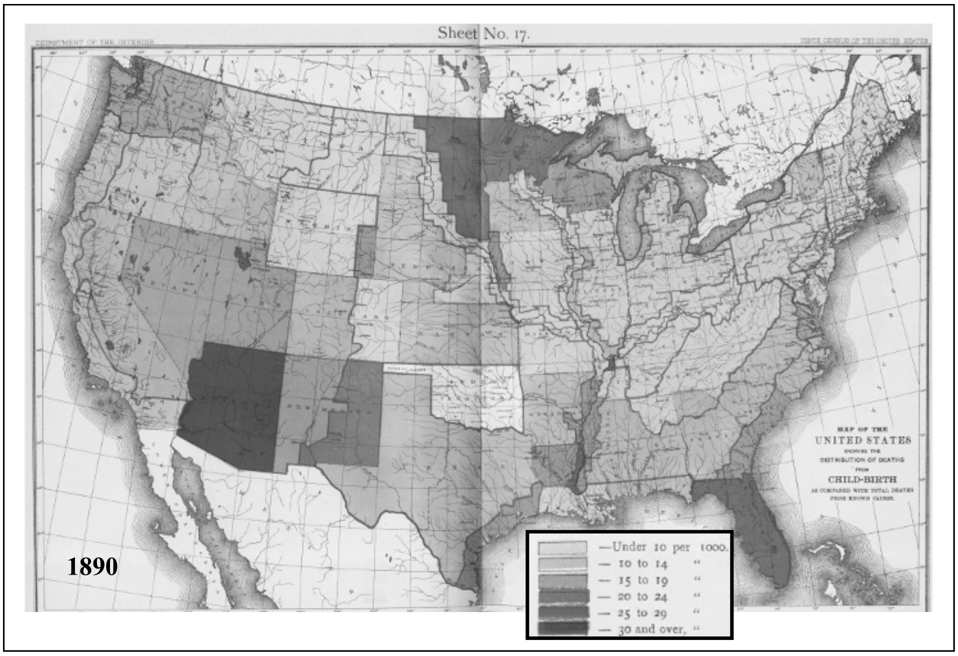
So where does geography and GIS come into this? First, geographers have long worked as a clearinghouse for other disciplines, searching for the spatial interconnections in the different investigations. The previously mentioned risk components have been extracted from disciplines including sociology, environmental toxicology, criminology, geography, and of course, public health. Second, GIS provides the resource to bring these different disciplines and ideas together in a single research frame, with research not being limited to the ivory tower, but being possible by local community health groups. Part of the mission of this book is to provide discussion points and ideas for these types of users.

Think again about the different risks mentioned, from smoking to stress. If we don't consider individual health data, such as asthma, lead screening, and behavioral pregnancy information, we can use the GIS to map where older housing can be found (as a proxy for lead exposure), where industrial polluters are located, where traffic accidents have happened, where different arrests have occurred, where clinics, schools, and stores are located, and if these fall on bus routes. We can also map a multitude of socioeconomic data, ranging from income levels to education. If we then place our pregnant program participants onto these surfaces, an invaluable impression of where these women walk can be gained.

The Geography of Health

As mentioned in the previous section, GIS now makes the (spatial) investigation of health issues an easier prospect. This is not to say that this type of conceptual approach is a recent phenomenon. It may have taken far longer to make the maps, construct the tables, and draw the histograms, but many of the forms of visualization (including overlays) found in a GIS can also be found in Board of Health reports from the later 1800s. As an example, Figure 3, displays infant mortality rates by state for 1890 (taken from the U.S. Mortality Census of that year).

Figure 3.

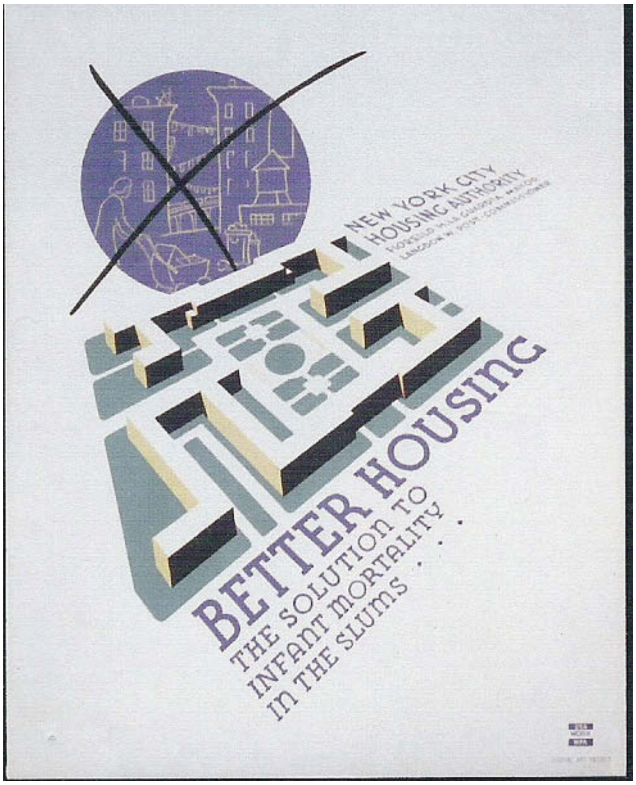


At the 2004 National Healthy Start Association's regional conference for the Southwest, one workshop introducing the "Healthy Start Guide to Program Excellence and Community-Based Interventions" quoted from a 1912 report on infant mortality, from the newly established "The Children's Bureau." The quote said:

…coincidence of a high infant mortality rate with low earnings, poor housing, the employment of the mother outside the home, and large families was indicated in these studies. They all showed great variation in infant mortality rates, not only in different parts of the United States but also in different parts of the same state and in the same city, town, or rural district. These differences were found to be caused by different population elements, widely varying social and economic conditions, and differences in the appreciation of good prenatal and infant care and the facilities available for such care.

The impact of geography is quite obvious in this quote. In fact, the statement could easily have been written as a comment on the current condition in the United States. The report also detailed the racial disparity in infant mortality rates, at a 2:1 ratio, with African Americans suffering an infant mortality rate of

Figure 4.



200/1,000. The rates may have gotten better, but that disparity still exists. The message of Figure 4 is also as applicable today as when this public health announcement was distributed in the pre-1940s, "Better Housing: The solution to infant mortality in the slums."

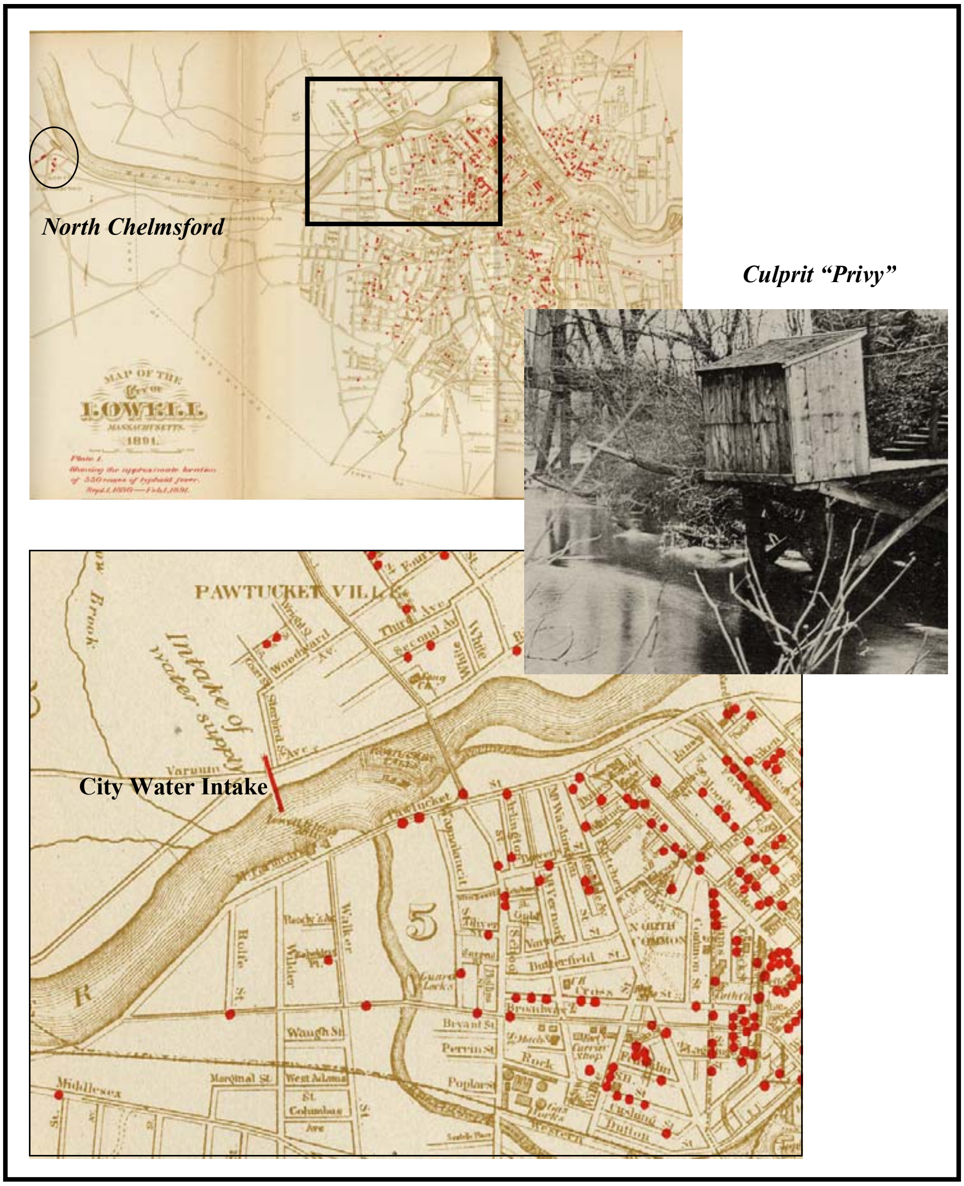
Early Board of Health reports often went beyond a simple visualization of disease events to using maps and overlays to gain an understanding of how geography affected health. In some ways this initial spatial investigation of health is analogous to current GIS use in the health sciences. While some simply use the technology to produce maps, others seek insight in the interrelationship of geographic features and health events. The reader must now bear with me as I give an example of how geography and health were brought together in health investigation.

By the end of the 1800s, the conceptualization of germ theory, in combination with an understanding that water could act as a spreading mechanism, led to the creation of some of the first exploratory spatial analysis maps. An underlying premise to these maps of spatial association was that geography played a role in disease occurrence, and that events did not occur randomly. This comes as no surprise to students of geography who are familiar with Tobler's laws which, when paraphrased, state that objects and events are connected across space, and that this connection is stronger the closer the object or event is. Even when the underlying science was wrong — for example, that poor health had a miasmatic causation (such as proximity to a marshy area) — the driving rationale was that space played a part in the illness, that there was a pattern to be found between the environment and sickness. Possibly the most famous example of environmental epidemiology, or exploratory spatial data analysis, occurred in London in 1854, when John Snow created a dot map of cholera distribution. Snow did not understand *Vibrio Cholera*, indeed his work was before germ theory, and yet through inductive reasoning, exploratory analysis, and shoe-leather epidemiology he identified a spatial process. He plotted cases of cholera and from the resulting spatial distribution ordered the Broad Street pump handle to be removed. A similar logic applies today: Use the map to understand the disease, and then either stay away, take precautions, or investigate the causation so it may be mitigated. John Snow's map has been reproduced countless times as an example of how spatial analysis can solve a spatial riddle; indeed, it adorns the cover of a recent spatial analysis text, *Geographic Information Analysis* (O'Sullivan & Unwin, 2002). However, it is by no means the only example of exploratory spatial analysis applied to disease in the 1800s.

The 1892 Board of Health report for Massachusetts describes a typhoid epidemic which hit the city of Lowell between 1891 and 1892. Although the water systems of the city were considered sufficiently polluted to warrant calling the area "endemic" with typhoid, this epidemic was particularly severe, in fact the worst in 40 years (550 cases, 92 deaths). The epidemic peaked in November (122 cases, 28 deaths), as compared to the more usual months of September and October. In order to discover how the epidemic happened, so that recommendations could be made for future prevention, a house-to-house survey was administered. This survey contained a time frame for the disease (either the doctor's first visit or the first time the victim became bed ridden), and a spatial frame (address of the victim). In actuality the data collected is extremely spatial in nature, and would have made a wonderful GIS. Other questions on the survey with spatial implication included present address, address during illness, place of work, business, or school, source of drinking water, previous cases in the house, other proximate cases, and whether the victim had a privy or WC.

These "attributes" could have easily been turned into a GIS. The map produced for the Board of Health report (Figure 5) displays many similarities with a modern investigation. Spatial attribute data had been collected by survey. These surveys were "address-matched" to a road network to create a dot distribution map. Additional key features were heads-up, which simply means adding each location into the GIS by a mouse click, digitized onto the map (such as the city water intakes, which were colored red). Five water supplies were analyzed physically by taking samples, and then spatially by comparing individual survey responses to the source of their water needs. These water supply data were then "overlaid"

Figure 5. Spatial epidemiology of the 1891 typhoid fever epidemic in Lowell, MA



onto a map of actual disease cases. This approach identified the primary source of infection as being the Merrimack River, which had been infected by disease cases in North Chelmsford during September and October, with excreta literally dropping from privies into the water below. As the epidemic took hold in Lowell, other water sources became infected as people's waste passed into the water supply of those waiting downstream. As the Board of Health report puts it…

…when the excreta from many cases, especially mild cases and cases in the early stages of the disease, poured in from the sewers on Walker and School streets, further enriched by the discharges from the Lowell Hospital and various privies, and finally added to the excreta from thousands of workers in the mills on the upper level, must have infected the canal water. (Twenty-Fourth Annual Report of the State Board of Health of Massachusetts, pp. 676-677)

Although this last section may seem a little tangential to a modern GIS analysis of infant health, it provides a useful example of the spatial epidemiological approach; indeed, the survey leading to a map and spatial analysis bears similarities to the process employed in the creation of the Baton Rouge Healthy Start GIS. Imagine the spatial queries that could have been performed back in the 1890s had such a GIS been available: *Identify all houses (by date) with a first case, identify all workplaces of houses with a first case, identify all water sources for houses with a first case.*

So, a geographic appreciation in epidemiology is not new. Nor is our general appreciation of the public's perception of proximity and disease occurrence. Back in the late 1800s city residents would avoid neighborhoods populated with immigrants, whom they considered to be unclean. The motto of the day was "Public Health is Public Wealth," and this included either cleaning, or more likely, staying away from the squalid and cramped conditions of the newly landed immigrants (or *strangers* as they were referred to in New Orleans neighborhoods). And there was a logic to their fear, as these "strangers" would not have been exposed to the local diseases and therefore would have no immunity. In addition, the unsanitary conditions and high population density would have provided suitable breeding grounds to the communicable diseases of the day. Today the inner city neighborhoods have replaced the immigrant slums in our psyche, with typhoid and cholera being replaced by HIV and infant health problems. Our environmental fears have moved from the bilious swamps to focus on cancer, with a constant outcry for researchers to study the dangers of living in *the cancer alley*, as the stretch of petrochemical plants following the Mississippi through Louisiana is euphemistically called. The point is that society has always had a well-developed appreciation of the role geography plays in health disparities, fueled in the past by inductive logic, and today by the constant reporting on CNN, and films such as *A Civil Action* and *Erin Brockovich*.

What has changed is our understanding of disease causation, and our ability to effect change. We now have the techniques and technology that allow us to go beyond a visual presentation, and perform sophisticated exploratory spatial data analysis in real time. Geographic Information Science (GISc), which includes the software of GIS and Remote Sensing, the science of spatial investigation, and peripheral developments such as Global Positioning Systems, has revolutionized the way academics can search for spatial patterns. And this revolution is slowly but surely working its way into all branches of federal, state and local government, including the offices of Public Health. Indeed, one of the two major goals of Healthy People 2010 is to eliminate disparities in health, and an objective to reach that goal is to "increase the proportion of all major national, state, and local health data systems that use geocoding to promote nationwide use of geographic information systems (GIS) at all levels" (http://www.healthypeople.gov/document/html/objectives/23-03.htm). This increase in GIS use is targeted to include 90% of all health units by 2010. Other objectives include the continued reduction of fetal and infant deaths, low- and very low-birth-weight deliveries, and preterm births.

Geography has always instinctively mattered, but now we are being *told* it matters. This book is designed to help nurture this geographic appreciation and explain how these new technologies and techniques can be used to understand patterns of poor birth outcomes, and in so doing actually affect change. Most of the GIS examples given will involve the creation and running of the Baton Rouge Healthy Start program. GIS was initially used to target the service zip codes of the program. GIS was then used to develop neighborhood risk profiles so as to help understand the pressures and problems the "program participants" faced. Once the program was up and running, a GIS was created to store all program participant data. Now this data-rich GIS is being used to gain an even greater insight into the problems facing the pregnant women (and to some degree the citizens in general) of its service region, as the program moves into its second stage.

References

Anderson, O. W. (1952). *Infant mortality and cultural factors: Historical trends and current patterns*. Washington, DC: U.S. Department of Health, Education, and Welfare.

Blackburn, C., Spencer, N., Bonas, S., Coe, C., Dolan, A., & Moy, R. (2003). Effect of strategies to reduce exposure of infants to environmental tobacco smoke in the home: A cross sectional survey. *British Medical Journal, 327*, 257.

Blackmore, C. A., Ferre, C. D., Rowley, D. L., Hogue, C. J., Gaiter, J., & Atrash, H. (1993). Is race a risk factor or a risk marker for preterm delivery? *Ethnicity and Disease, 3*(4), 372-377.

Clarke, L. L., Farmer, F. L., & Miller, M. K. (1994). Structural determinants of infant mortality in metropolitan and nonmetropolitan America. *Rural Sociology, 59*, 84-99.

David, R. J., & Collins, J. W., Jr. (1991). Bad outcomes in black babies: Race or racism? *Ethnicity and Disease, 1*(3), 236-244.

Eyles, J., & Woods, K. J. (1983). *The social geography of medicine and health*. Andover, Hants: Croom Helm.

Farmer, M. M., & Ferraro, K. F. (2005). Are racial disparities in health conditional on socioeconomic status? *Social Science and Medicine, 60*(1), 191-204.

Jones, K., & Moon, G. (1987). Medical geography: taking space seriously. *Progress in Human Geography, 17*, 515-524.

Kleinman, J., & Kessel, S. (1987). Racial differences in low birth weight: Trends and risk factors. *New England Journal of Medicine, 317*, 31-39.

Kleinman, J. C. (1986). State trends in infant mortality, 1968-83. *American Journal of Public Health, 76*(6), 681-688.

Krieger, N. (1993). Analyzing socioeconomic and racial/ethnic patterns in health and health care. *American Journal of Public Health, 83*(8), 1086-1087.

Krieger, N., Rowley, D. L., Herman, A. A., Avery, B., & Phillips, M. T. (1993). Racism, sexism, and social class: Implications for studies of health, disease, and well-being [Suppl.]. *American Journal of Preventative Medicine, 9*(6), 82-122.

LaVeist, T. A. (1990). Simulating the effects of poverty on the race disparity in postneonatal mortality. *Journal of Public Health Policy, 11*(4), 463-473.

LaVeist, T. A., Keith, V. M., & Gutierrez, M. L. (1995). Black/white differences in prenatal care utilization: An assessment of predisposing and enabling factors. *Health Services Research, 30*(1), 43-58.

Litva, A., & Eyles, J. (1995). Coming out: Exposing social theory in medical geography. *Health and Place, 1*, 5-14.

Mayer, J. D., & Meade, M. (1994). A reformed medical geography reconsidered. *Professional Geographer, 4*, 103-106.

Meade, M. S., & Earickson, R. (2000). *Medical geography* (2nd ed.). New York: Guildford Press.

Ogada, E., & Tran, T. (2002). *Louisiana maternal and child health data book*. New Orleans, LA: Louisiana Department of Health and Hospital, Office of Public Health, Maternal and Child Health Program.

O'Sullivan, D., & Unwin, D. (2002). *Geographic information analysis*: Hoboken, NJ: Wiley.

Pappas, G. (1994). Elucidating the relationships between race, socioeconomic status, and health. *American Journal of Public Health, 84*(6), 892-893.

Pearl, M., Braveman, P., & Abrams, B. (2001). The relationship of neighborhood socioeconomic characteristics to birthweight among 5 ethnic groups in California. *American Journal of Public Health, 91*(11), 1808-1814.

Rauh, V. A., Andrews, H. F., & Garfinkel, R. S. (2001). The contribution of maternal age to racial disparities in birthweight: A multilevel perspective. *American Journal of Public Health, 91*(11), 1815-1824.

Richardus, J. H., & Kunst, A. E. (2001). Black-white differences in infectious disease mortality in the United States. *American Journal of Public Health, 91*(8), 1251-1253.

Roberts, E. M. (1997). Neighborhood social environments and the distribution of low birth weight in Chicago. *American Journal of Public Health, 87*, 597-603.

Santelli, J. S., Lowry, R., Brener, N. D., & Robin, L. (2000). The association of sexual behaviors with socioeconomic status, family structure, and race/ethnicity among US adolescents. *American Journal of Public Health, 90*(10), 1582-1588.

Shannon, G., & Pyle, G. (1993). Infant Mortality. In G. Shannon, & G. Pyle (Eds.), *Disease and medical care in the United States* (pp. 82-92). New York: Macmillan.

Sheehan, T. J. (1998). Stress and low birth weight: A structural modeling approach using real life stressors. *Social Science Medicine, 47*(10), 1503-1512.

Townsend, P., Simpson, D., & Tibbs, N. (1985). Inequalities in health in the city of Bristol: a preliminary review of statistical evidence. *International Journal of Health Services, 15*(4), 637-663.

Wakefield, M., Reid, Y., Roberts, L., Mullins, R., & Gillies, P. (1998). Smoking and smoking cessation among men whose partners are pregnant: A qualitative study. *Social Science Medicine, 47*(5), 657-664.

Whaley, A. L. (2003). Ethnicity/race, ethics, and epidemiology. *Journal of the National Medical Association, 95*(8), 736-742.

Wright, R., Mitchell, H., Visness, C., Cohen, S., Stout, J., Evans, R., et al. (2004). Community violence and asthma morbidity: The inner-city asthma study. *American Journal of Public Health, 94*(4), 625-632.

Chapter II

An Introduction to GIS (All Things Data)

One of the purposes of this book is to introduce community health groups to the potential of GIS, a technology that can help in understanding the spatial landscape of prenatal risk. Therefore, one of the first steps is to provide a brief overview of GIS. This introduction will be split over the next two sections, with this chapter focused on data issues associated with using a GIS, while the next presents an introduction to the functions of a GIS that make it so powerful: the ability to analyze and visualize spatial data. These two chapters are not meant for experienced GIS users (though even for these a few points and references from the non-geographic literature may prove to be useful). It is also not meant to be a comprehensive introduction to the science; there are several other excellent texts serving that need. These next two chapters are meant to give a basic understanding, and inform enough to encourage the adoption of a GIS approach.

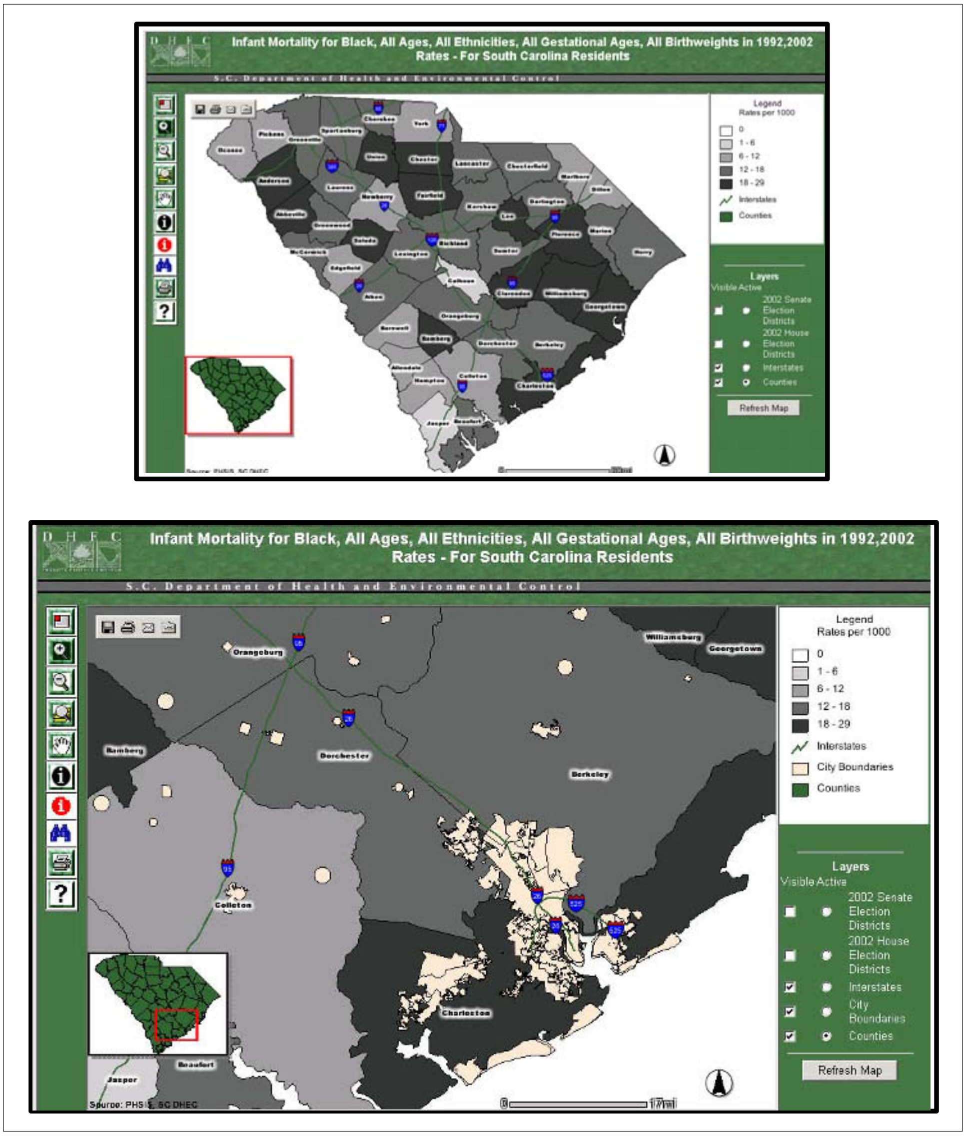
Most people reading this book will probably be using a vector GIS. There are, however, two basic GIS formats, raster being the other type. Raster GIS is best suited for surface or complete coverage data (for example, vegetation cover) because the spatial surface is transformed into a grid, with each cell or pixel containing a relevant geographic attribute (such as 1 = water, 2 = forest, etc.). These pixels are fixed in space, allowing multiple layers at a single location to be compared and analyzed. Vector GIS, which is a more useful GIS format for the type of investigation likely to be performed by a health unit, contains points, lines, and areas. Unlike in a raster GIS, each spatial object has its own geography. The mother's residence is a point on the city map, her bus route is a line, and she resides in a zip code, which in vector GIS-speak is a polygon. Most of the following descriptions will refer to vector GIS topology.

Many texts have been written defining and describing the constituent components of a GIS. There is no point in replicating these here, though the reader is recommended to (Longley, Goodchild, Maguire, & Rhind, 2001) if he or she wishes to gain a better understanding of the subject. For this book, I will use the definition I always give to my class, or at presentations to non-geographic audiences: A GIS is a tool to work with, analyze, and visualize spatial data. As its input data are spatial, the map is the central part of the system. On this map a variety of spatial layers can be turned on or off according to the needs of the investigation. It is rather like a multilayered cake. One can take a huge mouthful, which may result in the individual tastes of the layers being missed, and even some crumbs spilling out and being lost. Or, the cake layers can be separated and eaten one at a time, or in smaller combinations to satisfy the taste buds.

If this seems like a rather flippant analogy, let's use real examples. Figure 1 shows an online GIS available through the South Carolina Community Assessment Network (Department of Health and Environmental Control), otherwise known as SCAN. This GIS allows web visitors to select from a variety of databases to make charts, tables, and maps. Mappable data include the results of the South Carolina PRAMS (Pregnancy Risk Assessment Monitoring System) study (PRAMS will be described later in this chapter), pregnancy (live births, abortions, fetal deaths), mothers health and lifestyle, infant and child health, and infant mortality. If one of these data layers is selected, for example infant mortality, multiple GIS options become available as easy to navigate drop-down menus. Defaults are already checked on these menus for the easiest of map generations. The GIS user can select different subcategories of the outcome to be mapped — for example, race, ethnicity, age, birthweight, gestation, and ICD (international classification disease) codes. Figure 1 displays a snapshot of one of these maps. Any interested party could make maps for the years 1989-2002, save the images, and compare the outputs both spatially and temporally. Similarly, data can be outputted to tables (in Excel) for further analysis. An option is also available for more sophisticated analysis (such as chi-square, linear regression, and other forms of GIS analysis) through personal contact with the Division of Biostatistics and Health GIS.

A series of GIS operations are available on the bottom right of both maps. Notice that more options are available for the zoomed-in map. This is because the larger the scale of the map (the zoomed-in map is a larger scale than the state-wide map) the more spatial layers can be displayed. The user has the option to turn these spatial layers on or off as he or she desires. This is because more space is available on the larger scale map. If we tried to show all roads on a state level map the surface would be covered in a blackness of lines. We therefore make

Figure 1. Snapshots from South Carolina



"selections" of what we display based on scale. We also make the selection based on its appropriateness. If we have a limited map space, we want to display only those map elements that enhance the final map and aid our decisions.

For another example of how GIS can simplify the spatial understanding of a prenatal risk surface, let's consider an investigation into neighborhood-level hot spots of high infant mortality. In this investigation, in order to identify hot spots, the locations of all births (for the denominator) and all deaths (for the numerator) are needed. These data could be located in a GIS on a city map by heads-up digitizing, which simply means the analyst "clicks" a point on the map where the mother resides, collecting coordinates using a Global Positioning System (GPS), or more commonly by address matching, alternatively called geocoding. Again, this process of data entry will be described in more detail later in this chapter, but at this point it is enough to know that the residential address is placed proportionally on a street network containing address ranges (if the road segment goes from 1 to 99, and the address is 49, it will be approximately halfway along the road). At this stage, two spatial layers exist in our GIS: birth and deaths. Third and fourth layers are also likely to be present in terms of a road network and some sort of political or postal boundary, such as zip codes. It is now possible to perform one of a variety of analyses available in or associated with the GIS to identify hot spots of infant death. However, there are many other spatial layers that the user could add to the GIS, some of which might add more insight, and others which may just confuse the analysis. Examples of these other layers could include other political boundaries (such as county or state), postal boundaries (such as zip code), building locations (such as clinics), landfills, locations of crimes, locations of condemned housing, schools, job centers, etc. All of these spatial layers may have some relevance to the infant mortality hot spots; for example, the lack of access to clinics, the lack of access to education, the stress of living in a high crime neighborhood, and so forth. However, unless the right question is asked, and only the relevant layers are presented *for that question*, the mouth will be stuffed too full and important crumbs will spill out.

But how do we get to the point of having these multiple spatial layer options? As previously mentioned, when I teach any introduction to GIS class, I usually present a framework comprising four component parts of the GIS: data input, data manipulation, data analysis, and data visualization. This chapter will focus on the issue of data, its usefulness, its different types, and how to get it into the GIS.

Data Input

Two general types of data are needed for a community health group: health data and all the rest.

Health Data

For a GIS to function, one needs spatial data. This may seem like a trite comment, but obtaining good quality spatial data continues to be the bane of many spatial scientists. An event needs to have something that ties it to the spatial frame, whether that is the earth, a state, or a building complex. Examples of these spatial references include a coordinate (such as latitude and longitude), an address, a spatial name, or ID (for example *Louisiana* or Zip *70808*). It is not enough to have unique event identifiers (such as a person's social security number) unless one of these spatial references is included. Therefore, for data to be included within a GIS, an initial appreciation for space must occur at the information collection stage.

Most vital records and human surveillance data come with a spatial reference, not because of a more evolved appreciation of the impact of geography, but because of the societal importance of an address. Patients need to be contacted, and therefore the address, along with another personal identifier, is often the most common way of entering a patient's information into a database. It is for these reasons that health maps are often presented by zip codes, even though more socioeconomic information exists at the census tract/census block portioning of space. Data displayed by zip codes would usually be presented as a graduated color map of aggregations (how many people have condition X in zip code Y). If an actual address is included in the patient's information, then it is possible to display a point surface where each location can be queried for any of the patients' data fields, such as "show all women who received no prenatal care." A simple GIS operation, known as *point-in polygon*, can also extract other socioeconomic information from underlying political layers to each single address. In other words, the typical socioeconomic information of that census tract is attached to the person. The most common way of creating this point layer of data is by address matching, which links the vital records data to a network of roads (usually derived from census Topologically Integrated Geographic Encoding and Referencing [TIGER] files).

The most common and useful type of data for a community health unit will be vital statistics data, which for pregnancy outcome investigations usually means birth and death certificate data. In the event of an infant death, a death certificate is generated, which includes a residence. This address can be used to attach the certificate data into the GIS. All other data contained on the certificate (such as date of death, cause of death, etc.) will become "attributes" associated with the point (residence of death) in the GIS. Similarly, the birth residence is also extracted from the birth certificate. In the case of Louisiana, this record contains "attributes" such as place of delivery, mother's and father's place of birth, mother's age, marital status, total number of prenatal visits, education levels of the parents, the month prenatal visits began, number of mother's live births, number of live births that died, number of terminations, APGAR (Activity, Pulse, Grimace, Appearance, Respiration) score, medical risk factors, tobacco and alcohol use, weight gain, complications of labor and delivery, method of delivery, abnormal conditions of newborns, and congenital anomalies.

Of course it should always be remembered that the fields on birth and death certificates will always contain errors due to human fallibility, and sometimes blatant lying, two common examples being self-recorded substance use and highest educational level achieved (Clark, Fu, & Burnett, 1997; Gayle, Yip, Frank, Nieburg, & Binkin, 1988; Green, Nelson, Gaylor, & Holson, 1979; Kharrazi et al., 1999; McDermott, Drews, Green, & Berg, 1997; Reichman & Hade, 2001; Snell et al., 1992; Sorlie & Johnson, 1996; Walker, Schmunk, & Summers, 2004).

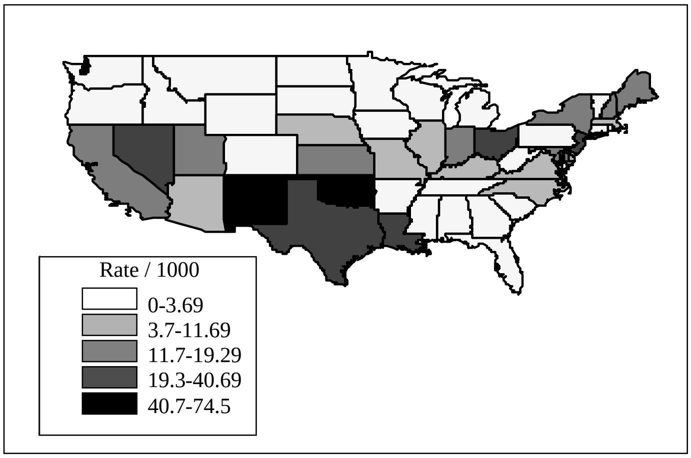
More insight into infant mortality patterns can be gained when the death certificate is "linked" to the birth certificate in order to construct pregnancy histories. This is imperative, as the majority of mortality outcomes are related to factors that are recorded on the birth certificate but not on the death certificate (such as the mother smoking, birth weight, weeks of gestation, etc.). Without such linking, the identification of a spatial infant mortality cluster has limited value, as conditions of the pregnancy are not known. If the data linking is performed by the office controlling vital statistics, the common field to "link" the certificates by would be a unique identifier such as a social security number. If these data are supplied unlinked (which usually means the unique identifier is missing for confidentiality reasons), then it is possible to query records based on address, mother's birth date, and previous birth information. However, as the mother's records often contain different residences on birth and death certificates (this will again be discussed in Chapter VII), linking is impossible. Simply layering a death and birth layer on top of each other in the GIS and joining the two is not accurate enough.

Since 1995, the Centers for Disease Control and Prevention (CDC) has been using a weighting mechanism to account for infant deaths that cannot be linked to birth certificates. This weight =

(number linked infant deaths + number unlinked infant deaths) / number linked deaths

For states where all infant deaths can be linked, the weight equals 1. The CDC recommends that the weighted infant mortality number be used as a more accurate total and characteristic of infant deaths. Again according to the CDC, in 1997 2.1% of all infant deaths could not be linked. Interestingly, there is even a geography to this linking. Figure 2 shows the rate of infant death certificates (for 2000) that could not be linked by each state. Why does the Southwest perform so poorly?

Figure 2. Infant mortality deaths unlinked to birth certificates (2000)



Confidentiality Issues

Data access is usually governed by bodies concerned with the larger issue of patient confidentiality. Unfortunately (and at the same time understandably), sometimes even the address is withheld from health data, making it impossible to create fine (point level) spatial resolution analyses and maps. The legal implications, along with GIS approaches to mask data, will be presented in more detail in Chapter VIII. One aspect of data control which is not understandable is when petty politics control data flow, either through a perceived need to control negative impressions of a health unit (a common complaint levied at hospitals not releasing data for Fetal and Infant Mortality Review [FIMR] panels), or because a body of doctors or researchers intends to maintain control of a data set for their personal gain. More trust is needed between public health offices and interested researchers, whether they be academic or from the local health unit. Anecdotal reports suggest that overstretched state-level public health units do not believe that the local office has the ability to analyze data, even though the data originated with them. In one instance, the local health office had to "purchase" its own data back. While the state head office drowns in data and tasks, the local office, with people who have a far better understanding of the local health concerns (such as the neighborhood characteristics of areas with high infant mortality), become more frustrated. Safeguards still need to be in place, but data release needs to be streamlined and expedited.

One way to circumnavigate the issue of data access, and gain a better understanding of local health issues, is for the local health office to maintain control of its data. This means constructing its own GIS to store and explore the data. This is no pipe dream, as soon most health units will have to collect data electronically in a GIS-friendly format (remember the Healthy People 2010 objective discussed in Chapter I). The Baton Rouge Healthy Start GIS provides an example of such a system. Currently each caseworker uploads, via a laptop, information about his or her program participants to a specially constructed database. These data are entered through hierarchically stacked windows containing drop-down menus and entry lines. In this instance, hierarchical stacking means if the caseworker checks a box on one window, further options, or even additional data windows, are activated. Chapter IX will describe the construction of this GIS in detail, while Chapter IV will present example data entry windows designed to collect prenatal risk information.

Soon these data inputs will occur through a secure Web site entry. Currently, the Louisiana Department of Health and Hospitals (DHH) is transferring children's immunization records to an Internet server so that any mother or heath care worker can access the data when needed. Similarly, in New Jersey an electronic birth certificate program has been implemented. Eventually data in the Baton Rouge Healthy Start will be entered through PDA hand-held units (such as the Dell Axiom series), and immediately uploaded through a secure connection to a Web page. Once the system is operating in (near) real time, other opportunities will be presented. For example, Chapter X will discuss how the GIS could become part of a larger bioterrorism syndromic surveillance system, which could mean community health groups tapping into federal Homeland Security dollars.

A more immediate benefit from creating such a (near) real time GIS is that the opportunity for nontypical data fields can be created. In this way, the analysis is not restricted to available vital statistics data, which usually means attributes found on a long-form birth certificate. Questions relevant to the program area can be asked, such as program participants' perception about racism in clinics. The Baton Rouge Healthy Start also allows for a more developed spatial understanding of those participating in the program. For example, what is an important address? Should all focus be placed on the mother's residency, where she works, her recreation spaces, or her travel routes? Chapter VII will consider the impact the mobility of the mother plays on GIS analysis. For example, if indigent populations frequently move, even during the period of the pregnancy, what is the important address to be included in the event of an infant death? Presuming this death is not from an accident, is the important residence the one at the time of death, the time of birth, or even at a particular stage of the pregnancy?

Address Matching/Geocoding

For point level data, whether accessed from vital records or generated within the health unit, a means is needed to input the data into the GIS. This process, called geocoding, assigns geographic coordinates (for example latitude and longitude or Universal Transverse Mercator [UTM] coordinates) to an address by matching the address number to an "address range" in a digital map (called a street reference map) and, by interpolating, estimates where the address is located between two coordinates that define the limits of the address (McElroy, Remington, Trentham-Dietz, Robert, & PA., 2003).

The most popular street reference file is the TIGER street map created and updated by the U.S. Census Bureau. TIGER street maps can be downloaded for free from the U.S. Census Web site: www.census.gov/geo/www/tiger/index.html. These data are provided in a variety of formats compatible with most GIS software packages. Commercial vendors also offer variations on street reference files, but these are relatively expensive. For example, Geographic Data Technology offers the Dynamap® Transportation road network. GDT-Dynamap® data provides a full range of addresses which are appropriately segmented for geocoding. They also include the impedance values (speed limit, direction, and time) necessary for network path analysis. For normal geocoding, TIGER files are sufficient; however, if a health organization wanted to use GIS in routing emergency services, or locating health facilities based on travel time, Dynamap® data would be required.

A critical component of geocoding is the number of addresses that are successfully matched to the street reference file. A perfect match (hit) rate would be 100%, meaning that all addresses are matched. In practical terms, match rates range from as low as 20% to as high as 100%, depending on factors such as the number of problematic addresses, quality of addresses, and type of street reference map used (McElroyet al., 2003). Typically, higher match rates are found with urban addresses because some rural routes and all post office boxes cannot be assigned appropriate (or accurately identifying) geographic coordinates. The challenge of tracing street addresses for post office box addresses are discussed in Hurley, Saunders, Nivas, Hertz, and P. (2003). Addresses that fail to match can be assigned to the centroid (center of gravity) of the addresses' zip code area (presuming one is supplied). Alternatively, unmatched addresses can be randomly distributed within their corresponding zip code areas, aggregated to the town center, or eliminated from the analysis. Ratcliffe (2004) estimates an "acceptable" minimum geocoding hit rate for crime data (and other point pattern datasets) to be 85%. This means that to generate a statistically reliable map, 85% of the points in a crime table must be geocoded. Even if an address has been successfully geocoded, there can still be a difference between the actual location of the address and the location of the geocoded address. This difference is referred to as the positional accuracy of the geocoded address and depends on the quality of the street reference map and the type of interpolation technique used. The positional accuracy of geocoded addresses is estimated to be within about 300 feet of true address locations (Rushton & Lolonis, 1996). A similar positional accuracy is reported by Bonner, et al. (2003) in a study comparing the location of addresses measured by Global Positioning System receivers (true address location) to positions geocoded with a commercially available street reference map (GDT-Dynamap®).

McElroy et al. (2003) used a five-step procedure to geocode addresses in a large, population-based health study. First, original addresses are standardized to the U.S. Postal Service format. This can be performed in-house through commercially available software, or sent to the Tennessee branch of the U.S. Postal Service for processing. This is necessary because geocoding tools in most GIS software require addresses in a standardized format. Second, standardized addresses are geocoded to two different versions (2000 and 1995) of the TIGER street map, using an 80% spelling score and 80% overall sensitivity score; these scores allow matches for addresses with minor deviations in spelling and format. Higher percentages would result in lower numbers of geocoded addresses and lower probabilities of placement errors; lower percentages would yield the opposite results. About three-quarters of all addresses from the health study are successfully geocoded after the first two steps. Third, individuals' telephone numbers or names from unmatched addresses, post office boxes, rural route and incomplete (partial or garbled) addresses are submitted to an Internet mapping engine (e.g., Anywho.com) to identify updated street addresses. Updated and edited addresses are subsequently entered into an Internet mapping engine (e.g., TeleAtlas.com, MelissaDATA.com, and the U.S. Census Bureau's Gazetteer) to find the respective geographic coordinates. For addresses that remained unmatched, participants' households are contacted by telephone in step four. If new street addresses do not match, then geographic coordinates are assigned to the closest street intersection. For all addresses that are still unmatched, the geographic coordinates of the zip code centroid (center of gravity) are used in step five. By the end, only 3% of all addresses from the health study are not assigned geographic coordinates.

Two geocoding examples for Baton Rouge include the original birth and death certificate data, split by month, used for the initial Baton Rouge Healthy Start project proposal (Table 1), and a more recent longitudinal investigation of (very) low birth weight and infant mortality in East Baton Rouge Parish, for 1990-2001 (Arthold, 2004). In this second study, an overall match rate of more than 98% was achieved using a priori and manually standardized addresses, the TIGER street map and local maps of the parish. In total, more than 66,000 addresses from birth records and 196 death record addresses were matched. The unmatched records (less than 2%) were omitted from the analysis.

Table 1. Geocoding metadata

1996	# address	# match	# PO Box	# unmatch	Tot unmatch
Jan	424	418	4	2	6
Feb	400	393	2	5	7
Mar	436	427	8	1	9
Apr	432	425	5	2	7
May	428	415	12	1	13
Jun	430	423	5	2	7
Jul	471	466	2	3	5
Aug	505	500	4	1	5
Sep	497	487	5	5	10
Oct	481	471	9	1	10
Nov	417	413	3	1	4
Dec	509	501	7	1	8
Total	6142	6016	94	32	126

Other Useful Data 1: Socioeconomic Data

U.S. Census Bureau

Although it is important to include health data in the GIS, preferably at the address level, there are several other spatial layers that are important for analysis and visualization. In the previous section, TIGER data were introduced as part of the geocoding process. This is just one of the datasets available through the U.S. Census Bureau. Without this data layer it would be extremely hard to place health events in the GIS. This is only one example, though, of useful census data. Although the Census Bureau conducts many important censuses and surveys (approximately 100 every year), the most well-known is the official decennial population census of the United States, collected in every 00 year. During each decennial census, the Census Bureau collects data from every household in the United States and its territories. By law, no one is permitted to reveal information from these censuses and surveys that could identify any person, household, or business. The information collected for each census and survey is summarized by geographic area and then published in a variety of formats, including printed reports, CD-ROM, DVD, and on the Internet (www.census.gov/).

Data from the decennial census are collected from all people and housing units (100% data) or from a 1-in-6 sample and weighted to represent the total population (sample data). The 100% data are the so-called Summary Files (SF) numbers 1 and 2, which include counts and information on age, sex, race, Hispanic/Latino origin, household relationship, whether residence is owned or rented, and so forth. The sample data, or Summary Files 3 and 4, include detailed population and housing data, such as place of birth, education, employment status, income, value of housing unit, year structure built, and so forth. The

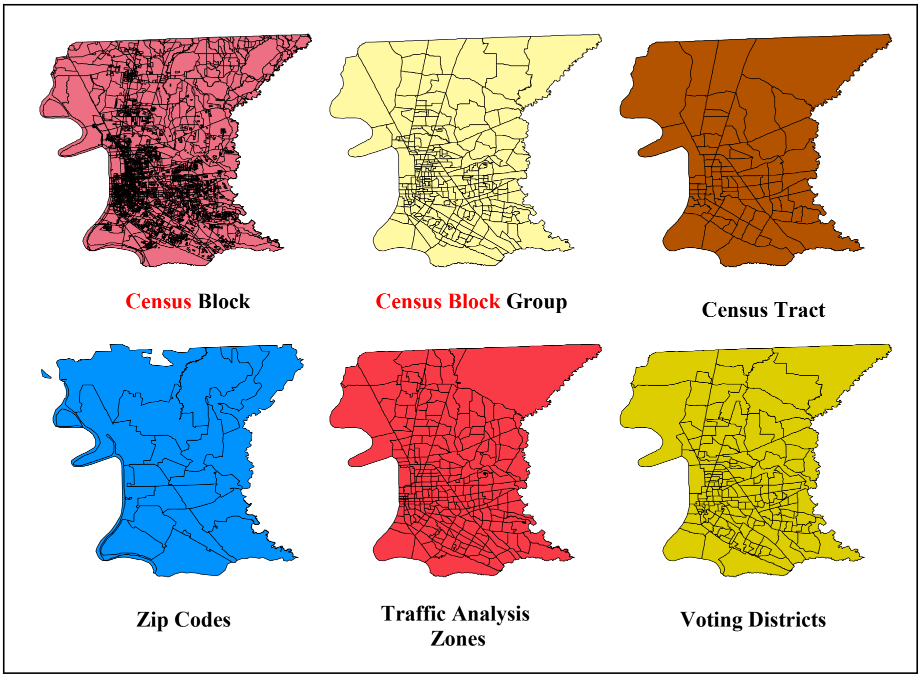
Census Bureau reports data for a wide variety of geographic areas, ranging from the entire United States down to a census block. Other enumeration units in between these two geographic areas are (from large to small): regions, divisions, states, counties/parishes, census tracts, and block groups. Additional geographic types include zip codes, school districts, traffic analysis zones, or metropolitan areas. These geographic types (boundary files) are important for GIS analysis because they serve as spatial references for the attribute data, collected in the censuses and surveys. They are available in digital format and as such can readily be used in a GIS. This makes it very easy to visualize any attribute data from the Census Bureau in the form of choropleth maps.

The importance of census data is that far more information can be gathered about a program participant's neighborhood, and even, after a little manipulation, about the program participant herself. Even with extensive data sets generated by programs such as the Baton Rouge Healthy Start, there will still be data holes that the decennial census can fill. These data can provide visual backgrounds, for example a choropleth map displaying areas of the city with old housing stock (which might reveal neighborhoods requiring lead exposure screening), to providing additional data fields in regression-based forms of analysis (as will be discussed in Chapter VI).

Other Useful Data 2: Boundary and Background Data

In the previous section, different geographic areas were mentioned as being available through the census. These areas, input into the GIS as boundary layers, are important for two reasons. First, they give the observer a frame of reference. If spatial data, such as the residences of infant mortality locations, are to be located on a map, the observer will need to be familiar with a geographic layer. Most people can identify their home state, and to a lesser degree their county shape. Unfortunately, at the city level traditional geographic areas, such as zip codes or census tracts are not as familiar. The second reason to include geographic boundary areas is as a means to display aggregated data. These data may originate from the decennial census, in which case census tracts and block groups are the most common form of aggregation, or be collected by another agency or facility (such as a hospital) in which case zip codes would be favored. In order to show how these boundaries differ, Figure 3 presents six common geographic areas for East Baton Rouge Parish. Researchers have also used combinations of these aggregations to form new areas, for example Roberts (1997) used community areas in his analysis of low birth weight in Chicago, these being comprised of census tracts. Similarly Rauh, Andrews, and Garfinkle (2001) used New York City health areas comprised of 4 to 6 census tracts, while Pearl, Braveman, and Abrams (2001) used combinations of census blocks.

Figure 3. Six common geographic boundaries



In Britain, an index of social deprivation was applied to census wards (similar to census tracts). By creating clusters of wards reaching a deprivation threshold, new "areas" meeting European Commission (EC) cutoffs for funding eligibility were created. Although these larger areas contained wards that exceeded the threshold, when "smoothed" into the larger cluster mass, poverty levels were such to meet the EC level (Boyle & Alvanides, 2004).

In order to improve the geographic frame of reference, it is often useful to turn to city background maps, which are available in raster format (Figure 3). Most vector GIS packages can also import these spatial data layers. Examples of these types of data include 1:24,000 Scanned Topographic Maps (DRGs), 1:100,000 and 1:250,000 Scanned Topographic Maps, Coastwide 2001 Scanned Images, and Digital Orthophoto Quarter Quadrangles (DOQQs).

Figure 4 displays the same section of Baton Rouge on both a DOQQ and a DRG. Both should be fairly recognizable to residents from these areas. The box indicates the approximate location of the Family Road building which houses the Baton Rouge Healthy Start Program. Women's Hospital, a major medical facility of the region can also be seen in the bottom right of both images.

Possibly the more useful spatial data layer for prenatal health analysis is the DOQQ, which is an infrared aerial photo. Each DOQQ, or tile, is approximately 4 by 4.5 miles. The resolution of the photo is 1 meter, though any zoom being

Figure 4. DOQQ and DRG of Baton Rouge Healthy Start

enlarged to a greater resolution than 20 meters on the ground will lose clarity. Even so, most buildings and roads can be identified from these images. These photos are scanned into the GIS, and "known" coordinate references, such as a building location, a road intersection, or field corner are embedded into the image (this process is referred to as georeferencing). Once the image has been georegistered, it now exists as a spatial layer in the GIS. The same process can be applied to any picture, such as an apartment building complex plan, or a spatially generated map from a non-government source. Once inside the GIS the map can be tied to other spatial data as long as a known coordinate system is used. The DOQQ in Figure 4 is embedded with Universal Transverse Mercator projection coordinates, which are expressed as meters. In order to make these images manageable a compression routine is often applied; the tile in Figure 4 has been translated using a MultiResolution Seamless Image Database, more usually referred to as MrSid.

The ability to pick places, buildings, roads, and general land cover from the DOQQ makes it a suitable source for creating new GIS layers by heads-up digitizing features. By clicking a mouse around a feature on the photo, the coordinates are transferred to the new layer which in effect sits on top as an overlay. For example, in a current project considering recreational activity spaces and indigent health, a DOQQ was used to create a GIS backdrop by digitizing key features such as paths, water fountains, lights, and rest areas in a city park. These maps are used by a second team, who would visit the park and assign quality attributes to these spatial features, such as “is the path broken,” “do the lights work,” and so forth. A similar approach could be used to identify key neighborhood features affecting a child’s health, such as a play area, proximity to busy roads, or the presence of known drug corners.

Internet Data Access

So where can a community health group get these data layers? In general, GIS users can access data in three ways: through national, free, Internet download sites, through local, free, Internet download sites, and through private vendors. One of the best Internet sites is www.geographynetwork.com, which acts as a clearing house for multiple types of data. Of particular interest to community health groups is the ability to download (for free) census socioeconomic data (only Summary File 1), a variety of different GIS-ready boundary layers, geocoding-ready road files, and other geographic features. The user simply specifies either a single data layer for the entire state, or zooms into a single county and picks multiple GIS layers. These are exported as a zip file (see Figure 5).

Alternatives to national data sites are state-level depositories, usually connected to state universities, which might be the only source for local data, such as recent DOQQs. For example, the Atlas Web site, housed in the CADGIS (College of Art and Design and the Department of Geography and Anthropology) lab at Louisiana State University offers a variety of national and local GIS data sets for download. Available data include: National Wetlands inventory, 1:24,000 Scanned Topographic Maps, 1:100,000 & 1:250,000 Scanned Topographic Maps, Coastwide 2001 Scanned Images, DOQQs, 1:24,000 Digital Elevation Models, and LIDAR (LIght Detection And Ranging) data (Figure 1). In addition, census data (2000 SF2Tables and LandView 5) are also available. These different data formats may be unfamiliar to readers, but most introductory GIS texts will provide definitions. The Atlas Web site also provides detailed data explanations and metadata.

Figure 5. Progression of data download windows from the geography network

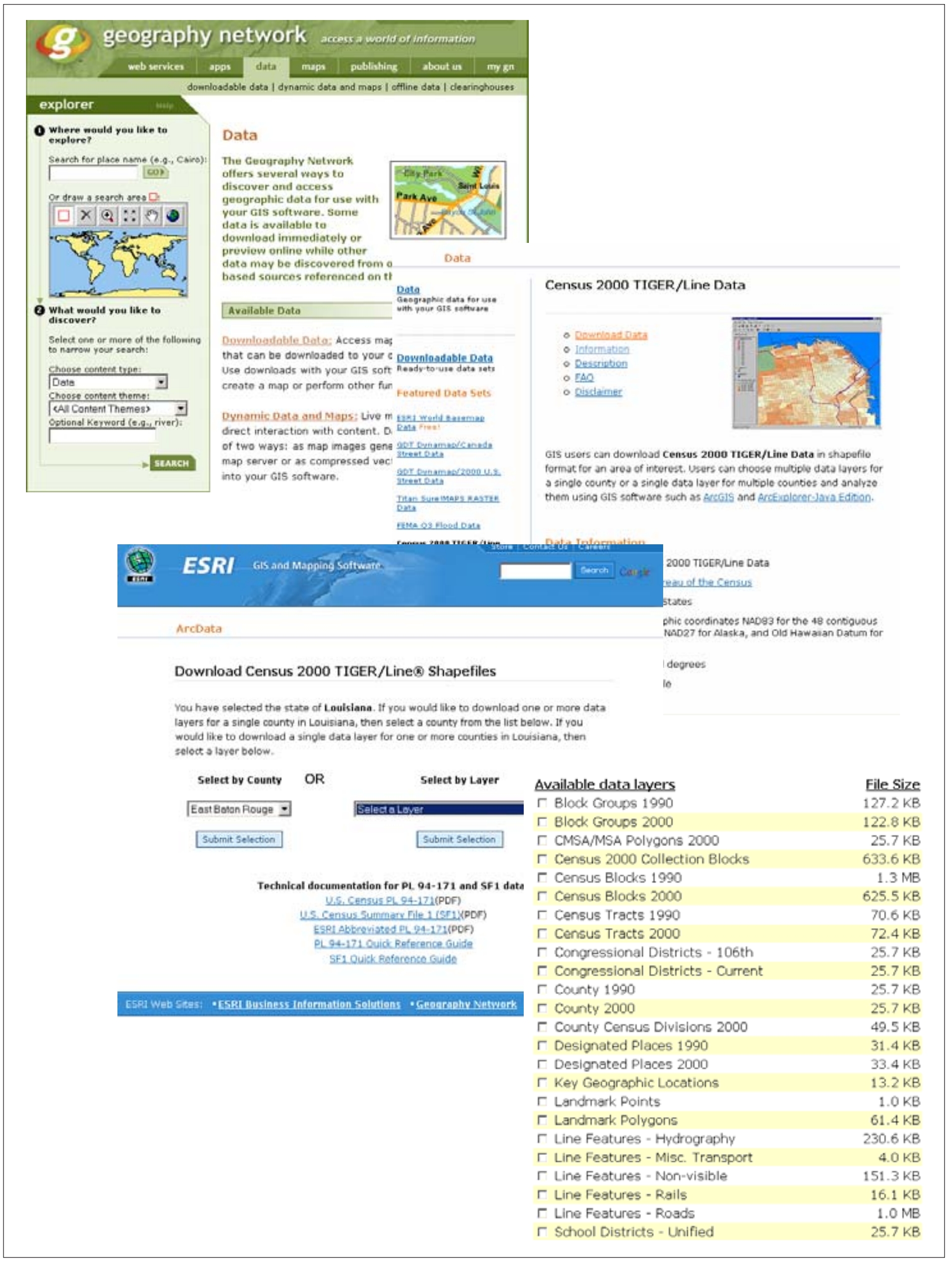
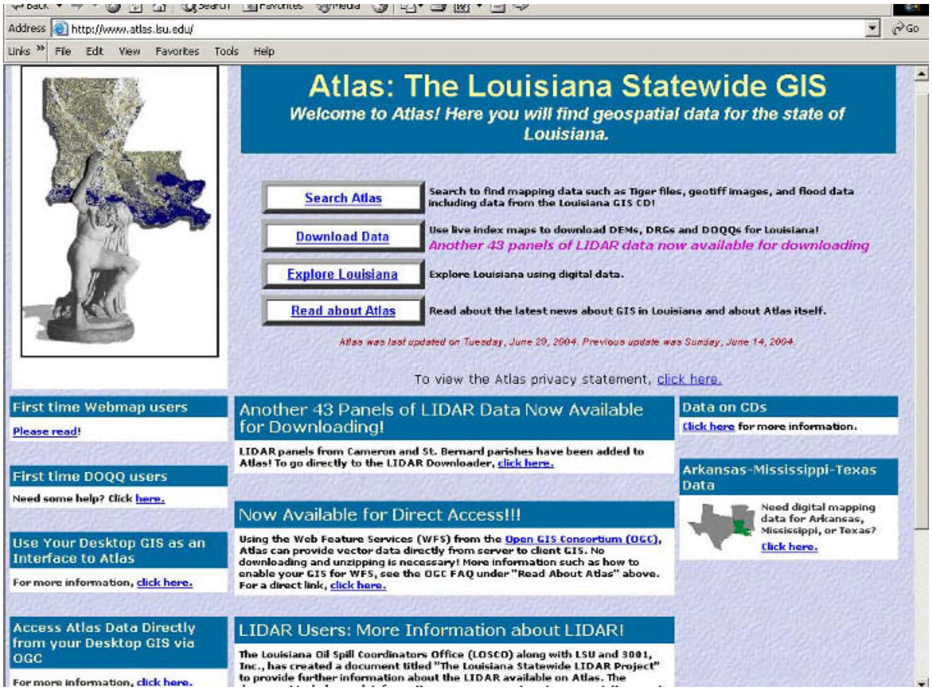


Figure 6. The Atlas Web site



Data Manipulation

The second key component of the GIS is as a tool to manipulate spatial data. There are many forms of spatial (and aspatial) manipulation that can be performed with a GIS, some of which, like extracting data through buffers, could be considered as analysis. One of the more common manipulations is using the GIS to display and analyze original data in alternative spatial units. For example, consider a spatial layer of addresses extracted from birth certificates. This GIS layer can be viewed as an aggregated layer which is manipulated in three general ways: by aggregating into spatial units, by reducing the data through spatial and aspatial queries, and combining attribute data to create new data.

Aggregating into Spatial Units

Point data (such as a birth residence) is easily aggregated into different geographic areas, such as a zip code or census tract. If we return to Figure 3, a point layer could be aggregated into any of these maps, and either displayed as a choropleth map or used as input units in an analysis. This is achieved either by using an existing spatial identifier in the record, for example a zip code in the address, or by performing a spatial join where the coordinate of each address is placed within a polygon (again, if using the zip code example, the zip code in which the address geographically falls). The GIS also allows for this second approach to create non-recognized spatial aggregations. For example, if the community health program wanted to analyze birth outcomes on either side of a political boundary, whether that is at a country (United States-Mexico), state (Louisiana-Texas), county or suburban level, a geographic shape can be created (usually a buffer following the boundary) and joined to the point layer. The smoothed surface of rates (such as infant mortality rates) created by spatial filtering discussed in Chapter VI uses a similar approach by joining points to an overlapping surface of circles.

There are several reasons why aggregating data is important. First, it is the accepted way of displaying (and releasing) health data so as to preserve patient confidentiality. The geographical unit of display (or analysis) depends on the health condition in question. Birth data is fairly common, so a relatively small underlying denominator population is needed, such as a census tract. For rare conditions, such as HIV/AIDS or infant mortality, a coarser geographic aggregation might be required, such as a zip code, or even a county. The second reason for aggregating point data to geographic units is so that an analysis can be performed whereby the dependent variable matches the same geography of available independent data. The most common example is a regression utilizing census data as independent variables, these data only being available in traditional census geographic areas. The third reason to aggregate point level data is to ease visual displays. Although it is possible to visualize points as sole entities (through a dot map), or as interpolated surfaces (usually by contours), the most common display is as a choropleth map. Not only does this preserve confidentiality, but it is also arguably the easiest way to show changing distributions (such as infant mortality) across a city.

Area analysis has become a popular means of investigation in the public health literature, especially with ecological studies and contextual or multilevel modeling using census data in the analysis (Diez-Roux, 2001). Multilevel modeling in particular approaches the analysis of health as a combination of both individual outcomes and neighborhood characteristics (e.g., income or race), these data usually being extracted from the census (for examples see Diez-Roux, 1998a, 1998b, 2000; Diez-Roux, Link, & Northridge, 2000; Duncan, Jones, & Moon, 1996; Matteson, Burr, & Marshall, 1998; Pickett & Pearl, 2001; Rauh, Andrews & Garfinkle, 2001; Subramanian, Acevedo-Garcia, & Osypuk, 2005; Subramanian, Chen, Rehkopf, Waterman, & Krieger, 2005; Subramanian, Kim, & Kawachi, 2002; Subramanian, Lochner, & Kawachi, 2003; Subramanian, Nandy, Kelly, Gordon, & Davey-Smith, 2004).

Unfortunately, there are some serious pitfalls that can befall the spatial aggregation of data (Soobader & LeClere, 1999). These pitfalls can broadly be split into problems with visualization, and variations that occur in analysis (P. A. Longley & Harris, 1999). The next chapter will revisit the issue of how one component of visualizing aggregates, namely the classification scheme chosen (how to break the data into different groups), can change a final map. Variations can also occur due to the units of aggregation. Consider Figure 7, which shows the number of single women who also have to care for young children. These data were originally point data and then aggregated to three different geographic areas: census tract, census block group, and zip code.

The three maps have been classified by standard deviations, with those areas exceeding three standard deviations being identified as areas of risk. If, for example, in a disaster, this particular risk population were to be identified for prioritized response, a map such as this could be used to target relief. However, depending on the geographic units chosen, these hot spots of vulnerability vary quite considerably.

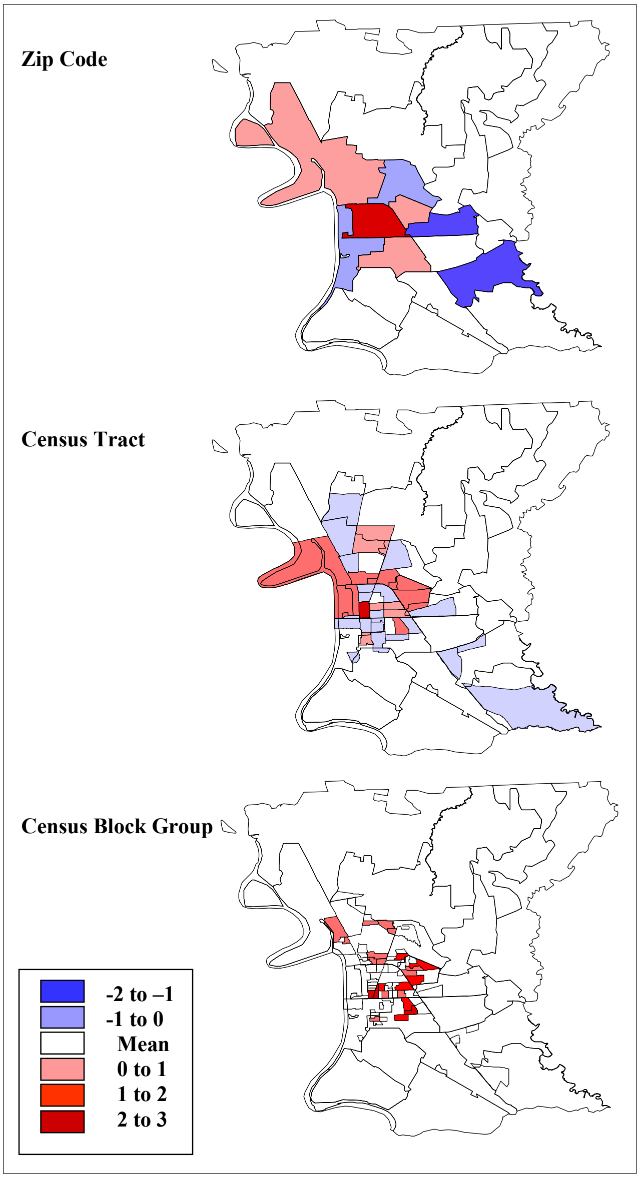
Another pitfall arrives when aggregated data is used in an analysis, or as part of a GIS manipulation known as point-in-polygon. The problem with aggregation and analysis is a variant of the ecological fallacy, which, simply put, states that aggregated data should not be disaggregated (Pattenden, Dolk, & Vrijheid, 1999). The spatial variant of this problem is known as the modifiable area unit problem (MAUP). Research has shown that if data is aggregated up to different configurations of space, different results will occur in analyses based on the scale (how many aggregations of space occur) and the zone (the shape of those aggregations). If a result is found — for example, a relationship between single women with children at home and successful prenatal visit completions — the strength of the relationship, and even the direction of the relationship, might vary on the scale and zone the original data were aggregated to. In this example the scale is the different size of areas (for example, census tract is larger than, and comprises of, census block groups).

If results do vary because of this aggregation effect, a question must be raised as to the validity of the analysis outcome.

A point-in-polygon analysis places a point (e.g., the address on a birth certificate) within a polygon (such as a census tract), and extracts all information from that polygon as additional attributes for the point. For example, the average income in the census tract can become the family income for all births falling inside that census tract. This technique would allow for investigations combining health outcomes and socioeconomic data, allowing for questions such as, "is there a relationship between low birth weight and families in poverty?" being posed.

Remember again that social environments have become important components of health analysis — and quite rightly so, as not every health outcome can be

Figure 7. Three aggregations displaying risk hot spots



attributed to an individual's circumstances (Diez-Roux, 1998a; Schwartz, 1994) — and that multilevel modeling is one of the techniques in favor. Multilevel modeling must employ a point-in-polygon approach to place individuals within their social environment. As an example of this approach, see Ahern, Pickett, Selvin, and Abrams (2003), who ran an analysis on the effect of individual smoking and neighborhood effects on preterm birth among African Americans.

The authors used "place" data taken at the census tract level, including (but not limited to) proportions of African Americans, adults without high school education, the unemployed, and persons in poverty (Ahern, Pickett, Selvin, & Abrams, 2003). In this paper the authors also go on to create an even more easily achieved manipulation within the GIS by calculating a change value between two decennial periods (1980 to 1990) for these variables.

Data Reduction

A common GIS approach is to reduce a dataset by a query, resulting in a selection. This selection can be extracted as a separate GIS layer or simply analyzed on the fly. The purpose of data reduction is to focus a research question on a subgroup of the entire population, such as, "select out all women who received no prenatal care." The query is built as a Boolean expression, which simply means asking a question of the database using algebraic language — for example, "prenatal = 0." This expression is usually constructed by double-clicking the appropriate columns and modifiers in a query builder. This query would search the attribute column "number of prenatal visits during pregnancy" for any "0." Only those records (births) that meet this query are selected out, and if necessary, saved as a new GIS layer which can be mapped showing the spatial distribution of women receiving no prenatal care. The data set has been "reduced" to meet this query. Statistical and spatial analyses can also be performed on the subset, such as finding the age distribution of women receiving no prenatal care.

The second type of query is spatial, whereby the question asked is one of geography. For example, if it is reported that racism exists at one clinic, all births within 1 mile can be queried so that a questionnaire will be distributed to those mothers about their delivery experience. Another approach could be to identify all births within 1 mile of three different clinics offering the same services (such as a fatherhood program) so that birth outcomes could be compared as an evaluation strategy regarding the success of each program. Similarly, environmental insults could be investigated; for example, births within 500 meters of a landfill, or within 50 meters of a busy road, the birth outcomes being extracted and compared to births in equal-sized neighborhoods of similar socioeconomic and racial makeup. For all these examples, the subsets can be saved as separate layers in the GIS, allowing for further analysis. None of the above examples should be taken as a standard reference for appropriate distances. This is one of the important considerations when performing a spatial query: What is an appropriate distance? The one piece of advice that can be offered is to let a hypothesis drive the distance used (for example likely, *acceptable* walking distance). If no such hypothesis can be ventured, then multiple distances should be chosen to see if different spatial patterns are revealed. A similar problem that will be addressed in the next chapter is, "what is the definition of a neighborhood?"

Both aspatial and spatial queries can combine multiple questions. For example, if we suspect that a landfill is causing negative birth outcomes, such as a high proportion of low-birth-weight deliveries, the initial spatial query could be, "select all births within 0.25 miles of the landfill." In order to compare the proportion of low-birth-weight deliveries in this "buffer" to the rest of the city, a difference of proportions t-test could be used (this will be described in more detail in Chapter V). However, if we want to control for the effect of poverty and race, we could create a further subset within the buffer where only white households making an income exceeding $60,000 were selected. The argument would be that births to these families should be similar both inside the buffer and for the rest of the city. However, if the sample inside the buffer had a statistically significantly higher proportion of low-birth-weight deliveries than for the rest of the city, it could be as a result of being proximate to the landfill. The rationale for using the sample described is that fewer confounding risks would be impacting middle class mothers as compared to an indigent cohort.

Creating New Data

The final manipulation is to use the GIS to create new data by combining or interacting with existing data. The most common example of this kind is the calculation of a rate. For example, aggregated birth and death data are combined (death divided by birth), and usually multiplied by 1,000, to give the infant mortality rate. Any other data column can be combined in the same way, using either the total population (for example, all births as the denominator), or a subset (all African American births as the denominator).

Calculating Deprivation Indexes

Many of the studies that use socioeconomic information in combination with individual outcome data investigate the degree of deprivation in the surrounding neighborhood, which could either have a negative influence in its own right, or act synergistically with other individual risks, such as smoking (Bancroft, Wiltshire, Parry, & Amos, 2003; Berman et al., 2003; Stead, MacAskill, MacKintosh, Reece, & Eadie, 2001). One argument is that an area-based deprivation measure is a better expression of the experience than individual measures (Pattenden et al., 1999). In some of these studies, hardship data extracted from the census is used unchanged (an example being unemployment and degree of poverty), or at least with a simple manipulation, such as turning the raw number into a proportion per geographic area. Some studies, however, use or create hardship indexes that attempt to combine multiple influences in order to gain a more holistic impression of the neighborhood. For example, Roberts (1997) uses an index for socioeconomic status (comprised of an average of z scores for percentage of white-collar workers, median family income, and median adult education level), and economic hardship (comprised of an average of z scores of percentage of unemployed adults, percentage of families in poverty, percentage of young residents under 18, and percentage elderly over 64 years). In another investigation into the relationship between lead and crime, Stretesky and Lynch (2004) use a "resource deprivation" index built on a principal components analysis, including percentage African American, natural log of median family income, a Gini index of family income inequality, percentage of families below poverty, and percentage of families headed by females. The Breadline Britain poverty index extrapolated survey responses to the larger population, with socioeconomic measures being paired from both sources. A logit regression estimated "poor houses" based on percentages of house owner status, car possession, single parents, labor skill, unemployment, and debilitating illness (Gordon & Forrest, 1995; P. R. Lee, Moss, & Krieger, 1995). Child-specific variables designed to capture needs assessment included children with disabilities, at family risk, in welfare families, who were low birth weight, had been expelled, and had been in trouble with the law. These data collected from different agencies were georeferenced and aggregated to enumeration districts. These indirect measures were broken into five categories and ranked from highest to lowest. The combination of ranks were then mapped to identify areas of risk (Craglia & Signoretta, 2004), a simple approach which has also been applied in Baton Rouge, though by using risk data extracted from the birth certificate and mapped at the census block group level.

Other poverty measures include the Carstairs Deprivation Index (Carstairs, 1995; Carstairs & Morris, 1990; Morris & Carstairs, 1991) which has performed well in health analyses; the UK Index of Multiple Deprivation and the Index of Local Deprivation (Department of Environment Transportation and the Regions, 2000); and the Townsend Index (Townsend, Phillimore, & Beattie, 1988). A review of different deprivation indexes can be found in Morris and Carstairs (1991).

Of course, these indexes are assigned to some form of geographic unit (Gordon, 1995; Gordon & Forrest, 1995; Harris & Longley, 2002; P. Lee, 1999) and occur spatially, so a GIS is an important tool in their creation (Boyle & Alvanides, 2004; Craglia & Signoretta, 2004; Harris & Longley, 2004; Harris & Frost, in press). However, as these indexes are assigned to either common political aggregations or to clustered masses, we are again faced with the issue of the modifiable area unit problem of the geographic scale at which deprivation exists (Harris & Longley, 2002, 2004). In a comparison study of deprivation measures, P. R. Lee et al. (1995) found that common findings were made at the more aggregate level, though more variation occurred when moving to the smaller scale. Similarly, Harris and Longley (2004) commenting on the Breadline Britain poverty index said that it is an incorrect assumption to believe that spatial aggregation has no impact on model parameter estimates. Spatial variation does occur between different aggregations, an important point to remember before embarking on any form of spatial analysis.

Another form of data creation includes the previously described point-in-polygon analysis, as data from the aggregated geographic unit is added to each point. In this way, socioeconomic information such as median income, employment type, and female as head of household could be included in an analysis of low-birth-weight patterns. Of course, it should be remembered that every point (birth) falling inside the same census tract would receive the same attributes.

Improving Health Outcome Information

One form of data manipulation used in pregnancy outcome investigations is the attempt to discern whether a mother has received adequate prenatal care. Although it is acknowledged that seven or fewer prenatal visits, and prenatal visits beginning in the second trimester or later, should be considered as risks leading to a potential negative birth outcome, more informed measures identify how many visits are needed given the starting point of prenatal care. As an example of acknowledged inadequacy of care, the Kessner Index compares gestational age to expected number of prenatal visits. Inadequate care in this index includes 14 to 21 weeks of gestation where no visit has been made, 22 to 29 weeks where one or no visit has been made, 30 to 31 weeks of gestation where two or fewer visits had been made, 32 to 33 weeks of gestation where three or fewer visits had been made, and with 34 or more weeks of gestation where four or fewer visits had been made. Also, any woman beginning care in her third trimester is also considered as receiving inadequate care under this index.

Of course, indexes such as these do not take into account the quality of the visit (something that can only be judged by extensive surveying), and quite often a degree of individual recall accuracy is required. An alternative approach, and one that can be constructed in the GIS by manipulating two common fields on the birth record, is the Kotelchuck Index, which combines when prenatal care began with the total number of prenatal visits. A ratio is calculated whereby the actual number of visits is compared to the expected number, as determined by the American College of Obstetricians and Gynecologists for different initiation periods of care. Inadequate prenatal care is determined as having a ratio less than 50%.

Perinatal Periods of Risk (PPOR)

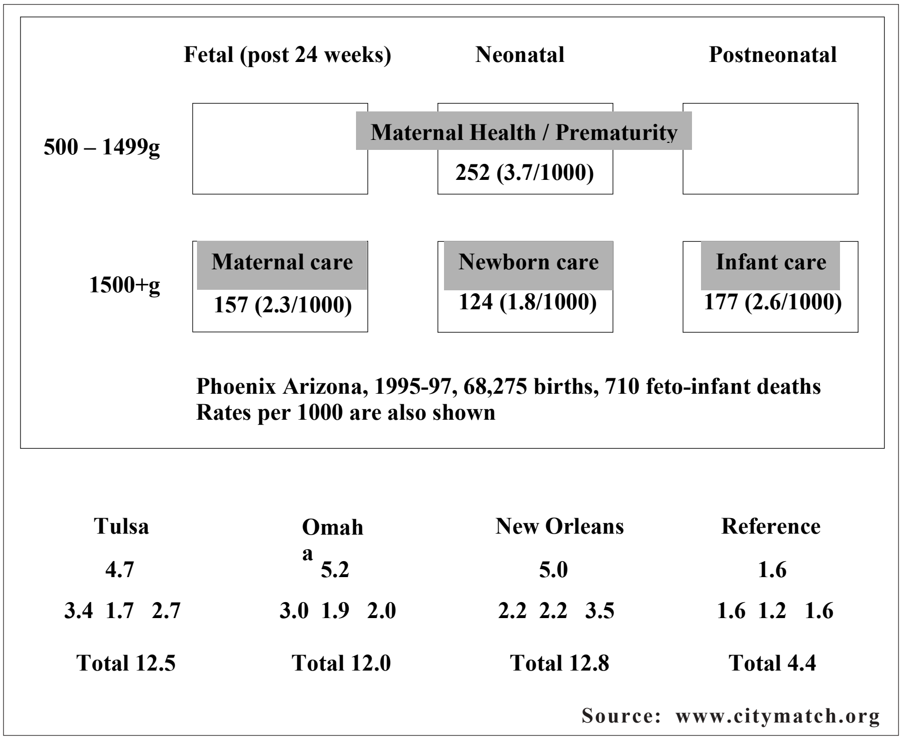
Originally designed by the World Health Organization for developing world countries, perinatal periods of risk (PPOR) was adopted by the CDC and CityMatCH as a tool for understanding fetal and infant mortality within U.S. urban environments (Perinatal Periods of Risk Work Group). A risk assessment tool, PPOR is designed to focus a partnership of community leaders, health care workers and program participants on specific components of causes leading to fetal and infant demise. An important component of this action plan is the mapping of fetal-infant mortality — not a spatial mapping, but a manipulation of outcomes based on age of death and birth weight. Instead of considering all infant deaths, deaths at different ages within the first year of life will generate different response solutions. Similarly, different birth weights will require different interventions. Therefore, the PPOR mapping attempts to determine local vulnerability by finding elevated levels within the map.

The categories are explained in the following way: 500- to 1,499-gram births are considered very low birth weight, births greater than 1,500 grams are at least moderate to low-birth-weight deliveries. Neonatal deaths are those that occur within the first month of life, and postneonatal deaths occur after the first month, but before the end of the first year of life. Two cutoffs are used to limit seriously impacted births and deaths with little chance for intervention, these being births weighing less than 500 grams, and fetal deaths occurring before 24 weeks. Also excluded are spontaneous and induced abortions (Perinatal Periods of Risk Work Group).

Figure 8 shows a modified version of the PPOR map available on the CityMatCH Web site. Six total categories are created from a combination of two birth weight thresholds (500 to 1,499 grams, and above 1,500 grams), and for three time periods of feto-infant death (beyond 24 weeks, first month of life, and before the end of the first year of life — all being mutually exclusive). These categories are further grouped into four, with the upper row becoming Maternal Health and Prematurity Group, while the bottom row includes Maternal Care, Newborn Care, and Infant Health. Data is also reported for Phoenix, Arizona, where out of 68,275 births, 710 feto-infant deaths occurred. The distribution of these deaths by category is also displayed.

Rates for each group can be calculated by taking the feto-infant deaths in each category, divided by all births plus feto-infant mortality deaths. These rates can be calculated for subcohorts, for example by race, socioeconomic status, or even by geographic area. These rates can be used to see how one group changes over time, or to judge differences between geographic areas. For example, Figure 8 also displays the PPOR for three other U.S. cities, Tulsa, Omaha, and New Orleans, to show how the distribution of rates can change. In all cases, Maternal

Figure 8. Perinatal periods of risk (PPOR)



and Prematurity records the highest feto-infant mortality rate, but the second highest rate varies between the cities. In New Orleans, with a high proportion of African American births, the largest problem is found in the infant care category. For this city, programs and resources should be targeted in this area (Perinatal Periods of Risk Work Group).

A reference population is used to calculate excess in any category by simply subtracting reference rates from the city rates. This approach identifies which of the categories have the highest differences, which should mean targeted resources could have the greatest impact. Again, in the New Orleans example, the greatest reduction in the excess can be found in targeting resources towards the Maternal Health and Prematurity Group, followed by Infant Health.

It is easy to achieve PPOR calculations within a GIS. As long as birth and death record data are linked, the age of death and the birth weight can easily be extracted. The hardest category to identify would be fetal deaths, as their recording (as fetal death certificates) varies from state to state. If fetal and infant mortality data are stored within the same geographic layer, a simple query could break the data into four categories, determining age at death and initial birth weight, as long as the baby had more than 24 weeks of gestation and weighed more than 500 grams.

One problem is the minimum feto-infant mortality number, which should be at least 60. In other words, no matter what geographic area or subpopulation is being investigated, at least 60 deaths need to be recorded. For the entire parish of East Baton Rouge, the number of infant deaths between 1996 and 1998 was 213. When these data are reduced (easily done in the GIS by a query) by births exceeding 500 grams and occurring after 24 weeks, this number drops to 122. If we now spatially reduce this number by the five zip code region of the Baton Rouge Healthy Start, only 56 deaths occur in the program area. If we extend the program area to include one further zip code, which will be added in Phase Two to the Baton Rouge Healthy Start, the number rises to 69. Even so, the PPOR calculation can be applied to the program study area, and this can be compared through time, but numbers are not sufficient to allow any neighborhood profiling.

Another source of pregnancy outcome information often used in analyses are Pregnancy Risk Assessment Monitoring System (PRAMS) data. These data include risk factors that are acknowledged by the CDC as affecting birth outcomes, including prenatal care, infant health care, substance use (including alcohol and tobacco), nutrition, mother's income, psychosocial support, stress, domestic violence, whether the pregnancy was planned and subsequent feelings, breastfeeding, and folic acid knowledge. Data is collected through surveying a sample of mothers, initially through mailings, but then by follow-up personal contact if necessary. Unfortunately, not all states participate in the PRAMS program, and even in the states that do, outcome data is only available for large geographic areas (usually the state), meaning it is not an appropriate dataset for local communities' GIS use. The questions asked, however, can be used as templates to guide the collection of similar data by local health units. This is in effect what has happened with the data collected for the Baton Rouge Healthy Start GIS.

References

Ahern, J., Pickett, K. E., Selvin, S., & Abrams, B. (2003). Preterm birth among African American and white women: a multilevel analysis of socioeconomic characteristics and cigarette smoking. *Journal of Epidemiology and Community Health, 57*(8), 606-611.

Arthold, M. (2004). *Local spatial variation in (very) low birth weight and other risk factors influencing infant mortality: A study from Baton Rouge, Louisiana*. Unpublished manuscript, Carinthia Tech Institute, Villach, Austria.

Bancroft, A., Wiltshire, S., Parry, O., & Amos, A. (2003). "It's like an addiction first thing...afterwards it's like a habit": Daily smoking behaviour among people living in areas of deprivation. *Social Science and Medicine, 56*(6), 1261-1267.

Berman, B. A., Wong, G. C., Bastani, R., Hoang, T., Jones, C., Goldstein, D. R., et al. (2003). Household smoking behavior and ETS exposure among children with asthma in low-income, minority households. *Addictive Behaviors, 28*(1), 111-128.

Bonner, M. R., Han, D., Nie, J., Rogerson, P., Vena, J. E., & J. L., F. (2003). Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology, 14*(4), 408-412.

Boyle, P., & Alvanides, S. (2004). Assessing deprivation in English inner city areas: Making the case for EC funding for Leeds City. In J. Stillwell, & G. Clarke (Eds.), *Applied GIS and spatial analysis* (pp. 112-136). Chichester, UK: John Wiley & Sons, Ltd.

Carstairs, V. (1995). Deprivation indexes: Their interpretation and use in relation to health [Suppl. 2]. *Journal of Epidemiology and Community Health, 49*, S3-8.

Carstairs, V., & Morris, R. (1990). Deprivation and health in Scotland. *Health Bull (Edinburgh), 48*(4), 162-175.

Clark, K., Fu, C. M., & Burnett, C. (1997). Accuracy of birth certificate data regarding the amount, timing, and adequacy of prenatal care using prenatal clinic medical records as referents. *American Journal of Epidemiology, 145*(1), 68-71.

Craglia, M., & Signoretta, P. (2004). GIS for joined-up government: The case study of the Sheffield children service plan. In J. Stillwell, & G. Clarke (Eds.), *Applied GIS and spatial analysis* (pp. 137-149). Chichester: John Wiley and Sons, Ltd.

Department of Environment Transportation and the Regions. (2000). *Indexes of deprivation 2000: Regeneration research summary No. 31*. London: DETR.

Diez-Roux, A. V. (1998a). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *American Journal of Public Health, 88*(2), 216-222.

Diez-Roux, A. V. (1998b). On genes, individuals, society, and epidemiology. *American Journal of Epidemiology, 148*(11), 1027-1032.

Diez-Roux, A. V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health, 21*, 171-192.

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., Link, B. G., & Northridge, M. E. (2000). A multilevel analysis of income inequality and cardiovascular disease risk factors. *Social Science and Medicine, 50*(5), 673-687.

Duncan, C., Jones, K., & Moon, G. (1996). Health-related behaviour in context: A multilevel modelling approach. *Social Science and Medicine, 42*(6), 817-830.

Gayle, H. D., Yip, R., Frank, M. J., Nieburg, P., & Binkin, N. J. (1988). Validation of maternally reported birth weights among 46,637 Tennessee WIC program participants. *Public Health Report, 103*(2), 143-147.

Gordon, D. (1995). Census based deprivation indexes: Their weighting and validation [Suppl. 2]. *Journal of Epidemiology and Community Health, 49*, S39-44.

Gordon, D., & Forrest, R. (1995). *People and places 2: Social and economic distinctions in England*. Bristol: School for Advanced Urban Studies and Bristol Statistical Monitoring Unit.

Green, H. G., Nelson, C. J., Gaylor, D. W., & Holson, J. F. (1979). Accuracy of birth certificate data for detecting facial cleft defects in Arkansas children. *Cleft Palate Journal, 16*(2), 167-170.

Harris, R., & Longley, P. (2004). Targeting clusters of deprivation within cities. In J. Stillwell, & G. Clarke (Eds.), *Applied GIS and spatial analysis* (pp. 89-110). Chichester: John Wiley & Sons, Ltd.

Harris, R. J., & Frost, M. (in press). Indicators of urban deprivation for policy analysis GIS: Going beyond wards. In D. Kidner, G. Higgs, & S. White (Eds.), *Socioeconomic applications of geographic information science innovations in GIS 9*. London: Taylor and Francis.

Harris, R. J., & Longley, P. A. (2002). Creating small area measures of urban deprivation. *Environment and Planning A, 34*, 1073-1093.

Hurley, S. E., Saunders, T. M., Nivas, R., Hertz, A., & P., R. (2003). Post office box addresses: A challenge for geographic information system-based studies. *Epidemiology, 14*(4), 386-391.

Kharrazi, M., Epstein, D., Hopkins, B., Kreutzer, R., Doebbert, G., Hiatt, R., et al. (1999). Evaluation of four maternal smoking questions. *Public Health Rep, 114*(1), 60-70.

Lee, P. (1999). Where are the deprived? Measuring deprivation in cities and regions. In S. Dorling, & S. Simpson (Eds.), *Statistics in society: The arithmetic of politics* (pp. 172-180). London: Arnold.

Lee, P. R., Moss, N., & Krieger, N. (1995). Measuring social inequalities in health: Report on the conference of the National Institutes of Health. *Public Health Report, 110*(3), 302-305.

Longley, P., Goodchild, M., Maguire, D., & Rhind, D. (Eds.). (2001). *Geographic information systems and science*. John Wiley & Sons.

Longley, P. A., & Harris, R. J. (1999). Towards a new digital data infrastructure for urban analysis and modelling. *Environment and Planning B, 26*, 855-878.

Matteson, D. W., Burr, J. A., & Marshall, J. R. (1998). Infant mortality: A multi-level analysis of individual and community risk factors. *Social Science and Medicine, 47*(11), 1841-1854.

McDermott, J., Drews, C., Green, D., & Berg, C. (1997). Evaluation of prenatal care information on birth certificates. *Paediatric and Perinatal Epidemiology, 11*(1), 105-121.

McElroy, J. A., Remington, P. L., Trentham-Dietz, A., Robert, S. A., & PA., N. (2003). Geocoding addresses from a large population-based study: Lessons learned. *Epidemiology, 14*(4), 399-407.

Morris, R., & Carstairs, V. (1991). Which deprivation? A comparison of selected deprivation indexes. *Journal of Public Health and Medicine, 13*(4), 318-326.

Pattenden, S., Dolk, H., & Vrijheid, M. (1999). Inequalities in low birth weight: Parental social class, area deprivation, and “lone mother” status. *Journal of Epidemiology and Community Health, 53*(6), 355-358.

Pearl, M., Braveman, P., & Abrams, B. (2001). The relationship of neighborhood socioeconomic characteristics to birthweight among 5 ethnic groups in California. *American Journal of Public Health, 91*(11), 1808-1814.

Perinatal Periods of Risk Work Group. *Perinatal periods of risk approach: The U.S. urban experience*. Retrieved October 12, 2004, from http://www.citymatch.org

Pickett, K., & Pearl, M. (2001). Multilevel analyses of neighborhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health, 55*, 111-122.

Ratcliffe, J. H. (2004). Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Science, 18*, 61-72.

Rauh, V. A., Andrews, H. F., & Garfinkel, R. S. (2001). The contribution of maternal age to racial disparities in birthweight: A multilevel perspective. *American Journal of Public Health, 91*, 1815-1824.

Reichman, N. E., & Hade, E. M. (2001). Validation of birth certificate data: A study of women in New Jersey’s Healthy Start program. *Annals of Epidemiology, 11*(3), 186-193.

Roberts, E. M. (1997). Neighborhood social environments and the distribution of low birth weight in Chicago. *American Journal of Public Health, 87*, 597-603.

Rushton, G., & Lolonis, P. (1996). Exploratory spatial analysis of birth defect rates in an urban population. *Statistics in Medicine, 15*, 717-726.

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.

Snell, L. M., Little, B. B., Knoll, K. A., Johnston, W. L., Jr., Rosenfeld, C. R., & Gant, N. F., et al. (1992). Reliability of birth certificate reporting of congenital anomalies. *American Journal of Perinatology, 9*(3), 219-222.

Soobader, M. J., & LeClere, F. B. (1999). Aggregation and the measurement of income inequality: Effects on morbidity. *Social Science and Medicine, 48*(6), 733-744.

Sorlie, P. D., & Johnson, N. J. (1996). Validity of education information on the death certificate. *Epidemiology, 7*(4), 437-439.

Stead, M., MacAskill, S., MacKintosh, A. M., Reece, J., & Eadie, D. (2001). "It's as if you're locked in": Qualitative explanations for area effects on smoking in disadvantaged communities. *Health Place, 7*(4), 333-343.

Stretesky, P. B., & Lynch, M. J. (2004). The relationship between lead and crime. *J Health Soc Behav, 45*(2), 214-229.

Subramanian, S. V., Acevedo-Garcia, D., & Osypuk, T. L. (2005). Racial residential segregation and geographic heterogeneity in black/white disparity in poor self-rated health in the U.S.: A multilevel statistical analysis. *Social Science and Medicine, 60*(8), 1667-1679.

Subramanian, S. V., Chen, J. T., Rehkopf, D. H., Waterman, P. D., & Krieger, N. (2005). Racial disparities in context: A multilevel analysis of neighborhood variations in poverty and excess mortality among black populations in Massachusetts. *American Journal of Public Health, 95*(2), 260-265.

Subramanian, S. V., Kim, D. J., & Kawachi, I. (2002). Social trust and self-rated health in U.S. communities: A multilevel analysis [Suppl. 1]. *Journal of Urban Health, 79*(4), S21-34.

Subramanian, S. V., Lochner, K. A., & Kawachi, I. (2003). Neighborhood differences in social capital: A compositional artifact or a contextual construct? *Health Place, 9*(1), 33-44.

Subramanian, S. V., Nandy, S., Kelly, M., Gordon, D., & Davey-Smith, G. (2004). Patterns and distribution of tobacco consumption in India: Cross-sectional multilevel evidence from the 1998-99 national family health survey. *British Medical Journal, 328*(7443), 801-806.

Townsend, P., Phillimore, P., & Beattie, A. (1988). *Health and deprivation: Inequality and the North*. London: Croom Helm.

Walker, D. S., Schmunk, S. B., & Summers, L. (2004). Do birth certificate data accurately reflect the number of CNM-attended births? An exploratory study. *Journal of Midwifery and Womens Health, 49*(5), 443-448.

Chapter III

An Introduction to GIS (All Things Spatial)

The last chapter provided an introduction to two of the key components in a GIS: getting the data in, and then manipulating them to answer questions. This chapter considers how these data can be visualized and analyzed. The analysis section in this chapter will be fairly simplistic, as this topic is revisited again in Chapters V and VI with the introduction of increasingly more sophisticated measures. The chapter will end with a discussion of an important topic integral to urban area analysis but which is often missed when discussing the GIS: What exactly is a neighborhood and how should it be spatially defined?

Visualizing the Data

The most widely used role of a GIS in public health (or for that matter in any government agency) is as a visualization tool. Data of a GIS is spatial in nature, and spatial data basically means maps. You can use the GIS to make maps, and simply put, people understand maps far better than tables of numbers.

An "introduction to cartography" lecture in a GIS course will often present the reasons why maps displaying the same general information can vary. For example, consider a map showing citywide infant mortality. Why would maps of such a specific topic differ? One reason is the skill of the mapmaker; for example, if the map displays graduated colors (a choropleth map), the choice of colors, data classification type and breaks chosen, and the geographic aggregation mapped will all result in different surfaces. Aggregation problems have already been discussed in Chapter II with another issue being addressed at the end of this chapter — namely, how to spatially define a neighborhood.

The second reason for map variation is the range of data available. For example, the relationship (or lack thereof) between the mapmaker and the Vital Statistics Department will influence how spatially disaggregated these data are. Address level data opens other forms of visualization, including hot spot displays and infant mortality rate surfaces (these will be discussed in Chapter V). If the data is only released by zip code, the resulting types of map display are limited. It should also be pointed out that if a vital statistics department only releases data at the zip code level, sizeable error can occur if they rely on the birth and death certificate, do not geocode the addresses, and then aggregate these up to the zip code level. In other words, the address listed on the certificate does not match the zip code listed. Table 1 displays a summary table for infant death data for the Baton Rouge Healthy Start Region over a 3-year period. The *zip named* column lists the frequency of that zip code being listed on the birth certificate (linked to the death certificate). The *geocode* column lists how many times a death was aggregated to a zip code by finding where the listed street address falls onto a zip code map. The *unmatch* column lists how many of these addresses listing one of the named zip codes could not be geocoded. The *% error* column shows the difference between using the named zip codes, as compared to actually mapping the addresses (unmatched records have been subtracted from the *zip named* column in the calculation). The resulting error is large (obviously we are dealing with small numbers), yet resulting maps of infant mortality will look quite different.

If the mapmaker is connected to a health organization, such as a Healthy Start program, data accessibility will again be improved, resulting in more spatially disaggregated information and more quality controls in data recording, resulting in greater confidence and accuracy in the findings.

The third reason for map variation results from the software and hardware available to the mapmaker. Ten years ago (early 1990s) the software costs and computing power needed to run most GIS packages (usually a workstation), would limit GIS use to academic departments. Today all costs have dropped to the point of widespread availability. A suitable PC, with printer, GPS unit, and

Table 1. Differences in zip code destination

	Zip Named	Geocode	Unmatch	% Error
70805	28	20	1	26
70806	17	11	1	31
70807	20	8	0	60
70811	12	9	0	25
70812	13	8	0	38

lower-end vector GIS package could all be purchased for less than $2,000. Where limits still exist in terms of mapmaking are the larger peripherals, such as large format plotters and scanners. Even so, as long as finished maps are only meant for presentation at larger gatherings (usually with the aid of a digital projector), and finished printing is A4 size, these technical limitations are probably the least of today's cartographic problems. Even so, the justification of that $2,000 price tag to agencies unfamiliar with GIS can still be somewhat of a tough sell.

The final reason for variation depends on the audience for the map. The Baton Rouge Healthy Start is an organization serving the local community. At any given point it could be presenting information to caseworkers inside the program, community representatives from the neighborhoods it serves, political representatives from the state or city, evaluation teams from the national office, or even to "professional" audiences at national conferences. Each one of these audiences (and this list is by no means exhaustive) will understand and respond to the same information in different ways.

Consider the two following examples, caseworkers and community representatives, as map audiences. Caseworkers are arguably the most important audience for a community health organization's GIS output. After all, what use is an analysis displaying neighborhoods of high numbers of women receiving no prenatal care if an outreach strategy cannot be formulated? In Chapter IX, an information system is presented that was developed for caseworkers in the Baton Rouge Healthy Start. The idea behind the information system was to create a user-friendly electronic guide that would allow each caseworker to extract relevant neighborhood information that could better prepare the worker in the field. Maps created for caseworkers should be rich in local detail; air photos should be as enlarged as possible. The issue of patient confidentiality is not a barrier with this audience, so even apartment buildings can be targeted. The information system contains background data such as what proportions of deliveries in a particular apartment building had been low birth-weight in previous years. Road names are also needed, as are driving directions. The caseworkers are not familiar with census boundaries, but a common geographic language between health workers is the zip code. Quite often health workers will refer to a specific zip as having a whole host of problems. So actually displaying these boundaries can be quite interesting for them as they make the connection between a geographic name (the zip code) and an actual area of the city, especially as most people have no idea what the shape of his or her own zip code is like. Census divisions are even more spatially unfamiliar, as most researchers would be hard-pressed to locate any given census tract on a city street map. In order to make a John Snow-style discovery, such as that infant mortality rates are highest close to major road intersections (*this is a hypothetical example and not a finding!*), we need to be able to understand and be familiar with the underlying geography.

It has been a worthwhile experience to sit and observe these local experts as they navigate through the information system, as they often provide insights into the urban geography that very few GIS experts will be aware of. For example, many neighborhoods in Baton Rouge go by colloquial names (Cadillac City), each bringing with it a local perception. These names are not shown on any map, but gather a group of caseworkers around an aerial photograph and soon they will be able to pick out the key features. As a final point, caseworkers do not need to see detailed outputs of spatial models. For example, the output of a spatial filter, which will be presented in Chapter VI, is a contour map. This is a usual spatial expression for academics, but for caseworkers, a generalized area covering the "hot spot" is all that is needed.

In comparison, a map presented to the community should be informative, displaying general patterns of what is happening in the neighborhood, while at the same time preserving program participant confidentiality. Aerial photos are a useful way for most people to gain a spatial reference frame, as they can "see" key features such as churches or road intersections in their neighborhood. Such reference points and other important city landmarks, such as major streets, malls, etc. should be labeled. However, the map should not be enlarged beyond a certain scale, as buildings displaying actual health events would compromise confidentiality. If health "locations" are needed, such as actual birth addresses, grossly exaggerated dots covering a portion of a city block, and still displaced (e.g., to the nearest street intersection), have been used in the past, though current HIPAA regulations even make these displays problematic. A better approach would be to once again generalize the findings of spatial models (such as hot spots) and lay these onto the map. There would be no need for outlines of census tracts or census blocks, though again this audience would be familiar with the concept of zip codes. Additional infrastructure displays could include bus routes, libraries, clinics, and so forth.

For both of these maps, a clear legend, north arrow (to indicate directions), and scale (so distances can be estimated between features) should be included.

The maps just described contain the two general types of data that comprise all cartographic displays: qualitative data, which means important locations are mapped (such as bus routes), and quantitative data, whereby intensities such as the infant mortality rate are displayed (in the previous example as contour lines).

Chapter I discussed the importance of finding spatial associations in relation to a health outcome, the traditional example being John Snow's cholera concentration around the Broad Street Pump, which could also be called an exploratory spatial analysis. The same type of point pattern map, or pin map as they are commonly referred to in non-GIS circles (referring back to the old, wall-based city map with pins literally being stuck in), can reveal a great deal about underlying spatial processes. An informed observer, such as a caseworker, could look at residences of women receiving no prenatal care and visually identify high intensities. By drawing on contextual information, such as knowledge of the city's neighborhoods, and an appropriate background image, such as an aerial photo, the caseworker can begin to formulate an outreach strategy. Obviously, we now have the techniques to help better inform the caseworker when making this judgment, but the point is that most people can respond effectively to a map if effective cues (spatial references such as road intersections) are used.

Sometimes it is not enough to simply show key locations, we have to show quantitative surfaces. Consider the following paragraphs extracted directly from the Maternal and Child Health Bureau Division of Perinatal Systems and Women's Health grant application instructions for eliminating disparities in perinatal health (2004):

A project area is defined as a geographic community in which the proposed services are to be implemented. A project area must represent a reasonable and logical catchment area, but the defined areas do not have to be contiguous. Communities are broadly defined so that multi-county projects serving racial/ethnic or other disparate groups (e.g., Hmongs, Native Americans, etc.) would be eligible.

…the project area for which the applicant is applying;…that the infant mortality rate (IMR) for the proposed area is at least 10.58 deaths/1,000 live births (which is one and a half times the national infant mortality rate for the period 1999 through 2001).

2. Community Assessment

Provide a clear description of the current status, capacity, and needs of the proposed geographic project area and the current perinatal system serving that area. Include demographic and health statistics to support the presentation and to demonstrate current prevalent disparities. For comparison to other applications, applicants must present data minimally from (3-year average) 1999-2001. If more current data is available, for example 2000-2003, it may also be included. Describe (by race/ethnic origin) the perinatal health indicators including, 3-year averages (1999-2001) for live births, infant deaths (under 1 year of age), neonatal and postneonatal mortality rates, as well as the incidence of low birth weight, SIDS, births to teenagers 18 years and younger, trimester of initiation of prenatal care and adequacy of prenatal care. Highlight current trends in morbidity, including such areas as birth defects, infant/child abuse and neglect, accidents, AIDS, other communicable diseases and other prevalent factor(s) affecting the project area.

3. Identification of the Targeted Population(s)

Briefly describe the size, demographic characteristics, prevalent norms, and health behaviors of the targeted population(s). Please include for women of child bearing age: data on poverty, average education level, employment status, and major industries.

Data required by these instructions, though often appearing as tables in the proposal, are by their very nature spatial and therefore begging to be mapped, especially given the availability of GIS. However, although most GIS software packages make mapmaking a rather simple task, there is still a need to follow sound cartographic design principles. The following section will provide some basic cartographic guidelines that should be considered when compiling a map with a GIS, with rule number one being that every good map should include a map title, legend, scale, north arrow, and map credits, which provide information about the data shown on the map and the people or organization compiling the map.

Most map types used in public health are so-called *quantitative thematic maps*. In a thematic map, the quantitative data will be visualized on top of a geographic or base map. The base map may simply include the boundaries of some administrative or statistical unit, such as a zip code or census tract. Among the thematic maps discussed in the following section, the choropleth map is more often used for the display of public health data than any other map. For this reason, the choropleth map is discussed in more detail. The interested reader is also encouraged to consult with the references at the end of this section, which provide additional information.

Choropleth Map

The choropleth map is a common technique for representing enumeration data. The International Cartographic Association (ICA) defines it as "a method of cartographic representation which employs distinctive color or shading applied to areas other than those bounded by isolines. These are usually statistical or administrative areas" (Dent, 2002). The choropleth map is a very popular mapping technique because it is easy to construct and to understand. In addition, it can readily map data collected by the U.S. Census, because such data are mostly based on administrative areas (census tracts, counties/parishes, etc.). Several examples have already appeared in the first two chapters, for example, Figure 6 in Chapter II displayed choropleth maps at three different spatial aggregations.

To compile a choropleth map, data need to be collected at statistical or administrative areas (for example census tracts, counties/parishes). These data are subsequently classified into groups. Finally, an areal symbolization scheme (shadings of gray or color hue, or a black and white pattern) is devised and applied to the classified data. Each data value will be represented in the map according to the class (with resulting color shade) in which its value falls.

The choropleth map provides an overall picture of the spatial distribution of the data being mapped. It is based on the assumption that data are equally distributed within each enumeration unit. This is, of course, a simplification of reality, because we know that, for example, the infant mortality rate is not equally distributed across the East Baton Rouge Parish or even across a particular census tract within this parish. In general, the smaller the enumeration unit, the closer the representation will be to reality, and vice versa. Since in choropleth mapping data is being equally distributed within each enumeration unit, it is impossible to make inferences about unequal distributions within the same enumeration unit. Geographers refer to this phenomenon as the *Ecological Fallacy* — the same problem that occurs when neighborhood-level information is attached to people who live inside that neighborhood, the point-in-polygon procedure described earlier, and an important component of multi-level modeling. One way to get around this problem in mapping is by adding additional information to a choropleth map in order to refine and more accurately depict the spatial distribution of the original data. For example, the spatial distribution of hydrological features (rivers, lakes, bayous) could be added to a choropleth map showing the infant mortality rate in East Baton Rouge Parish on the basis of census tracts. Since we know that people do not usually live in rivers, lakes, or bayous, we can safely assume that the infant mortality rate at these locations must be zero. Accordingly, original infant mortality rates outside rivers, lakes, and bayous must increase somewhat. The amount of increase depends on the size of these water features. The larger these features are, the higher the increase and vice versa. It is obvious that the more geographic information is added to the same choropleth map, the more accurate the spatial distribution of the original data will be. The new and revised choropleth map is called a dasymetric map (McCleary, 1969). An alternative approach would be to use remotely sensed images or high-resolution air photography to key out unpopulated areas (Harris, 2003; Harris & Frost, in press; Longley & Mesev, 2002). Fortunately, both of these approaches are easily achievable within a GIS.

Although choropleth mapping is often used, it is almost as frequently misused. A few guidelines will assure an appropriate use of this technique. First, choropleth maps should only be used for visualizing derived data values, such as rates (infant mortality rate), densities (number of doctors per mile squared), or percentages (proportion of mothers with no prenatal care visits). There are very few occasions when this technique should be applied to map actual or absolute values (number of births).

Second, use an appropriate number of class groups for the data. Cartographers have recommended somewhere between four and eleven classes, though I strongly suggest never using more than five to seven classification groups due to the difficulty in distinguishing between color gradations. An often-cited general rule of thumb suggests seven, give or take two, classes (Gilmartin & Shelton, 1989). Among other things, the number of classes selected will depend on the total number of enumeration units, whether areal symbols are depicted in black and white or in color, and the complexity of the map and the medium of presentation (paper or computer screen). Independent of the number of classes chosen, what is important is to make sure that the areal symbols can accurately be identified and are easily discernible by the map reader.

Third, use an appropriate classification method. This is easier said than done, because numerous classification methods exist. The selection of an appropriate classification method is important, since the look of the choropleth map can change significantly with different classification methods. Most GIS software will offer at least the following four classification methods:

Equal steps:	Data values are grouped into classes with constant intervals (0-10, >10-20, > 20-30, etc.).
Quantiles:	Puts an equal number of data values in each class.
Natural breaks:	Creates natural groups of similar values.
Standard deviation:	Creates constant class intervals with class width often equal to one standard deviation.

Of these classification methods, natural breaks is always a good choice. It guarantees that the numerical differences within groups are less than the differences between groups. If choropleth maps were used to compare different variables for the same study area (one map per variable), then classifying by standard deviations would be most appropriate. Whatever classification method is selected, it should be mentioned on the map — for example, underneath the map legend.

Finally, use an appropriate areal symbolization scheme. Most common GIS software will offer a wide array of such schemes. Cartographic guidelines suggest going from light (light gray, color, or pattern) to represent classes with the smallest values to dark (dark gray, color, or pattern) for classes with the largest values. Brewer (1994) refers to these areal symbolization schemes as sequential color schemes. A diverging color scheme is usually the combination of two different sequential color schemes. Diverging color schemes are usually applied when some enumeration units show an increase in values while others show a decrease. A second example would be to visualize values falling above and below a certain threshold.

There is much more that could be said about choropleth mapping. Usually, introductory cartographic textbooks devote one entire chapter to this map type. Interested readers should consult the introductory cartographic textbooks listed in the reference section for further information. It should also be kept in mind that the cartography and mapping terminology sometimes used in GIS software might be different from the terminology used in these textbooks (for example, a choropleth map is sometimes referred to as a graduated color map).

Common Dot Map

A second, less frequently used map type is the common dot map. This involves the selection of an appropriate point symbol to represent occurrences of the data being mapped. The number of point symbols changes in proportion to the number of occurrences. In one-to-one mapping, one dot represents exactly one occurrence (the typhoid map shown in Chapter I is an example), in one-to-many mapping, one dot represents more than one occurrence. Dot maps can be used to display absolute values, such as the number of infant deaths or the number of women receiving no prenatal care. Such values are usually collected for enumeration units, similar to choropleth maps. The precise location of each dot symbol within each enumeration unit is critical. From a cartographic design perspective, a spatially random distribution of dot symbols within each enumeration unit is preferred over a regular distribution. Regular distributions look rather boring and suggest that geographic phenomena are uniformly distributed across space, which rarely happens. Random distributions do not suggest this, and they also look more interesting. Usually, the smaller the enumeration unit, the more accurate the dot placement will be. On the other hand, a very accurate dot placement (in one-to-one mapping) might not be desirable either, because this might lead to the identification of an individual and might thus compromise his/her individual privacy.

The "look" of dot maps varies with the chosen dot size (how large each dot is) and the unit value (the count represented by each dot). If there are too few dot symbols and the chosen dot size is too small, then the dot map might look empty. On the other hand, if there are too many dots and the size of each dot is too large, then the map might appear overcrowded. A good dot map shows a balance between these two extreme cases. To construct dot maps by hand is very tedious and time-consuming. Fortunately, most GIS software allows for the computer generation of dot maps, with the user specifying the dot size and the unit value.

Isarithmic (Isoline) Map

An isarithmic map is a planimetric graphic representation of a three-dimensional volume with quantitative line symbols. These line symbols are commonly known as isolines and points along the same isoline have the same value. Isolines are sometimes named after the phenomena they represent, for example, an isohypse (a contour) represents elevation above mean sea level, or an isobar represents atmospheric pressure. Because isoline maps produce a continuous surface, they have commonly been used to represent continuous phenomena. These are phenomena that can occur anywhere, at any point, on the surface of the earth. For example, many climatic phenomena are continuous, such as temperature, precipitation, and barometric pressure. Most social phenomena are discrete, which means that they occur at specific locations. Examples are the residence of a pregnant mother or the location of a health care center. Nevertheless, isoline mapping can also be applied to social data that have been collected at points (street addresses) or aggregated over geographic areas (census tracts, counties/parishes). In the case of an area, a point will be placed at the midpoint (center of gravity or the center of the largest city) of the enumeration unit with all data collected for this area being attached to the point. Isolines are constructed by spatially interpolating these data points. Isolines can map absolute values (infant death locations with each location being represented by the mother's residence), or derived data values, such as rates (infant mortality rate), densities (number of medical doctors per mile squared), or percentages (proportion of mothers with no prenatal care visits). Mapping only absolute values (such as infant death locations) is straightforward but might be misleading, because of a missing background population (e.g., total number of birth locations). It is therefore better and more realistic to map derived data values as isolines. Finally it should be noted that the accuracy of any isoline map is a function of the interpolation method used, the quantity of points or enumeration units (areas) collected and the size of these units. Usually, the more points or areas that are collected or the smaller the sizes of these enumeration units, the higher the accuracy of the final isarithmic map.

Proportional (Graduated) Point Symbol Map

To construct a proportional point symbol map, a symbol form (circle, square, triangle, etc.) is selected and its size is changed across space in proportion to the quantities it represents. It is very flexible because it allows mapping of absolute, as well as derived, data values. This map type is selected when data occur at points (number of infant deaths in large cities across the United States) or when they are aggregated at points within areas (infant mortality rate for each parish in Louisiana). For several decades, researchers (mostly cartographers and psychologists) tried to understand the mechanisms by which map readers perceptually scale quantitative symbols. An important finding of this research is that most people underestimate circle sizes and this underestimation increases with larger circle sizes. Today, cartographers suggest mapping proportional point symbol by range grading. With this approach, data are first classified into groups and each group is represented by a proportional symbol that is clearly distinguishable from other symbols in the series. The overall goal is the discrimination of symbol sizes rather than magnitude estimation. Range-graded circle series can be found in any introductory cartography textbook. Symbols can be filled or not, as long as they provide enough contrast to the underlying map information. As a general rule, when symbols overlap, smaller ones should be placed in front of (or on top of) larger ones.

The reader is now pointed towards the following introductory cartography texts: Campbell (2000); Dent (2002); Muehrcke, Muehrcke, and Kimerling (2001); Robinson, Morrison, Muehrcke, Kimerling, and Guptill (1995); and Slocum, McMaster, Kessler, and Howard (2004): though there is no substitute to taking a computer cartography course at a local university.

Spatial Analysis

The most underutilized aspect of a GIS is in using the system to perform a spatial analysis. Although there is a great deal that can be gained from being able to understand the spatial patterns of risk facing pregnant mothers in a program study area, the problem is that many people with access to a GIS do not know how to perform appropriate analyses. Many government agencies employ GIS users who are well-versed in heads-up digitizing, who know how to program for a Web-based GIS, who even know the quirks of different projection systems, but ask them to solve a simple spatial problem…. I recently posed a question of my GIS class for their first test:

Four addresses have stockpiles of vaccines (20 vaccines at each stockpile), and these vaccines cannot travel more than 2 miles. If smallpox cases were found at the following addresses, and the number of people living inside each were X, which stockpile should serve each house? Would anyone be missed? And if we were to make the question a little more imposing, let us consider the numbers of people living in neighboring houses, and actual road distance travel time.

In many ways these students are in a similar situation to those early epidemiologists searching for typhoid patterns. We have a tool that can visualize a spatial problem, and these maps may indeed provide an insight into the problem, but the ability to actually *find* a relationship is often missing.

Of course, this does not mean techniques are not available. In fact, there has been an explosion of spatial analysis techniques and approaches made possible because of GIS. Several books and even Web sites, some offering free-download software that can be linked to common GIS packages, are now available in an area that is often referred to as the GIS/Spatial Analysis Interface. As this is such an important aspect of GIS use, Chapters V and VI will be dedicated to the spatial analysis of health data, with the first of these presenting simple solutions, while the second provides an overview of the latest techniques.

I always tell my students to keep it simple, if possible. Any GIS can be used to create basic summary or descriptive statistics of the data. These include sum, range, mean, minimum, maximum, variance, and standard deviation. By combining these summary statistics with some of the queries mentioned in the previous section, powerful insights can be gained into spatial patterns of pregnancy outcomes. For example, if all pregnant women are selected within 1 mile of each clinic, the summary statistics will show, for any attribute field, how similar the birth experiences were. By a simple manipulation, all women exceeding, for example, two standard deviations within this area, could easily be identified and extracted. The attribute in question may be the result of a questionnaire investigating the perception of environmental stress (such as a high rodent population), or data extracted from a birth certificate, like weeks of gestation. By comparing this "risk" group to similarly sized neighborhoods around the city, specific risk problems can be identified and mitigation strategies defined. As another example, due to budget cuts in Louisiana many community-based health units were closed, with rural areas being hardest hit. A concern was that birth outcomes would suffer around these closed clinics, as women would have to travel further for services. By comparing birth outcomes in a 5-mile radius both before and after the closure, an impression can be gained as to how the closure may have impacted the region. A further comparison with a similar-size area around clinics that had not closed could provide data needed for statistical analysis. In both cases, those births falling inside the buffer area would be extracted and their basic statistics summarized.

Most federally funded pregnancy programs are already following this approach, even though they may not be thinking explicitly about the spatial dimension. For example, the Baton Rouge Healthy Start is asked to provide summary statistics for its annual report at both its program region (a five zip code area) and the parish as a whole. This summary information can be used to compare program performance against the wider area, other Healthy Start programs, or even against itself once enough years of data have been collected. The GIS also allows for the service region to be further subdivided, to see which of the five zip codes have been performing at expected levels, and which still remain problematic.

Chapter V takes this approach a step further by presenting a difference of proportions t-test to identify statistically significant variations between neighborhoods. Other tests that could be employed include cross-tabs, analysis of variance, and so on. Although many books have been written describing how such tests could be encoded within a GIS (O'Sullivan & Unwin, 2002), a basic familiarity with statistics and spatial analysis, and of course its associated pitfalls, is often all that is needed. The importance lies in the skill of the analyst contemplating how the GIS could be used to answer the question in hand, rather than what techniques in easy-to-use drop-down menus are available.

Having said this, there has also been a growth in more academically oriented, exploratory spatial analysis techniques made possible because of GIS. Increased computing power allows for larger data sets to be manipulated in real time. Analysis results can immediately be displayed as maps and charts, with some programs now allowing for data to be entered and removed on-screen in order to see how modifications (such as the removal of outliers) affect the analysis. These software packages are either in the public domain or can be purchased from an academic institution or a private vendor. If software packages are in the public domain, they can usually be downloaded from a Web site free of charge. If the software development is at a university, then it can be purchased for a reasonably low price. Costs associated with a private vendor are higher. Software packages can be extensions of existing GIS software (e.g., Spatial Analyst) or they are stand-alone packages that perform the analysis, but do not have capabilities to visualize the results. In cases like this, analysis results need to be imported into a GIS for mapping.

All software packages that will be briefly discussed below are stand-alone packages and most of them can be downloaded from a Web site free of charge. These software packages are listed in alphabetical order, which does not reflect their importance toward analysis and mapping of health-related data. This list is by no means comprehensive and due to the rapid development in the field of spatial analysis, will become outdated quickly.

CrimeStat®

CrimeStat® is a spatial statistics program for the analysis of crime incident locations, developed by Ned Levine & Associates under grants from the National Institute of Justice. The program is Windows-based and interfaces with most desktop GIS programs, including ArcView®, MapInfo®, Atlas*GIS™, Surfer® for Windows, and ArcView Spatial Analyst©. The purpose is to provide supplemental statistical tools to aid law enforcement agencies and criminal justice researchers in their crime mapping efforts. Although CrimeStat® is a spatial statistics program for the analysis of crime incident locations, it can be successfully applied to analyze any incident locations (e.g., health-related data) in the form of street addresses or already expressed as x- and y-coordinates. CrimeStat® should be of interest to the health community because it can be used to identify spatial and space-time cluster, and interpolate incident locations to a continuous density surface. The current version is 3.0. CrimeStat® is in the public domain and can be downloaded for free from the following Web site: www.icpsr.umich.edu/NACJD/crimestat.html. This Web site includes the software application, sample data, a comprehensive manual, and a reference section.

GeoDa™

Anselin, Syabri, and Kho (forthcoming) describe GeoDa™ as "a user-friendly and graphical introduction for non-GIS specialists. It includes functionality ranging from simple mapping to exploratory data analysis, the visualization of global and local spatial autocorrelation, and spatial regression. A key feature of GeoDa™ is an interactive environment that combines maps with statistical graphics, using the technology of dynamically linked windows." GeoDa™ is a reinvention of the original SpaceStat package (Anselin, 1992) and the prototype version of GeoDa™ was originally known as DynESDA. GeoDa™ is in the public domain and can be downloaded for free from the following Web site: http://sal.agecon.uiuc.edu/geoda_main.php. This Web site includes the software application, tutorials, and sample data.

Geographically Weighted Regression (GWR)

GWR is a method of analyzing spatially varying relationships and is a technique for exploratory spatial data analysis. It involves fitting a model to predict the values of one variable (response or dependent variable) from a set of one or more independent (predictor) variables (A. S. Fotheringham, Brunsdon, & Charlton, 2002). While "normal" regression assumes that the relationship between dependent and independent variables holds everywhere in the study area, GWR allows the regression parameters to vary locally. The software, including the manual, can be purchased directly from the authors. For more information go to the following Web site: http://www.ncl.ac.uk/geps/research/geography/gwr/.

SaTScan™

SaTScan™ was developed by Martin Kulldorff to analyze spatial, temporal and space-time data (Kulldorff et al., 1997, 1998). It can be used to identify (1) spatial or space-time disease clusters, (2) whether a disease is randomly distributed over space, over time or over space and time, (3) to evaluate the statistical significance of disease cluster alarms, and (4) to perform repeated time-periodic disease surveillance for the early detection of disease outbreaks. The data may be either aggregated at the census tract, zip code, county, or other geographical level, or there may be unique coordinates for each observation. SaTScan™ can be downloaded for free from the following Web site: www.satscan.org/. This Web site includes the software, sample datasets, a bibliography, and the software manual. The current version 5 of SaTScan™ was released on September 21, 2004.

The techniques offered in many of these programs can broadly be divided into either exploratory or confirmatory analyses. Chapter VI will discuss the following point-pattern techniques: Spatial Filter (Rushton & Lolonis, 1996), Kernel Density (Levine, 2004), and nearest neighbor hierarchical spatial analysis (Levine, 2004). In general, these techniques seek to identify spatially significant clusters of points (such as infant deaths, cases of cancer) and compare them to a known (theoretical) distribution. Tests of statistical significance are often generated by a Monte Carlo approach (the generation of thousands of simulated data sets to form a comparison distribution), which is now easily achieved within the GIS environment. The GIS allows for these results to be quickly visualized onscreen. Although these techniques do not find spatial associations, they do provide useful exploratory tools to investigate where further analysis is needed. Once this exploratory analysis has been completed, confirmatory analysis searching for causations can be undertaken. In other words, actual questions can be asked and hypotheses ventured, such as whether high infant mortality clusters fall in neighborhoods with high crime rates and liquor availability (both variables suggesting a neighborhood stress association).

In the social sciences, especially outside of geography, the most common spatial analysis method would be a variant of multiple regression calibrated on a surface of aggregated units (such as census tracts or zip codes). A dependent variable (low birth weight) could be regressed against independent variables (race, income, age of mother, etc.). By mapping the residuals, outliers (either areas of the city doing well or poorly) could be identified. Investigating residuals of the analysis could give insights into why some neighborhoods did not "fit" into traditional expectations, such as being correlated to income. Associations between any two variables are likely to vary across space (between towns, zip codes, even neighborhoods). The residual approach can unearth some of these outlier areas, but the relationship is still "global" in nature. For example, a traditional citywide model may find a negative relationship between education and low-birth-weight deliveries. However, this relationship for some neighborhoods may not be statistically significant due to the presence of a third variable (Anselin, 1995; Fotheringham and Rogerson, 1993), such as a strong church presence. In order to capture these "local effects," techniques such as Spatial Regression or Geographical Weighted Regression (Fotheringham, Brunsdon, & Charlton, 1998, 2002) can be used. Again, these methods will be discussed in more depth in Chapter VI.

A further criticism of traditional models (such as regression analysis) being applied to spatial data, and also a justification for using one of the spatial autoregressive models, is the possibility of model violation, as it is a requirement that each unit of space is independent of the other, which is simply not the case (one of Tobler's Laws of Geography states that there is an inverse relationship between distance and similarity, which in effect means they are not independent observations). For example, are we to assume that social conditions in one census block (such as crime levels) do not have a spill-over effect onto neighboring census blocks? The second problem is again the ecological fallacy, or the spatial version known as the Modifiable Area Unit Problem (MAUP). Although this problem has been previously discussed, it is important to restate that results can vary according to the way the aggregation of space (county, census tract, zip code), is decided upon, both due to scale (the size of the aggregation of space), and the configuration of that area. An analysis of infant deaths at the zip code level could find no significant relationship between infant mortality and birth weight, as affluent mothers would be able to afford the care needed to cope with medical conditions arising from the low-birth-weight delivery. The changing of scale to a census block group would find no ameliorating influence of wealth, and low-birth-weight babies born to indigent mothers would more likely result in an infant death.

One last type of analysis that has not been mentioned here, and will not be covered in depth in this book, is that of efficient accessibility. Location allocation routines using a heuristic approach can locate the most efficient placements of population serving centers, such as clinics or ambulance stations. There is an extensive literature on the effectiveness of these algorithms (Hodgson, 1988; McLafferty & Broe, 1990; Mohan, 1983; Scarpaci, 1984), and most of the commercial GIS packages contain location tools. However, it is also possible to gain an understanding of gaps in service provision by using other, simpler procedures. From basic economic geography theory, concepts such as range, threshold, and break point can be used to identify service areas. The range can be defined as the furthest distance people are willing to travel. The threshold is the required number of inputs needed to sustain a center (for example, patients within "range" of a clinic). The break point is the middle distance between two servicing centers, which is halfway between the two if no weight is added, or is derived as a proportion if weight (such as number of beds, number of doctors, etc.) and separating distance are included in the calculation. These concepts can be approximated by several functions within a GIS analysis, such as density analysis or proximity analysis. Using density analysis as an example, each center (such as a clinic), would become the center of a density calculation, the radius of which could be defined by the range (in this instance, how far people are willing to travel). Where these density windows overlap, the color changes. The resulting surface of the city is a patchwork of these densities, with areas poking through the surface where no coverage exists. Further complexity can be added by using a distance decay function associated with the clinics, which can be achieved by running a kernel density instead of standard density, the general premise being that although people may travel up to 2 miles for a service, they are more than twice as likely to travel if the distance is 1 mile. In addition, the range may be varied according to the underlying socioeconomic characteristics of the population. Indigent populations are less likely to have private transport, and are likely to have smaller ranges than suburbanites. Although more sophisticated location allocation analysis would be advisable for the construction of a new clinic, the approach described here could be used to identify which neighborhoods outreach workers should target.

GIS as a Management Information System

Although the last two chapters have detailed the potential of incorporating a GIS into an analysis of health data, it should also be remembered that we are working with a Geographic *Information System*. Indeed, Chapter IX will describe how the GIS can be used as a management information system (MIS), organizing all data for an entire community-based health care program. The MIS can be used to collect, synthesize and coordinate multiple data layers, resulting in informed decision-making strategies. In a more sophisticated MIS, these strategies can also be automated as spatial decision support systems. An example is provided in Chapter X, where a syndromic surveillance system for bioterrorism is suggested. However, the technology could also be used to efficiently manage resources throughout a public health system, such as the distribution of flu vaccines. Each recognized center administering these vaccines could be linked into a central GIS, with each given dose being subtracted from the system stockpile. When dose levels fall below a given threshold, back-up doses could be ordered, or even sequestered from other neighboring centers. Similar systems are already in place for retail chains. This type of enterprise GIS is certainly the future for health care, though it is debatable whether such a sophisticated system is needed for community-level centers. For centers of this size, automation could be used for the generation of a phone call to remind a program participant that a clinic appointment is due. The East Baton Rouge Parish Office of Homeland Security and Emergency Preparedness currently uses a community alert system, called CAL, that can dial any number needed, for any geographic area, with a prerecorded message. This system can spatially target all phones within an area with "advice" during an evolving emergency. The same technology could also be linked to particular subsets of program participants extracted by the MIS, such as calling all first-time mothers, or all teenage mothers, about a special class being offered at the program center.

What is a Neighborhood?

This chapter has so far presented an overview of the visualization and analysis of spatial data. Most of the traditional causations ventured for low-birth-weight deliveries, or infant deaths, create spatial footprints. These "risks" will be discussed in more detail in Chapter IV, but before we get to this stage it is useful to briefly consider the "geographies" of the analysis or visualization. Chapter II also discussed how different spatial aggregations can affect both analysis and visualization. A related consideration is, "What is the geography we are working with?" Or more specifically, how do we determine appropriate spatial areas, which are usually referred to as neighborhoods.

There have been several studies relating birth outcomes to neighborhood effects (Pearl, Braveman, & Abrams, 2001; Rauh, Andrews, & Garfinkel, 2001), these being only part of a larger approach to health data analyses running through the 1980s and 1990s, which considered the representation of an individual's "place" (Carstairs & Morris, 1989; Diehr et al., 1993; Humphreys & Carr-Hill, 1991; Krieger, 1992), or more specifically their "neighborhood" (Anderson, Sorlie, Backlund, Johnson, & Kaplan, 1997; Diez-Roux et al., 2001; Duncan, Jones, & Moon, 1999; Kleinschmidt, Hills, & Elliott, 1995; LeClere, Rogers, & Peters, 1998; O'Campo, Xue, Wang, & Caughy, 1997; Robert, 1999; Roberts, 1997). As most GIS analyses performed in community health units will have to consider spatially aggregated data, it is useful to briefly consider "what is a neighborhood?"

The driving belief behind many of these health investigations is that social variables, which are usually expressed as spatial aggregates (by census units), can have as much of an influence on health as behavioral characteristics (Caughy, O'Campo, & Muntaner, 2003; Diez-Roux, 1998; Pickett & Pearl, 2001; Reagan & Salsberry, 2005; Schwartz, 1994; Subramanian, Acevedo-Garcia, & Osypuk, 2005; Subramanian, Chen, Rehkopf, Waterman, & Krieger, 2005; Subramanian, Lochner, & Kawachi, 2003). For example, crime could be considered an expression of the neighborhood, and yet it can also lead to individual health consequences, such as elevated lead exposure, asthma, or the missing of prenatal visits (Robert, 1999;

Yen & Kaplan, 1999; Yen & Syme, 1999). In another related example, obesity may result from a lack of community exercise opportunities (such as parks), a lack of available healthy food, and even a lack of education. Therefore, is obesity an individual or a neighborhood factor? How can we extract individual choice from this equation? If results from an analysis show that a neighborhood has a high proportion of women not making an adequate number of prenatal visits, is this collective choice? If two “similar” women come from different neighborhoods, one with a bus line, and one without, a simple explanation could be the presence or absence of a bus line. But what if the difference between the women is a fear of venturing out? This is far harder to identify within the GIS. And even if we could find a variable for this fear, if a high proportion of women miss prenatal visits because of the same reason, is this an indication of a neighborhood effect, or merely that similar women live in the same area? Or is the distinction between individual and neighborhood less distinct, or at least more dynamic in nature? In another example, the presence of liquor stores is often used as an indicator of local stress (often a surrogate for poverty); while without the store the drinking would not be as high, the store is there because the demand is there.

It is obviously a very complicated relationship, and the disciplines of sociology and geography have produced volumes on the societal (spatial) shaping of lives; for example, the effect racial segregation has on poverty (Massey & Eggers, 1990; Massey, Gross, & Eggers, 1991; Subramanian et al., 2005) and how it affects the quality of living.

Including Geography in the Analysis

In order to fully realize the potential of GIS, an appreciation for the way data connects spatially is needed. At the heart of GIS is the spatial science of geography, and geographers have traditionally synthesized research from other disciplines. For example, research that might be relevant to understanding and solving health problems associated with birth outcomes could come from the disciplines of sociology, social work, anthropology, political science, racial studies, public health, medicine, epidemiology, nursing, and public health delivery. Other technical disciplines that have relevance to research involving a GIS include cartography, medical statistics, computing, and information science. The role of the geographer would be to link these different disciplines together.

Currently, relatively little has been done to find how these risks interact spatially. To think that a set of risk factors are common to all spatial locations is naïve; indeed, the risk surface is likely to vary considerably within the same city, let alone among cities. Think back again to the quote in Chapter I about infant mortality in 1912. Although these risk factors are important, how much of the real picture is lost without considering the role space plays? Although it is only scratching the surface, the GIS created for the Healthy Start Program is the first step towards creating this comprehensive approach.

This is not to say that spatial data has been missed from all previous health analyses. Indeed, three broad categories of analysis in the public health literature tend to include a neighborhood space, these being ecological studies, contextual or multilevel models, and neighborhood comparisons, which could also be considered neighborhood profiling (Diez-Roux, 2001).

Ecological studies associate health outcomes to a geographic area, which is usually politically defined. For a typical community-based health program the "holy" three levels of aggregation tend to be census tract, census block group, and zip code, though Chapter II (see Figure 3 in the chapter) displayed other possible units. For example, as discussed in Diez-Roux (2001), Rauh et al. (2001) used health areas in New York City which were comprised of four to six census tracts, whereas Pearl et al. (2001) used areas built around census blocks. Although GIS allows us to create areas or neighborhoods in any size or configuration we choose, as most social data originates from the census, census tracts and block groups are often the default units of analysis. One of the benefits of collecting individual level data, such as with the Baton Rouge Healthy Start, is that this spatial straightjacket is removed and analysis can be performed at any neighborhood definition, whether using existing neighborhood boundaries or distances from a central feature (such as a clinic).

So how do we define a neighborhood? Three common spatial expressions often used interchangeably are neighborhood, community, and area; indeed, all three have been used throughout this book. Neighborhoods can be defined historically, by demographic similarities, by political boundaries, and even by perceptions. A church congregation area, even a large housing project could be a valid "neighborhood." These definitions will vary according to the question being asked. If the question is the effect a community clinic has on those it serves, the service region of the clinic would be an appropriate neighborhood. A political boundary, such as the city boundary, may cut across a locally defined neighborhood, yet city-provided resources may only be distributed to those residing on the *city side* of the neighborhood. If the question involves an environmental threat, such as proximity to a landfill or interstate, then an even different shape may result. One of the more obvious units of analysis would be the school district, especially as so many studies either use education as a marker for deprivation, or recommend that a school-based education program be used to alleviate a particular health outcome.

A further limitation of a purely spatial or ecological analysis is that without considering the individual process it is not possible to see whether a mapped surface of infant mortality reflects variations in the individuals living in the different areas, or whether the area itself exerts an influence. As Diez-Roux (2001) asks, are individuals confounders or mediators to an area effect? In order to understand these different impacts, but at the same time combine their influence in a complete model, an approach used since the beginning of the 1990s is the contextual or multilevel model (Acevedo-Garcia, Lochner, Osypuk, & Subramanian, 2003; Diez-Roux, 1998, 2000, 2002; Diez-Roux, Link, & Northridge, 2000; Duncan, Jones, & Moon, 1996; Pickett & Pearl, 2001; Rauh et al., 2001; Subramanian, 2004; Subramanian, Kim, & Kawachi, 2002; Subramanian, Nandy, Kelly, Gordon, & Davey-Smith, 2004). These simultaneous regression models combine individual level data, such as birth certificate data, and social or neighborhood data. More often than not the individual-level data is medical, and the neighborhood data is social, either census-based or created as hybrids of data sources, such as deprivation indexes (Carstairs, 1995; Morris & Carstairs, 1991). The output from these models allows for the geographic interpretation of health events; that is, to what degree a condition (low birth weight) occurs because of a common experience.

Despite a large number of these models being found in the literature, no accepted "neighborhood" standard has been developed. Doubt must be raised as to the reduction of data to convenient political units, such as zip codes or census tracts (Diez-Roux, 2001). The question must also be raised whether the results are reflective of the neighborhood, or just the aggregation chosen. However, this question cannot completely be laid to rest until individual-level data is collected and aggregated into different configurations. Remember that census data are still summary aggregates assigned to everyone living in that spatial unit.

Holistic Neighborhood Investigations

The final research approach attempts a holistic understanding of how the neighborhood operates, with its character being the sum of several interlinked parts: political, social, and historical (Ellaway, Anderson, & Macintyre, 1997; Macintyre, MacIver, & Sooman, 1993; Phillimore & Morris, 1991). Although these investigations are usually qualitative in nature, it is possible to combine these into a GIS. For example, an ongoing study extending from the north gates of the LSU campus to downtown Baton Rouge is defining African American neighborhoods based on archeology, primary document and oral histories, literature, and sociology. The relevance of this approach to current African American problems, such as high proportions of negative birth outcomes, is that you cannot divorce modern problems from their root causes. The GIS in the project is more of an information system, with maps associated with the area, irrespective of time, being scanned and georegistered, and then used as a base layer for other information, such as written descriptions and oral histories. Although it has as not yet been investigated, it would be interesting to place current birth outcome clusters into this information system, in effect overlaying (for example) neighborhood low birth weight surfaces onto their histories.

This neighborhood profiling approach also brings up another issue, that of longevity of effect. A large number of GIS analyses are snapshots in time, or at least small aggregations of time such as a 3-year average. In order to truly understand the impact of a neighborhood, a longitudinal analysis is needed to investigate temporal stability as neighborhoods will change over time, as will their risks (Diez-Roux, 2001). This issue will again be considered in Chapter VII, along with the related concept of program participant mobility.

Spatially Synthesizing Previous Research

All of the concepts and GIS approaches discussed in this chapter are reliant on one thing: quality spatial data. Without good spatial data, the GIS is useless. In some ways this is a self-fulfilling cycle. As the spatial component of the data improves, the simplest form of which is all health data having an accurately listed residential address, the quality of the analysis will improve. As the outcomes of these analyses are implemented as public health improvements, so the appreciation for quality spatial data leads to better quality data. As the GIS becomes a more integrated component in public health, so more researchers and agencies from outside of geography will begin to incorporate the tool in their investigations. As their research develops, so they will become more familiar with the manipulations, models, errors, and visualization capabilities of spatial data. As this spatial understanding improves, so the pressure will mount for not only the improvement of existing surveillance data, but suggestions as to what other data should be collected. To this end, Chapter X comments on an interesting step forward and tangent in real-time GIS health analysis.

References

Acevedo-Garcia, D., Lochner, K. A., Osypuk, T. L., & Subramanian, S. V. (2003). Future directions in residential segregation and health research: A multilevel approach. *American Journal of Public Health, 93*(2), 215-221.

Anderson, R. T., Sorlie, P., Backlund, E., Johnson, N., & Kaplan, G. A. (1997). Mortality effects of community socioeconomic status. *Epidemiology, 8*(1), 42-47.

Anselin, L. (1992). *SpaceStat, a software program for analysis of spatial data*. University of California, Santa Barbara: National Center for Geographic Information and Analysis (NCGIA).

Anselin, L., Syabri, I., & Kho, Y. (forthcoming). GeoDa: An introduction to spatial data analysis. In the *Journal of Geographical Analysis*.

Brewer, C. (1994). Color use guidelines for mapping and visualization. In Taylor (Ed.), *Visualization in modern cartography* (pp. 123-147). Oxford, New York, Tokyo: Elsevier Science.

Campbell, J. (2000). *Map use and analysis* (4th ed.). Boston: WCB/McGraw-Hill.

Carstairs, V. (1995). Deprivation indices: Their interpretation and use in relation to health [Suppl. 2]. *Journal of Epidemiology and Community Health, 49*, S3-8.

Carstairs, V., & Morris, R. (1989). Deprivation and mortality: An alternative to social class? *Community Medicine, 11*(3), 210-219.

Caughy, M. O., O'Campo, P. J., & Muntaner, C. (2003). When being alone might be better: Neighborhood poverty, social capital, and child mental health. *Social Science and Medicine, 57*(2), 227-237.

Dent, B. D. (2002). *Cartography, thematic map design* (5th ed.). Boston: WCB/McGraw-Hill.

Diehr, P., Koepsell, T., Cheadle, A., Psaty, B. M., Wagner, E., & Curry, S. (1993). Do communities differ in health behaviors? *Journal of Clinical Epidemiology, 46*(10), 1141-1149.

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *American Journal of Public Health, 88*, 216-222.

Diez-Roux, A. V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health, 21*, 171-192.

Diez-Roux, A. (2001). Investigating neighborhood and area effects on health. *American Journal Public Health, 91*(11), 1783-1789.

Diez-Roux, A. V. (2002). A glossary for multilevel analysis. *Journal of Epidemiology and Community Health, 56*(8), 588-594.

Diez-Roux, A. V., Kiefe, C. I., Jacobs, D. R., Jr., Haan, M., Jackson, S. A., Nieto, F. J., Paton, C. C., Schulz, R., & Roux, A. V. (2001). Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. *Annals of Epidemiology, 11*(6), 395-405.

Diez-Roux, A. V., Link, B. G., & Northridge, M. E. (2000). A multilevel analysis of income inequality and cardiovascular disease risk factors. *Social Science and Medicine, 50*(5), 673-687.

Duncan, C., Jones, K., & Moon, G. (1996). Health-related behavior in context: A multilevel approach. *Social Science and Medicine, 42*, 817-830.

Duncan, C., Jones, K., & Moon, G. (1999). Smoking and deprivation: Are there neighbourhood effects? *Social Science and Medicine, 48*(4), 497-505.

Ellaway, A., Anderson, A., & Macintyre, S. (1997). Does area of residence affect body size and shape? *International Journal of Obesity and Related Metabolic Disorders, 21*(4), 304-308.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. E.. (1998). Geographically weighted regression: A natural evolution and expansion method for spatial data analysis. *Environment and Planning A, 30*, 1905-1927.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. E. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester, UK: Wiley.

Gilmartin, P., & Shelton, E. (1989). Choropleth maps on high resolution CRTs: The effects of number of classes and hue on communication. *Cartographica, 26*, 40-52.

Harris, R. (2003). Population mapping by geodemographics and digital imagery. In V. Mesev (Ed.), *Remotely-sensed cities* (pp. 223-241). London: Taylor and Francis.

Harris, R. J., & Frost, M. (in press). Indicators of urban deprivation for policy analysis GIS: Going beyond wards. In D. Kidner, G. Higgs, & S. White (Eds.), *Socioeconomic applications of geographic information science innovations in GIS 9*. London: Taylor and Francis.

Hodgson, M. J. (1988). A hierarchical location-allocation model for primary health care delivery in a developing area. *Social Science and Medicine, 26*(1), 153-161.

Humphreys, K., & Carr-Hill, R. (1991). Area variations in health outcomes: Artifact or ecology. *International Journal of Epidemiology, 20*(1), 251-258.

Kleinschmidt, I., Hills, M., & Elliott, P. (1995). Smoking behaviour can be predicted by neighbourhood deprivation measures [Suppl. 2]. *Journal of Epidemiology and Community Health, 49*, S72-77.

Krieger, N. (1992). Overcoming the absence of socioeconomic data in medical records: Validation and application of a census-based methodology. *American Journal of Public Health, 82*(5), 703-710.

Kulldorff, M., Athas, W. F., Feuer, E. J., Miller, B. A., & Key, C. R. (1998). Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos. *American Journal of Public Health, 88*, 1377-1380.

Kulldorff, M., Feuer, E. J., Miller, B. A., & Freedman, L. S. (1997). Breast cancer clusters in the northeastern United States: A geographical analysis. *American Journal of Epidemiology, 146*, 161-170.

LeClere, F. B., Rogers, R. G., & Peters, K. (1998). Neighborhood social context and racial differences in women's heart disease mortality. *Journal of Health and Social Behavior, 39*(2), 91-107.

Levine, N. (2004, May). CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (Version 3.0). Houston, TX: Ned Levine & Associates; Washington, DC: National Institute of Justice.

Longley, P. A., & Mesev, T. V. (2002). Measurement of density gradients and space-filling in urban systems. *Papers in Regional Science, 81*, 1-28.

Macintyre, S., MacIver, S., & Sooman, A. (1993). Area, class, and health: Should we be focusing on places or people? *Journal of Social Policy, 22*, 213-234.

Massey, D., & Eggers, M. (1990). The ecology of inequality: Minorities and the concentration of poverty, 1970-1980. *American Journal of Sociology, 95*, 1153-1188.

Massey, D., Gross, A., & Eggers, M. (1991). Segregation, the concentration of poverty, and the life chances of individuals. *Social Science Research, 20*, 397-420.

McCleary, G. F. (1969). *The dasymetric method in thematic cartography*. Unpublished doctoral dissertation, University of Wisconsin-Madison.

McLafferty, S., & Broe, D. (1990). Patient outcomes and regional planning of coronary care services: A location-allocation approach. *Social Science and Medicine, 30*(3), 297-304.

Mohan, J. (1983). Location-allocation models, social science and health service planning: An example from North East England. *Social Science and Medicine, 17*(8), 493-499.

Morris, R., & Carstairs, V. (1991). Which deprivation? A comparison of selected deprivation indexes. *Journal of Public Health and Medicine, 13*(4), 318-326.

Muehrcke, P. C., Muehrcke, J. O., & Kimerling, A. J. (2001). *Map use* (4th ed.). Madison, WI: JP Publications.

O'Campo, P., Xue, X., Wang, M., & Caughy, M. (1997). Neighborhood risk factors for low birthweight in Baltimore: A multilevel analysis. *American Journal of Public Health, 87*, 1113-1118.

O'Sullivan, D., & Unwin, D. (2002). *Geographic information analysis*. Hoboken, NJ: Wiley.

Pearl, M., Braveman, P., & Abrams, B. (2001). The relationship of neighborhood socioeconomic characteristics to birthweight among 5 ethnic groups in California. *American Journal of Public Health, 91*(11), 1808-1814.

Phillimore, P. R., & Morris, D. (1991). Discrepant legacies: Premature mortality in two industrial towns. *Social Science and Medicine, 33*(2), 139-152.

Pickett, K., & Pearl, M. (2001). Multilevel analyses of neighborhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health, 55*, 111-122.

Rauh, V. A., Andrews, H. F., & Garfinkel, R. S. (2001). The contribution of maternal age to racial disparities in birthweight: A multilevel perspective. *American Journal of Public Health, 91*(11), 1815-1824.

Reagan, P. B., & Salsberry, P. J. (2005). Race and ethnic differences in determinants of preterm birth in the USA: Broadening the social context. *Social Science and Medicine, 60*(10), 2217-2228.

Robert, S. A. (1999). Socioeconomic position and health: The independent contribution of community socioeconomic context. *Annual Review of Sociology, 25*, 489-516.

Roberts, E. M. (1997). Neighborhood social environments and the distribution of low birth weight in Chicago. *American Journal of Public Health, 87*, 597-603.

Robinson, A. H., Morrison, J. L., Muehrcke, P.C., Kimerling, A. J., & Guptill, S. C. (1995). *Elements of cartography* (6th ed.). New York: John Wiley and Sons.

Rushton, G., & Lolonis, P. (1996). Exploratory spatial analysis of birth defect rates in an urban population. *Statistics in Medicine, 15*, 717-726.

Scarpaci, J. L. (1984). A methodological note on location-allocation models. *American Journal of Public Health, 74*(10), 1155-1157.

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.

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2004). Thematic cartography and geographic visualization. In *Thematic cartography and geographic visualization* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Subramanian, S. V. (2004). The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Social Science and Medicine, 58*(10), 1961-1967.

Subramanian, S. V., Acevedo-Garcia, D., & Osypuk, T. L. (2005). Racial residential segregation and geographic heterogeneity in black/white disparity in poor self-rated health in the US: A multilevel statistical analysis. *Social Science and Medicine, 60*(8), 1667-1679.

Subramanian, S. V., Chen, J. T., Rehkopf, D. H., Waterman, P. D., & Krieger, N. (2005). Racial disparities in context: A multilevel analysis of neighborhood variations in poverty and excess mortality among black populations in Massachusetts. *American Journal of Public Health, 95*(2), 260-265.

Subramanian, S. V., Kim, D. J., & Kawachi, I. (2002). Social trust and self-rated health in US communities: A multilevel analysis [Suppl. 1]. *Journal of Urban Health, 79*(4), S21-34.

Subramanian, S. V., Lochner, K. A., & Kawachi, I. (2003). Neighborhood differences in social capital: A compositional artifact or a contextual construct? *Health Place, 9*(1), 33-44.

Subramanian, S. V., Nandy, S., Kelly, M., Gordon, D., & Davey-Smith, G. (2004). Patterns and distribution of tobacco consumption in India: Cross sectional multilevel evidence from the 1998-9 national family health survey. *British Medical Journal, 328*(7443), 801-806.

U.S. Department of Health and Human Services, Health Resources and Services Administration, Bureau, M. a. C. H., & Health, D. o. P. S. a. W. s. (2004). *Eliminating Disparities in Perinatal Health (General Population) — HRSA-05-013 Program Guidance*. Retrieved November 9, 2004, from http://www.hrsa.gov/grants/preview/guidancemch/hrsa05013-05037.htm

Yen, I. H., & Kaplan, G. A. (1999). Poverty area residence and changes in depression and perceived health status: Evidence from the Alameda County Study. *International Journal of Epidemiology, 28*(1), 90-94.

Yen, I. H., & Syme, S. L. (1999). The social environment and health: A discussion of the epidemiologic literature. *Annual Review of Public Health, 20*, 287-308.

Chapter IV

The Geography of Health Risks

This book has so far provided an introduction to GIS in terms of its use as part of a community health program. Subsequent chapters will describe a selection of more detailed GIS techniques and approaches. This chapter can be considered as an interlude, an attempt to set the scene by painting a backcloth of risks facing many city neighborhoods and the pregnant women living in them. Of course, at the risk of repeating oneself, there have been whole books written about single risk factors, so the task of compressing all risks into a single chapter is extremely difficult. This is made even harder because many of the situations described are not directly related to pregnancy outcomes, but instead are classed as neighborhood risks, contributing to the overall vulnerability of a person living in these environments. Although some may question why not concentrate solely on pregnancy-related risks, a more holistic understanding of the social environment can help place context into data, a movement away from the earlier criticized "structural functionalism" approach (Litva & Eyles, 1995). At this point it might be worth briefly mentioning that debate continues within the field of medical geography as to the degree in which pure analytical approaches ignore the social relevance of actions (a political economy approach), or how an individual's experience shapes his or her actions (Dorn & Laws, 1994). For a review of these critical literatures in association with infant mortality see Gesler, Bird, and Oljeski, (1997).

Although this is not the forum to extend this debate, I am sympathetic to many of the points raised. There is a danger of overreliance on a single variable within a GIS analysis — indeed, we fall victim to that approach throughout this book. What must happen eventually is to combine both quantitative and qualitative approaches as we move through a general methodology of investigating health related problems, which includes identifying spatial patterns of the problem, finding which variables are associated with that pattern, and understanding how the process works, whether that is a biological or social pathway to an outcome.

That final stage is a complex undertaking, as vulnerability is a function of individual, behavioral, social, neighborhood, and environmental risks, many of which are inextricably linked and result in a negative health outcome (Anderson, 1952). If the reader remembers back to Chapter I, a benefit of geography is that it can act as an overview of disciplines, bringing together disparate data from multiple sources through the connection of space.

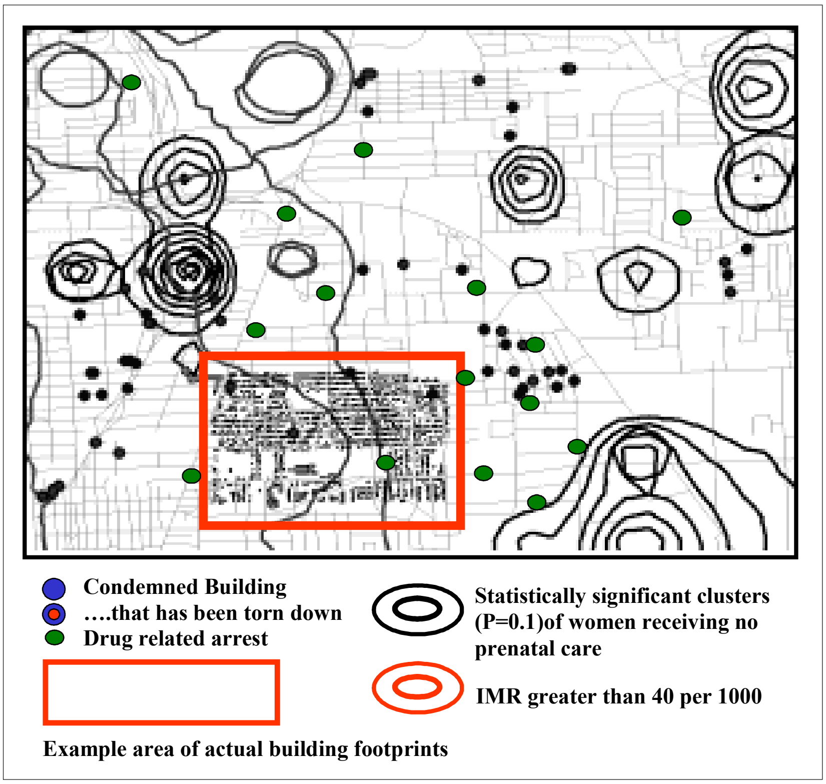
Some risks are common to a social cohort, a good example being smoking. The perils of smoking are well documented, not only in terms of what tobacco can do to a fetus, but in general individual and societal terms (Committee on Substance Abuse, 2001). Just one cigarette a day can increase the progression of atherosclerosis, while nicotine can cause the narrowing of arteries (American Heart Association, 2004). Health problems are not just restricted to the smoker, for example Gergen, Fowler, Maurer, Davis, and Overpeck (1998) found that environmental tobacco smoke (secondhand smoking) exposure increased the prevalence of asthma, wheezing, and chronic bronchitis in children aged 3 to 5 years. It therefore makes sense to stop smoking, but how is this achieved? Is it simply a matter of choice? Are publicly funded smoking cessation programs available? Is smoking actually a coping mechanism for other stresses in the neighborhood, such as violent crime? Is it a question of education; are these women not warned about the dangers of smoking, and similarly are children adequately educated about the risks in school? Even if the program participant is persuaded to stop smoking, what about secondhand smoke in the household? Many of these conditions are likely to show up as patterns on a map because they have a spatial setting; these outcomes are part of what could be called a spatial cohort. Also, several factors influencing the decision to smoke will be geographic: the state and its smoking cessation programs, the clinic and its pregnancy advice, the middle school and its classroom instruction, or the number (and location) of drug offences or violent crimes (if smoking is a coping mechanism).

Chapter II discussed the movement in public health to ecological studies which recognize that place can be as important as individual factors in explaining an outcome (Diez-Roux, 2001), and yet the use of GIS in understanding the geographic processes at work is still somewhat lacking. GIS allows us to bring these multiple layers of risk together as we search for the spatial connections of causation. For example, crime hot spots, such as drug arrests, can be overlaid onto prenatal risk hot spots. The crime patterns could be used to identify neighborhoods of chronic stress where smoking or alcohol use is likely to be higher. An environmental risk layer can also be added, such as pre-1950 housing stock, a traditional warning factor for lead exposure. If we are interested in understanding the risk of stress and its health manifestations (such as depression during pregnancy), then women screened for mental health problems can also be overlaid in order to validate the problem neighborhoods.

Figure 1 displays how the GIS could be used to visualize multiple risk surfaces. Hot spots of women receiving no prenatal care and an infant mortality rate (IMR) above 40 per 1,000 are shown as contour lines. The other spatial layers include drug related arrests, and evidence of urban blight (those buildings which had been condemned by the city, and those which had already been torn down).

This chapter categorizes risks for the ease of presentation, but at the same time it generally follows the disease ecology approach of creating a comprehensive model involving biological, social, economic, behavioral, and environmental factors (Mayer & Meade, 1994; Meade & Earickson, 2000). It should be remembered, though, that these categories blend into one another, with many being potential contributors to each other. For example, certain negative birth outcomes are considered risks for other negative birth outcomes. It has been

Figure 1. Various "risks" for a neighborhood in Baton Rouge



shown that low-birth-weight babies have a higher probability of becoming an infant death (Iyasu, Becerra, Rowley, & Hougue, 1992). Similarly, preterm deliveries are a leading cause of perinatal mortality (Berkowitz & Papiernik, 1993) and neurodevelopment problems. Therefore, a risk of a low-birth-weight delivery is also indirectly a risk for an infant death, and a risk for a preterm delivery is a secondary risk for a low-birth-weight delivery. Similarly, substance abuse during pregnancy is also associated with low birth weight. However, both substance abuse and low birth weight may be caused at least in part by other social conditions in the neighborhood, such as stress (Sheehan, 1998). Wilcox (2001) suggests that the link between low birth weight and infant mortality may not be causative but both are symptomatic of other (social) factors. It is therefore important to consider these risks not in isolation but as components of a complex birth situation involving both individual and neighborhood influences, including access to health care (Gessner & Muth, 2001).

Although many of the risks identified in this chapter are presented aspatially, which means that connections are made between a birth outcome and a causative factor irrespective of location, the reader should always be thinking of how space enters into the relationship. To believe that a set of risk factors are common to all spatial locations is naïve (Macintyre, MacIver, & Sooman, 1993; Sooman & Macintyre, 1995). Indeed, the risk surface will vary spatially within cohorts of the same city, let alone between cities. For example, this very point has been made in reference to air pollution exposure (American Heart Association, 2004).

Of course, this does not mean all investigations have been aspatial — think back to the discussions on ecological studies and neighborhood definition in Chapter II. There are spatial and temporal patterns to prenatal risk (Bode, O'Shea, Metzguer, & Stiles, 2001; Crosse, Alder, Ostbye, & Campbell, 1997). Although these spatial patterns will occur at least partly due to variations in neighborhood impact (Fang, Madhavan, & Alderman, 1999; O'Campo, Xue, Wang, & Caughy, 1997; Pearl, Braveman, & Abrams, 2001; Roberts, 1997), relatively little has been done to consider spatial patterns explicitly (Chapter V will discuss some examples, especially involving the spatial filter methodology).

The risks discussed in this chapter will be presented in two ways: first, as negative birth outcome specific risks, focusing on the big three of infant mortality, low birth weight, and short-gestation deliveries. This section will be followed by a more holistic approach to risks using a six-category (disease ecology) determination, including medical risks, such as evidence of a previous low-birth-weight birth and being HIV positive; behavioral risks, such as substance abuse, smoking, alcohol, and diet; cohort or social risks, including race, age, and socio-economic status; neighborhood risks, such as clinic coverage and provider performance in the role of health care delivery; and environmental risks, such as proximity to a chemical works, land fills, polluted water sources, or high concentrations of toxins in the atmosphere.

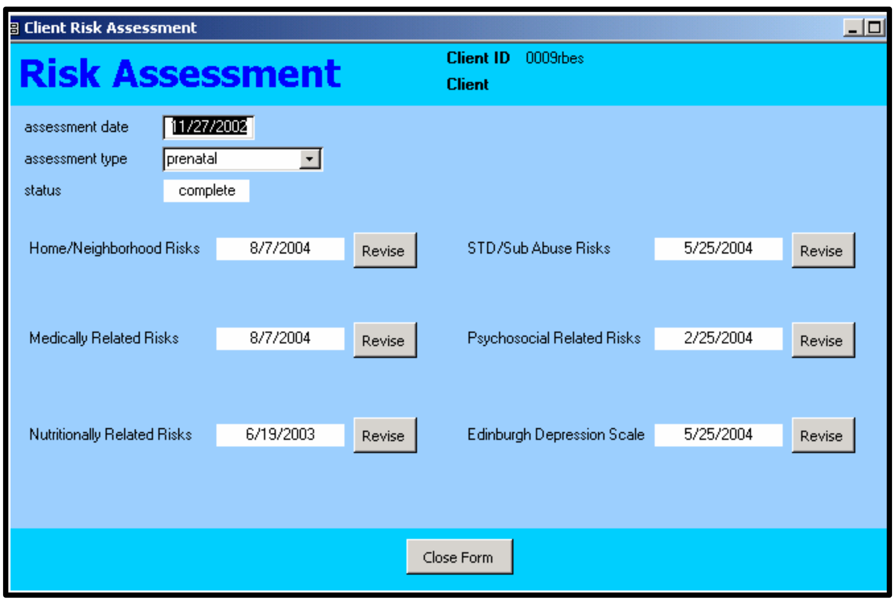
Of course, these categories are largely artificial and will definitely generate arguments as to where a certain risk should be listed. For example, lead exposure will be linked to crime patterns, which in turn will lead to stress and coping outcomes — where should these risks be placed? Similarly, where should domestic risks, which would include the living environment, such as overcrowding, inadequate amenities, and domestic abuse be placed? The point is to make these interconnections and not get too attached to category labels. In addition, many of the described risks are not traditionally linked to pregnancy outcomes, though their inclusion is because of possible indirect effects (especially through stress generation), and may even have direct yet undiscovered associations.

Figure 2 displays the initial risk assessment form for the caseworkers at the Baton Rouge Healthy Start.

Infant Deaths, Low Birth Weight, and Short Gestation Deliveries

Before we introduce each of the general risk categories, it is useful to show how three of the most discussed negative birth outcomes — infant mortality, low birth weight, and preterm deliveries — are interconnected, both in terms of one outcome leading to another, and in sharing what are believed to be the same risk

Figure 2. Caseworker generalized risk categories



causations. For example, in 1985 low birth weight (defined as a baby at birth weighing less than 2,500 grams) was identified by the Committee to Study the Prevention of Low Birth as the most important factor associated with perinatal mortality in the United States. Similar results have been found in Luke, Williams, Minogue, and Keith (1993), McCormick (1983), Paneth (1995), Wilcox and Skjoerven (1992), and Witter (1993). Therefore, risks associated with low birth weight are indirectly associated with infant mortality. Risks associated with a low-birth-weight delivery fall into two general groups: Either there is inadequate intrauterine growth, or the baby is a preterm (36 weeks or less) or very preterm delivery (32 weeks or less). Of these two general causations, further risks associated with intrauterine growth include smoking, which explains about 20% to 30% of all cases (Chomitz, Cheung, & Lieberman, 1995), low maternal weight gain and low prepregnancy weight (Alexander & Korenbrot, 1995).

Now let us consider risks associated with preterm deliveries which have also been associated with infant mortality (Bohin, Draper, & Field, 1999; Davis & Sandman, 1998; Draper, Manktelow, Field, & James, 1999; Emery, Eaton, Grether, & Nelson, 1997; Field & Draper, 1999; Hoffman & Bakketeig, 1984; Lumley, 1993; Mercer et al., 1999; Tin, Wariyar, & Hey, 1997a, 1997b). There has been considerable research into causations of preterm deliveries (Adams, Berg, Rhodes, & McCarthy, 1991; Adams, Delaney, Stupp, McCarthy, & Rawlings, 1997; Collins & David, 1990, 1993; Kiely & Kleinman, 1993; Kleinman & Kessel, 1987; Kleinman & Madans, 1985; Mangold & Powell-Griner, 1991; McGrady, Sung, Rowley, & Hogue, 1992; Rich-Edwards et al., 2001; Rowley, 1994; Shiono & Klebanoff, 1986; Shiono, Rauh, Park, Lederman, & Zuskar, 1997), with biological causations including maternal or fetal distress (Wadhwa, Culhane, Rauh, & Barve, 2001; Wadhwa, Culhane, Rauh, Barve et al., 2001), and infections (Madianos et al., 2001; Offenbacher et al., 2001). However, specific connection between these risks and a short-gestation birth remain unclear (Goldenberg & Rouse, 1998; Lee & Silver, 2001; Tucker et al., 1991), with biological and behavioral factors such as violence (Covington, Hage, Hall, & Mathis, 2001) not providing adequate answers (Krieger, Rowley, Herman, Avery, & Phillips, 1993). This has led to a shift in emphasis in the role “place” plays (Fang et al., 1999; Krieger, 1991; O’Campo et al., 1997; Pearl, Braveman, & Abrams, 2001; Pickett & Pearl, 2001; Roberts, 1997).

Similarly and unsurprisingly, given the strong association between low birth weight and short gestation, one neighborhood characteristic, that of social deprivation, has been estimated to be associated with 30% of low-birth-weight deliveries (Pattenden, Dolk, & Vrijheid, 1999). Typical neighborhood variables resulting in this deprivation measure include community economic hardship, housing costs, and unemployment (Roberts, 1997). This relationship is hardly unexpected, as having fewer resources will mean fewer health options, fewer choices of services and even day-to-day necessities (such as nutritious food). In addition, community level economic hardship could mean fewer options, even if the program participant has disposable income, such as few stores selling fresh produce, fewer drug stores to fill prescriptions, and fewer community level resources that can also improve quality of life and help relieve stress (ranging from a safe park or playground, to just a neighborhood “feel”).

Finally, living in an impoverished neighborhood can result in an increased number of environmental hazards, partly through these neighborhoods being disproportionately located close to polluting facilities (a question of environmental injustice), and partly through these populations not having the financial resources to deal with the consequences of the exposure (an issue of social justice). It is a complex situation, with risks and outcomes all being linked together. To summarize, short-gestation births may be causative of low-birth-weight deliveries, which in turn may result in an infant death. Therefore, causative factors for any of these three outcomes may be directly or indirectly linked to the others. For example, a pregnant woman who uses drugs during pregnancy may have a full-term baby which then develops complications associated with that substance. The pregnant woman may also deliver a preterm baby because of the substance use, the preterm delivery eventually resulting in death due to a preterm-related cause such as the underdevelopment of the immune system or a brain hemorrhage (BBC News, 2002). The remaining sections will consider these causative factors.

Medical Risks

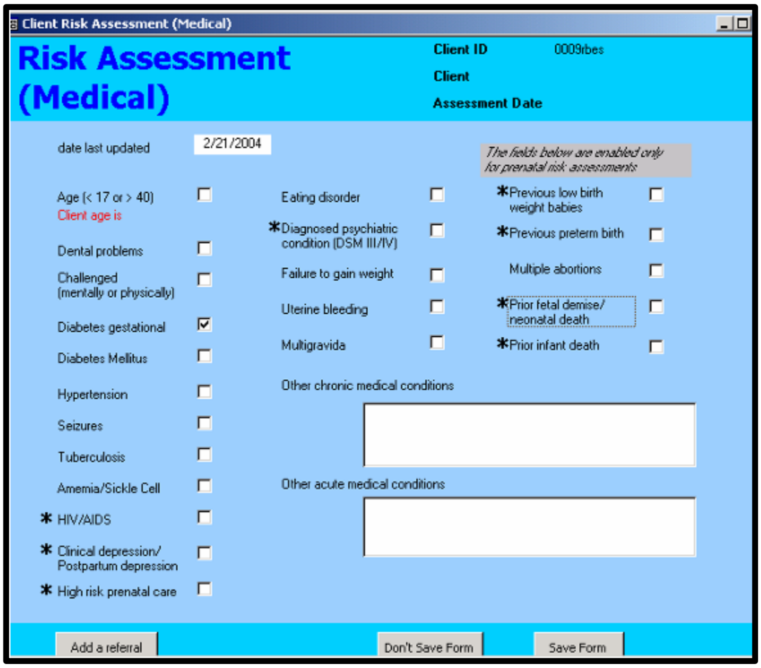
Medical risks linked to birth outcomes — for example, infections leading to a low-birth-weight delivery (Cotch et al., 1997; Goldenberg et al., 1996; Hillier, Krohn et al., 1995; Hillier, Nugent et al., 1995; Joesoef et al., 1995; Krohn, Hillier, Lee, Rabe, & Eschenbach, 1991; Watts, Krohn, Hillier, & Eschenbach, 1992; Watts et al., 1993) — are difficult to capture and analyze within a GIS environment, as they are usually individual in nature. However, spatial patterns may exist if an outside factor either exaggerates or acts as a catalyst with a medical condition. For example, women with recurring bacterial vaginosis (BV) may actually increase the likelihood of an episode if they smoke, BV being a risk for a preterm delivery (Goldenberg et al., 1996; Hauth, Goldenberg, Andrews, DuBard, & Copper, 1995; Hillier, Krohn et al., 1995; Hillier, Nugent et al., 1995; Joesoef et al., 1995; Offenbacher et al., 2001; Wadhwa, Culhane, Rauh, Barve et al., 2001). Further, African American women are more likely to suffer recurring BV (Goldenberg et al., 1996; Hillier, Nugent et al., 1995). Therefore, two spatial patterns may exist for BV: neighborhoods with high levels of smoking and/or high proportions of African Americans.

It is well documented how different substance uses during pregnancy can lead to negative birth outcomes, some of which are behavioral and developmental problems, while others manifest in tangible medical conditions. For example, alcohol use during pregnancy can result in a variety of negative outcomes, including Fetal Alcohol Syndrome (FAS) which can result in pre- and post-growth retardation, facial anomalies, microcephaly, and mental retardation; alcohol-related neurodevelopmental disorders (ARND), which can result in learning and control disabilities at school, including hyperactivity; alcohol-related birth defects (ARBD), which can result in physical defects, such as heart, kidney and hearing problems; and an elevated risk of spontaneous abortion (Chomitz et al., 1995). Kesmodel and Olsen (2001) found that women having five or more drinks per week during the pregnancy were more than eight times as likely to suffer a stillbirth as a result of feto-placental dysfunction. Illegal drug use also results in tangible medical outcomes, such as placental abruption and premature rupture of membranes, both of which have been statistically linked to cocaine use (Addis, Moretti, Ahmed Syed, Einarson, & Koren, 2001). Other outcomes such as malformations, skull development problems, low birth weight, and prematurity have been linked to cocaine use but could not be extracted from other confounding substance uses, which are often concurrently present in a cocaine-using mother-to-be.

The impact stress can have on a birth will be discussed later in this chapter, though it is worth mentioning how knowledge of a medical condition can also lead to anxiety. Studies have generally been inconclusive in terms of establishing a connection between anxiety and birth outcomes (Copper et al., 1996; Peacock, Bland, & Anderson, 1995; Rini, Dunkel-Schetter, Wadhwa, & Sandman, 1999). However, if a mother is aware of a medical situation, either through clinical or self-diagnosis, then it is likely to cause her anxiety. For example, what is the mental state of a woman with HIV who knows she is pregnant? Unfortunately, it is hard to establish what added negative impact this anxiety has on the birth because of “confounding by indication” (Salas, Hofman, & Stricker, 1999). Is it the anxiety or the medical risk itself that causes the problem?

A further advantage GIS use can have when dealing with medical risks is as an information system. GIS can be used to track a program participant’s pregnancy, making sure she follows a medication regime and keeps clinic visits. In addition, GIS, through the analysis of previous birth records, can also be used to “tag” current pregnancies if a previous risk also has implications for the present; for example, smoking in a previous pregnancy is also a predictor of risk for future pregnancies (England, Kendrick, Gargiullo, Zahniser, & Hannon, 2001; England, Kendrick, Wilson et al., 2001). Other risks can be treated similarly; for example, it has also been found that a women who has had a previous low-birth-weight birth is at greater risk of a subsequent low-birth-weight birth (Bakketeig, Hoffman, & Harley, 1979). A previous preterm delivery can also predict future low-birth-weight deliveries (Adams, Elam-Evans, Wilson, & Gilbertz, 2000;

Figure 3. Medical risks



Cnattingius, Granath, Petersson, & Harlow, 1999; Winkvist, Mogren, & Hogberg, 1998), particularly for African American teenagers (Blankson, Cliver, Goldenberg, Hickey, & Dubard, 1993). Similarly, the degree of previous preterm delivery is a predictor of subsequent risk (Bakewell, Stockbauer, & Schramm, 1997; Bakketeig et al., 1979). Even a previous birth interruption, whether planned or spontaneous, can have a residual effect. All of these data are captured on the birth certificate, and it would be relatively easy for a decision support system to be created to check any new program participant entering a community health program using previous birth records as input data. The decision support system would raise "risk" flags if any previous negative outcome were detected.

Figure 3 Although *Medical Risks* are found on several forms in the Baton Rouge Healthy Start GIS, this figure provides a good indication of how a variety of medical and *previous birth experience* risks are captured.

Behavioral Risks

Behavioral risks are sometimes defined as poor "choices" made during pregnancy. For example, the risks associated with perinatal mortality include substance abuse (Chasnoff, 1991), smoking (Bartal, 2001; Wisborg, Kesmodel, Henriksen, Olsen, & Secher, 2001), alcohol (Kesmodel, Wisborg, Olsen, Henriksen, & Secher, 2002), caffeine (Brooke, Anderson, Bland, Peacock, & Stewart, 1989; Wisborg, Kesmodel, Bech, Hedegaard, & Henriksen, 2003), and maternal nutrition (Kramer, McLean, Eason, & Usher, 1992). Among "choices" associated with preterm or low-birth-weight deliveries are smoking (Jaakola, Jaakkola, & Zahlsen, 2001; Shiono, Klebanoff, & Rhoads, 1986), young maternal age (Ancel, Saurel-Cubizolles, Di Renzo, Papiernik, & Breart, 1999), socioeconomic status of women (Ancel et al., 1999; Berkowitz & Papiernik, 1993; Luginaah, Lee, Abernathy, Sheehan, & Webster, 1999), low pregnancy weight-gain (Schieve, Cogswell, et al., 2000), and short periods between pregnancy (Adams et al., 1997; Brody & Bracken, 1987; Ferraz, Gray, Fleming, & Maia, 1988; Lang, Lieberman, Ryan, & Monson, 1990; Shults, Arndt, Olshan, Martin, & Royce, 1999). Many of these "choices" also display as patterns on a map due to similar cohorts, or neighborhoods experiencing similar lifestyles and activities. A more specific spatial reason could also mean a lack of appropriate education in terms of the harmful effect smoking or drinking can have on a fetus (these education initiatives being based, for example, at a school-based health center). Access to courses designed to wean mothers away from risk substances, such as smoking cessation classes, could also play a part (in general, the further the distance to such a course, the less impact the course has). Maternal nutrition can also display spatial patterns not only in terms of similar cohorts, but also due to a lack of education (for example, the importance of folic acid in the diet).

However, to think that these risks are solely one of choice is both naïve and limiting in terms of developing intervention strategies. For example, smoking is bad, end of story, both in terms of the general population and especially during pregnancy (Ahern, Pickett, Selvin, & Abrams, 2003), resulting in twice the likelihood of a low-birth-weight delivery, an average reduction of 200g in birth weight (Walsh, Lowe, & Hopkins, 2001), and twice the likelihood of a stillbirth (Kesmodel et al., 2002). Smoking both during the pregnancy and in the postpartum period also leads to an increased likelihood of infantile colic (Sondergaard, Henriksen, Obel, & Wisborg, 2001). Yet some estimates have smoking during pregnancy for at-risk cohorts being as high as 10%-20% (Mathews, 1998, 2001). So the obvious answer is to target initiatives at city areas where pregnant women choose to smoke. Klesges, Johnson, Ward, and Barnard (2001) acknowledge that socially disadvantaged pregnant women could be reached by an office-based intervention by doctors. Walsh et al. (2001) note that at least a 50% reduction in smoking is required to make any change for the better on the pregnancy outcome, or as England, Kendrick, Wilson, et al. (2001)found, less than eight cigarettes per day, while Wisborg et al. (2001) found that quitting smoking by the sixteenth week of the pregnancy could result in a drop of 25% of all stillbirths and 20% of infant deaths. However, although cessation schemes are obviously necessary, the reasons for smoking are often deep-rooted and beyond simple interventions. Smoking may be a coping mechanism for an individual faced with multiple chronic stresses, such as neighborhood violence. Besides, all family members are likely to suffer the same stresses, and use similar coping strategies. Therefore, any smoking cessation strategy aimed solely at the mother will only provide a partial solution. At the end of this chapter, several "risk" maps are displayed for the Baton Rouge Healthy Start area (Figure 9). Notice how many more secondhand smoking residences there are than smoking mothers, though obviously there will always be an underreporting of self-smoking (Ford, Tappin, Schluter, & Wild, 1997).

Although it has been found that African Americans are less likely to smoke than whites during adolescence (Epstein, Botvin, & Diaz, 1998; Nelson, Giovino, et al., 1995) and tend to start smoking later (Griesler & Kandel, 1998; Griesler, Kandel, & Davies, 2002; Headen, Bauman, Deane, & Koch, 1991), with a potential explanation being the presence of a strong family and religious influence (Catalano et al., 1992; Clark, Scarisbrick-Hauser, Gautam, & Wirk, 1999), these are probably not associations found in the poorest neighborhoods with the worst birth outcomes. These are the neighborhoods most likely to experience the greatest stresses, and have the weakest family and religious support structures. It is for this reason that the actual risk associated with smoking appears to be greater for African Americans than whites (Ekwo & Moawad, 2000; Rolett & Kiely, 2000; Shah & Bracken, 2000), because in those neighborhoods where smoking is prevalent, it is caused and compounded by multiple other negative neighborhood experiences. It is a vicious cycle: A woman already facing a higher stress level because of her local environment smokes to cope, which increases the chance of a potential negative birth outcome. And of course there are many other stresses that can lead to her coping through smoking, arguably the worst of which is domestic violence (McCauley, Kern, Kolodner, Derogatis, & Bass, 1998; McCauley et al., 1995; Tilden et al., 1994). It is far too simplistic to think that smoking is a simple matter of choice.

Tobacco use during pregnancy is only one harmful substance screened for during prenatal visits. Alcohol and illegal drug use are also substances of concern. Kesmodel et al. (2002) found that women were almost three times as likely of having a stillbirth if they consumed five or more drinks per week during the pregnancy. The National Pregnancy and Health Survey found that of 4 million women who gave birth between 1992 and 1993, 757,000 used alcohol, 820,000 smoked tobacco, and 221,000 used illegal drugs during pregnancy (National Institute on Drug Abuse [NIDA] Info Facts, 1997). As previously mentioned, the rate of alcohol and smoking were higher for whites than for African Americans, though this was reversed for illegal drugs.

Of illegal drugs, marijuana (119,000 women) and cocaine (45,000 women) were the most frequently used. Women under 25 were more likely to use marijuana, while those over 25 used cocaine. A general profile of a pregnant woman most likely to use illegal drugs is a single mother, who is unemployed, and has less than 16 years of formal education. This profile fits a large portion of the Baton Rouge Healthy Start population.

Other apparent individual choices can also have external influences that would allow for spatial risk maps to be created. For example, not making enough prenatal visits (the most common minimum threshold is seven visits) or starting these visits at a late stage of the pregnancy (during the second trimester or later) are widely acknowledged as risks for a poor birth outcome (Alexander & Korenbrot, 1995; Kiely & Kleinman, 1993). Violence in the surrounding neighborhood may act as a deterrent to leaving the house. The mother-to-be may not have family or friends who could care for her other children during the visit. She may not have access to transportation to make the visit. She may not have the financial support necessary to take time off work to make the visit.

Other risks such as being single (Pattenden et al., 1999), having a child at an early age (Geronimus, 1986), and drinking during pregnancy (Verkerk, 1998; Verkerk, van Noord-Zaadstra, Florey, de Jonge, & Verloove-Vanhorick, 1993) have also been found in Baton Rouge to create identifiable map patterns, meaning that there are background factors that are influencing these individual choices.

It is also naïve to think that interventions will be totally effective by reaching a woman just during her pregnancy. Obviously, the incentive to change individual behavior is greatest at this point, but many choices have developed over the course of a lifetime. Indeed, many of these choices will have an impact during the pregnancy irrespective of intervention success. For example, associations have been found between coronary disease and elevated levels of neonatal and postneonatal mortality (Stein et al., 1996). I am sure the reader is well aware of the popular reports (and captured in documentaries such as *Super Size Me* [Spurlock, 2004]), detailing how we are, as a society, moving toward obesity.

Nutrition advice can certainly be beneficial; for example, the Special Supplemental Nutrition Program for Women, Infants and Children (WIC) has traditionally been associated with improving birth outcomes due in part to its focus on improving eating habits and providing referral to medical services (Ahluwalia, Hogan, Grummer-Strawn, Colville, & Peterson, 1998; Buescher & Horton, 2001; Buescher et al., 2003; Buescher, Larson, Nelson, & Lenihan, 1993; Heimendinger, Laird, Austin, Timmer, & Gershoff, 1984; E. T. Kennedy & Kotelchuck, 1984; Kotelchuck, Schwartz, Anderka, & Finison, 1984; Moss & Carver, 1998). It is also important to stress the importance of folic acid in the diet. Similarly, breast-feeding improves infant outcomes, ranging from the biological advantage of shared immunity to the development of a strong mother-child bond (Beaudry, Dufour, & Marcoux, 1995; Howie, Forsyth, Ogston, Clark, & Florey, 1990; Shu et al., 1999), hence the 2004 U.S. Department of Health and Human Services advertisement that “Babies were born to be breastfed.”

However, in order for a community health program to be effective, women need to be targeted with these messages *before* they become pregnant. This is not easy, because even dietary habits, the most obvious expression of which may be obesity, may be more a reflection of the neighborhood or spatial cohort than just an individual choice. In a violent neighborhood, children may be kept inside, resulting in enforced sedentary lifestyle. The crime rate may also limit local shopping trips and store (and product) availability. And so these children are more predisposed to obesity and its associated health outcomes, such as asthma (Camargo, Weiss, Zhang, Willett, & Speizer, 1999; Stenius-Aarniala et al., 2000). Any education strategy designed to address the problem of diet must be aware of these other contributing factors in the neighborhood.

Many behavioral risks are captured on the Baton Rouge Healthy Start GIS caseworker forms. Figure 4 provides an example focused on sexually transmitted disease and substance use. Notice the grayed-out areas of the form — these will only become active if the program participant reports to her caseworker that she is indeed a smoker or uses drugs. Although these fields are captured on the birth certificate, it is extremely common for these data to be underreported. Underreporting may also occur for Healthy Start program participants, though a close working relationship between the program participant and caseworker will hopefully generate the degree of trust needed for more accurate responses.

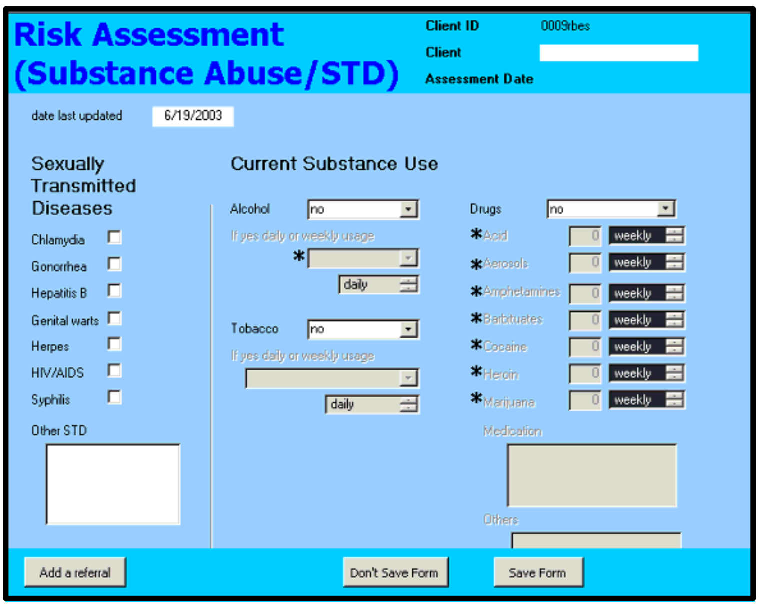
So What Can We Do With GIS?

A common theme running through this section is that a behavioral or individual risk may actually have causative factors originating from the neighborhood or social spatial cohort. As such, there is justification to identify these individual patterns using a GIS. For example, if we again consider prenatal care, spatial patterns will exist if (1) women have little opportunity to access a clinic (for example, the clinic is too far away), or (2) a clinic is accessible but some condition inside the clinic acts to dissuade use, such as a racist attitude of the caregivers, or (3) other individual or neighborhood influences prevent women from attending the clinic (such as a fear of leaving the home). Although studies combining both individual and place risks, such as multilevel modeling, have been mentioned in Chapter II, insight can be gained by simply mapping the locations of these variables. For example, a current investigation in Louisiana is interested in whether there has been an increase in low-birth-weight deliveries in areas where rural-based clinics have been closed due to state-wide budget cuts. The hypothesis states that accessibility has an impact on prenatal visits — in other words, a distance decay of interaction. The further the woman has to travel, the less likely she is of making the visit. By comparing births before and after the closures, differences in the proportion of low-birth-weight deliveries can be mapped and statistically analyzed.

Chapter VII will present another investigation within Baton Rouge identifying neighborhoods with traditionally high levels of “no prenatal visit births.” Obviously, identifying women who make no prenatal visits during their pregnancy is by its very definition an extremely hard task to achieve during the pregnancy. It is also a prime objective of many community health programs that are attempting to serve this particular group. Although the decision to not make a visit may be a personal one, if there is a social or neighborhood contributing factor, then identifying spatial and temporal hot spots from previous birth data could help direct intervention.

A GIS can also be used to “map” risks by assuming smoking is indicative of other stressors. Self-reporting smoking on birth forms, or through more sophisticated tools such as the Four and 4P’s Plus, which is asked at the first prenatal visit, can be linked to a residence to produce a spatial layer. An ongoing screening project in Baton Rogue involves two prenatal care centers using the 4P’s Plus screening questionnaire and intervention into standard prenatal care. The addresses of these women will be address-matched to identify areas of the city with elevated tobacco and alcohol use. These data will be validated using screening responses collected from the Baton Rouge Healthy Start GIS (information collected in Figure 4). Those areas of the city with elevated rates for both screening tools will be used as indicators of neighborhood risk, with screening tool responses being viewed as a sample for other pregnant women from the area. The questions asked on the 4 P and 4 P’s Plus screening tool include:

Figure 4. Behavioral risks



The 4 P's:

1. Have you ever used drugs or alcohol during pregnancy?
2. Have you had a problem with drugs or alcohol in the past?
3. Does your partner have a problem with drugs or alcohol?
4. Do you consider one of your parents to be an addict or alcoholic?

And 4P's Plus:

1. Does your partner have a problem with drugs or alcohol?
2. Do you consider one of your parents to be an addict or alcoholic?
3. What is your prior smoking history?
4. How much alcohol, drugs, or tobacco have you used in your past?
5. How much alcohol, drugs, or tobacco have you used during this pregnancy?

Obviously, a positive response to any of the last three questions will need a follow-up. In order to gain a more representative answer for question 5, the question can be asked about her substance use immediately prior to her knowing of her pregnancy, thus reducing the stigma of admitting using during pregnancy.

For many successful pregnancy intervention programs designed to educate or target specific risk behaviors, a GIS is not used (or at least described in papers presenting their work). For example, one successful program in California, Family Planning, Access, Care, Treatment (Family PACT) has reduced an estimated one in four pregnancies to adolescents. Although results from this program are presented for the state, as with the 4P's Plus approach described above, how much more insight might have been gained from address-matching the participants? In this way, spatial pockets of success could have been identified, as what works for one area may not work elsewhere.

This is not to say GIS is a silver bullet — it must be used in combination with expert and locally informed caregivers. There will never be a substitute for the information gained from personal interviews and focus groups. In Baton Rouge, GIS analysis has identified 18 neighborhoods as having excessively high teenage pregnancy rates. As local education programs aimed at high school students are generally perceived to be successful, why is there this spatial variation? Interviews with neighborhood groups revealed that teenage births often result from older men (in their early to mid-twenties) having sex with teenage girls. Older men prefer younger women, who are seen to be "clean" (an important factor in a city with a high HIV rate). The younger women are entranced by the mystique and purchasing power of the older man and are more susceptible to being coerced into unprotected sex. Although teenage pregnancies per se are not correlated with infant deaths in Baton Rouge, they can be related to a later start with prenatal care, and a problematic home environment in which the child is raised (reflect on all the risks mentioned in this chapter associated with being single). Although pregnancy intervention programs at local schools are still important (though abstinence-only approaches are just plain naïve, as has been shown with the PACT program), an important group to reach is local area males, especially those who have already left school. Although other confounding factors may exhibit spatial patterns, I'm not sure how this particular connection would have ever been captured using a GIS.

As reported in Gatrell (2002), a team of researchers in Huddersfield, England, investigating childhood accidents moved away from a purely statistical analysis, as results would "describe and not explain" (Sparks, Graven, & Worth, 1994). This certainly is in keeping with one Baton Rouge doctor's comment that we need to "put ourselves in the shoes of our program participants" to understand their lives and living spaces so we can help improve them. All researchers are sometimes rightly criticized for not understanding or caring about the health problem being investigated, or the needs and rights of the local community (Weijer, 1999; Weijer & Emanuel, 2000).

Quinn states, "Not infrequently, investigators may enter a community without full comprehension of relevant issues and previous research demands and depart when their grant draws to a close" (2004, p. 922).

I don't deny reductionism exists, and I suggest any GIS analyst should spend some time in a community-based health unit and talk with the nurses, doctors, social workers, and program participants themselves. My golden rule for performing any GIS-based medical or public health analysis is that I can answer most questions concerning the topic under investigation, at least to the degree one would expect without spending many years in medical school. Although some of this knowledge can be gained from the literature, far more useful (and interesting) insights are gained from talking with those on the front line.

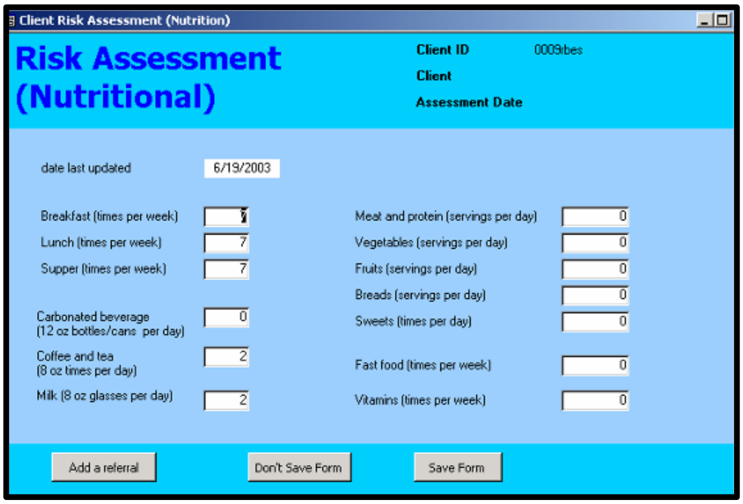
I also have no time for those who dismiss all quantitative analyses. How can a community health program facing 6,000 births annually prioritize care — there are simply not enough caseworkers available to interview everyone. In addition, it is very hard for personal interviews to reveal neighborhood or environmental problems. For that we need pattern analysis, we need to know which neighborhoods are different and then search for explanations. There is no reason why quantitative and qualitative approaches should be mutually exclusive. The Baton Rouge Healthy Start has complimented GIS analysis with focus groups, and even the program participant database, though standardized and mainly numeric in organization, was built with the input of as much local expertise as possible so as to capture individual outcomes as data fields. For other locations without the benefit of a collaboration with a community health unit, community advisory boards (CAB) should be involved in the identification of appropriate research questions and the design needed to address those questions (Weijer & Emanuel, 2000). Hopefully, neighborhood residents could be included in the research process (Kone et al., 2000), their help ranging from collecting data through interviews and focus groups to actually being trained with a PDA running a GIS/GPS. Collaboration with CAB also benefits the research process as it increases trust and participation (Israel, Schulz, Parker, & Becker, 1998; Kone et al., 2000), bringing a neighborhood expertise to the interpretation of results and identifying further research that benefits the community, and not just an academician's CV.

Cohort or Social Risks

As was mentioned in the previous section, many individual or behavioral risks are likely to present as spatial patterns because of the cohort in which the woman lives. Role models, day-to-day experiences, and education initiatives all play some part in molding choices and actions. It is therefore conceivable that any risk can be mapped. For example, Figure 5 displays nutritional questions asked of the Baton Rouge Healthy Start program participants. However, any one of these categories can be mapped by the mother's address to display hot spots of fast food intake, for example. Is this a true spatial pattern? Is the choice of what we eat truly individual, in which case the pattern is spurious (akin to throwing coins in the air and having some fall into "clusters"), or are there social/family influences at work? And again, is the pattern reflective of the availability of healthy food options in the immediate neighborhood?

A further reason why risks may display as spatial patterns for populations that do not geographically move between the generations is "programming," which simply put means that the health of the mother becomes the health of the child. The poor health of a parent can result in a negative birth outcome, which not only has childhood implications in terms of developmental problems, but even later life health effects, such as heart disease (Barker, 1992). Without appropriate health intervention this cycle will continue, with neighborhoods of poor health remaining neighborhoods of poor health (Barker, 1992). The same pattern can be caused by generational "stress" cycles, where the stresses of the parents are passed on as anxieties in the children. The scale of these generational stresses vary from a city neighborhood to whole swaths of the United States, such as the stresses facing the poverty-stricken African American communities in the rural Delta (which will be revisited in Chapter XI), or "hillbilly" communities in Appalachian

Figure 5. Nutrition risks



Eastern Kentucky. Federal commissions have been tasked with trying to improve the living conditions, economic outlook, and health of these communities, in an attempt to break this cycle. Readers interested in these two particular geographic areas should watch the following documentaries: *American Hollow* (1999) and *Lalee's Kin: The Legacy of Cotton* (2001).

One problem in determining an individual risk, whether medical or behavioral, is how much is truly hereditary (a passed-on gene) and how much is socially hereditary (as a result of where a person lives). Is the mother prone to heart disease because of an hereditary condition or because her educational or family environment did not provide the right role models or advice as to why fast food is an unhealthy option, or because she was a short-gestational birth due to her mother's stress coping mechanism during the pregnancy, which was alcohol and tobacco? It is a complex situation.

Social Risks: Disparities in African American Neighborhoods

Although this book has been aimed at using GIS to reduce negative birth outcomes, the Baton Rouge mission is largely one of eliminating racial disparities. In a situation of minority disparity that is mirrored in many American inner cities (Reagan & Salsberry, 2005), the worst birth outcomes are disproportionately experienced by African American communities (Lane et al., 2001). The Baton Rouge profile was described in Chapter I, but to recap the disparity in terms of the infant mortality rate (IMR) for 1998 was almost 5:1 between African

Americans and Caucasians, and the disparity in the low birth weight rate was approximately 3:1. Several studies have found similar disparities for perinatal mortality (Adams et al., 1991; L. A. Schieve & Handler, 1996), low birth weight (Kessel, Kelinman, Koontz, Hogue, & Berendes, 1988; Kleinman & Kessel, 1987; McGrady et al., 1992), and short gestation deliveries (Rolett & Kiely, 2000; Rowley, 1994). In addition, previous outcomes also act as predictors of current outcomes; for example, African American teenage mothers are more susceptible than whites to a short-gestation delivery if a previous birth had also been preterm (Blankson et al., 1993). One explanation for this is the lack of ability to change the negative factors facing the mother and the general lack of quality in perinatal care (Kiely & Kleinman, 1993). These negative influences are found time and again to disproportionately affect inner-city African American cohorts (Hulsey, Levkoff, Alexander, & Tompkins, 1991; Kempe et al., 1992; Shiono et al., 1997), ranging from the domestic situation, including maternal education (Nordstrom & Cnattinguis, 1996) and paternal influence (Klebanoff, Mednick, Schulsinger, Secher, & Shiono, 1998), to quality of life in the neighborhood, with African Americans being more likely to live in unsafe neighborhoods, have less shopping and recreational choices, and suffer persecution by law enforcement (Blackmore et al., 1993; David & Collins, 1991).

African Americans also suffer disproportionately from environmental hazards, with neighborhoods often being found proximate to pollution emitters (Hird & Reese, 1998). Environmental pollutant exposure results in elevated African American asthma levels (Wright et al., 2004) and lead exposure, with a compounding problem of this population being less likely to be screened and treated (Brody et al., 1994; Pirkle et al., 1994). A similar situation is found with the ability to detect and treat the stress levels that individuals in African American neighborhoods face due to elevated risks and racism (Stancil, Hertz-Picciotto, Schramm, & Watt-Morse, 2000). As a cohort, African Americans have frequently been denied the education to understand or the resources to treat these chronic stresses and the psychological distresses that result (Dole et al., 2004).

Of course, as has been previously mentioned, there is a belief by some that a racial disparity is really an economic disparity (Krieger, 1993; Lillie-Blanton & Laveist, 1996; Lillie-Blanton, Parsons, Gayle, & Dievler, 1996; Pappas, 1994). For example, Ahern et al. (2003) found that African American women with no insurance living in impoverished areas (as measured by high unemployment) were at a greater risk of a preterm birth. However, in some studies, even when the other traditional risk factors (including economic measures) are controlled for, race still provides the highest explanation of low birth weight (LaVeist, Rolley, & Diala, 2003; LaVeist, Sellers, & Neighbors, 2001; Polednak, 1991; Roberts, 1997; Schoendorf, Hogue, Kleinman, & Rowley, 1992). Roberts also found that African American risks are reduced in neighborhoods with high proportions of minorities. In the same vein, Pickett, Collins, Masi, and Wilkinson (2005) found that the additional expected benefits of higher income potential and education were countered by the racism faced in relatively wealthier mixed neighborhoods. A further explanation could be that the support network is likely to be stronger within minority neighborhoods (Stack, 1974), as extended families and neighbors participate in the raising of the child. This would seem to make sense, for making a prenatal visit would often involve childcare and transport, two factors which can be impediments for a single mother without a strong support mechanism. This explanation is further strengthened by the finding of a negative association between crowding and low birth weight, which is somewhat counterintuitive but is explained by the density of people acting as a care network. The spatial proximity of friends and family members is important, as they play a large support role (Hogan, Hao, & Parish, 1990; Stack, 1974). However, this last association was found when measures of poverty were held constant. Impoverished African American neighborhoods suffering from overcrowding are still likely to be at high risk to a host of social problems. As a counterpoint, though, other research has found varying degrees of negative health outcomes linked to racial residential segregation (Laveist, 1993; Polednak, 1991; Subramanian, Acevedo-Garcia, & Osypuk, 2005).

Spatial Cohort

Unfortunately, many of the risks described in this chapter impact indigent and/or minority populations, yet risk-mapping should not just equal poverty- or race-mapping, as other risk categories such as single mothers would be missed in such a shotgun approach (Pattenden et al., 1999). Spatial variation can also exist between “similar neighborhoods.” In an Australian study which mirrors the situation found in many American cities, the highest IMR was found in neighborhoods with a high aboriginal concentration, which also happened to coincide with economic poverty (Turrell & Mengersen, 2000). However, this study also showed that in comparing IMR between Adelaide and Brisbane, differences existed even though the underlying social and aboriginal patterns were similar. Geography is important. In other words, similar cohorts will experience different health outcomes. Therefore, a better term to use than “cohort” is “spatial cohort,” which can loosely be defined as one neighborhood, or several closely tied neighborhoods, drawn along racial, cultural, and economic lines. These spatial cohorts will potentially display elevated levels of the same risk due to common experiences (Betemps & Boucher, 1993). For example, residents of these neighborhoods may face similar environmental risks in the home due to building material, such as exposure to lead or asbestos, or even to a commonly used pesticide, as will be mentioned later in this chapter. Environmental hazards could also occur in the neighborhood, such as contaminated groundwater (Bukowski, Somers, & Bryanton, 2001; Sloss, 1999) and landfill sites (Elliott et al., 2001). A neighborhood's infrastructure can result in common cohort risks, such as proximity to heavily trafficked roads, lack of medical facilities, poor access to healthy foods, availability of green spaces (parks and their activities), or even a lack of quality schools. All of these factors can easily be displayed within a GIS. Indeed, a previously mentioned project in Louisiana involves the spatially precise recording of parks and their facilities (right down to the smelliness of trash cans and whether a path was broken or not) being linked to the surrounding public perception of the parks' usability. A perception of a lack of recreational space could result in a more sedentary lifestyle, resulting in health problems such as obesity or asthma.

Of course, there will also be variation within any neighborhood. For example, it has been found that a woman on Medicaid experiences more risks in a poverty stricken neighborhood than someone not on Medicaid because she has a general lack of resources to counter other hardships, (Ahern et al., 2003). In other words, even within a racially defined impoverished neighborhood, some will fare worse than others. However, enough shared experiences and similar negative health outcomes still make the neighborhood an important unit of analysis.

Neighborhood Risks

Neighborhood risks can be quantified for GIS analysis as long as we cheat and define neighborhoods by commonly defined political units such as a census tract, census block, or zip code (though this last unit is fairly useless at describing a neighborhood because of its large size). A good deal of social data is available for these political units, with even health data being more quickly released by health departments at these spatial aggregations. Chapter II broached the question of neighborhood definition and the concerns associated with applying different aggregations of space. This chapter also provided an overview of previous analyses of place, which are commonly referred to as ecological studies in the public health literature (Anderson, Sorlie, Backlund, Johnson, & Kaplan, 1997; Diez-Roux et al., 2001; Duncan, Jones, & Moon, 1999; Kleinschmidt, Hills, & Elliott, 1995; LeClere, Rogers, & Peters, 1998; O'Campo et al., 1997; Pickett & Pearl, 2001; Robert, 1999; Roberts, 1997). The general rationale for including the neighborhood as a contributing influence on a person's health is that the social and physical structure of the immediate environment can cause actual medical conditions, or will accentuate other preexisting conditions (Collins & David, 1992; Gorman, 1999). Typical neighborhood (and domestic) risks that can negatively impact health include racism and violence (Rich-Edwards et al., 2001), abuse (Murphy, Schei, Myhr, & Du Mont, 2001), and prevalence of drugs

(Singer, Hawkins, Huang, Davillier, & Baley, 2001), while infrastructure deficiencies include the availability and quality of local health care, such as the location of intensive care resources (Goodman, Fisher, Little, Stukel, & Chang, 2001), or accessibility to and therefore involvement in WIC (Buescher & Horton, 2001)

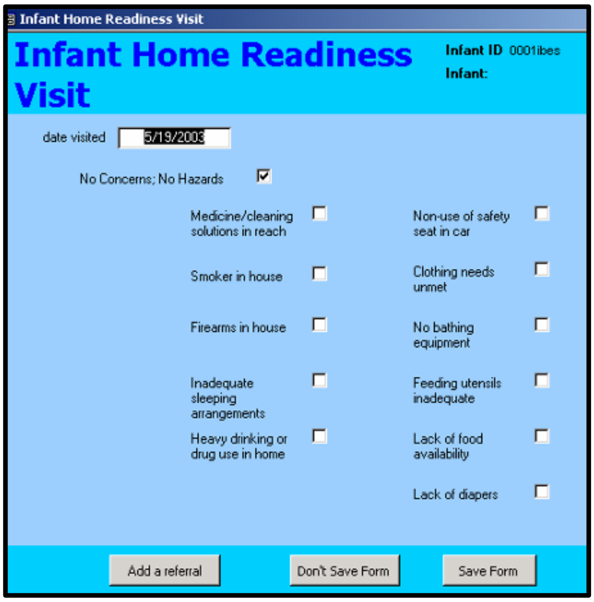
Risks in the neighborhood can also have a direct impact on pregnancy outcomes, (Collins & David, 1992; Gorman, 1999; Pearl et al., 2001; Rauh, Andrews, & Garfinkel, 2001). For example, low birth weight causation has often been linked to different neighborhood socioeconomic measures (Fang et al., 1999; Kirby, Coyle, & Gould, 2001; O'Campo et al., 1997; Pearl et al., 2001; Rauh et al., 2001; Roberts, 1997). Unfortunately a common finding in many studies is that there is a considerable social injustice within American cities. What has been described as an urban underclass (Jeneks & Peterson, 1991) has developed, where the word "disproportionately affected" becomes a phrase relevant to all conditions in the neighborhood. To synthesize these risks into a holistic understanding for a neighborhood is an immense undertaking, especially as the risk combinations are likely to vary spatially between neighborhoods of the same city (Politzer et al., 2001). The new analytical approach called data mining may offer one methodological approach to finding pattern in this complexity.

Figure 1 in Chapter V displays a data entry form for the Baton Rouge Healthy Start, which is designed to capture several domestic and neighborhood risks, such as the presence of rodents, the degree of overcrowding, neighborhood "fears," and even proximity to major roads. An interesting aspect of this data collection tool is that the perception of risk is also recorded for both the program participant and the caseworker. In some instances, the program participant may not know she lives in a risky situation, while in other circumstances she may believe to be at risk even though this view is not shared by the caseworker. Although, if risks are stress-inducing and harmful during the pregnancy, then even perceived risks are of concern.

Suffer the Children

Many of the program participants served by the Baton Rouge Healthy Start already have young children at home. Unfortunately, the type of neighborhood these program participants come from contain many risks that can harm their children (Davidson, Durkin, O'Connor, Barlow, & Heagarty, 1992; Durkin, Kuhn, Davidson, Laraque, & Barlow, 1996). Figure 6 displays another of the caseworker entry forms from the Baton Rouge Healthy Start GIS that identifies risks found in the home that could affect the health of the newborn baby, or any child living in the house for that matter.

Figure 6. Risk in the home



In Chapter X, conditions that make a population more vulnerable in a disaster will be identified. One such vulnerability is having young children in the home, especially if the woman is the head of the household. Although this vulnerability is usually meant to indicate the difficulties involved with the management of the children during a disaster, a similar situation exists during a pregnancy. The mother needs to be able to provide for the children, to have them cared for when she makes a clinic visit, and she needs to make sure they are safe at all times. The amount of danger present in the neighborhood will affect all of these points. If there are too many risks, the mother will be more wary of leaving her children, and even if she does, she may suffer chronic and acute stress. In other words, the safety of her children will have a direct impact on her pregnancy.

Therefore, neighborhood dangers that could affect children are likely to add to the chronic stress burden of the mother, whether she is pregnant or not. Anyone who has lived in a large city, or has at least watched the news, knows that a disproportionate amount of violence befalls inner-city neighborhoods (K. Sheehan, DiCara, LeBailly, & Christoffel, 1997). These neighborhoods also tend to be economically impoverished and racially defined (Fitzpatrick & Boldizar, 1993). Violence and crimes in general are an indicator of the social decay of a neighborhood and the lack of societal normalcy (B. P. Kennedy, Kawachi, Prothrow-Stith, Lochner, & Gupta, 1998; K. Lochner, Kawachi, & Kennedy, 1999; K. A. Lochner, Kawachi, Brennan, & Buka, 2003). Crime also has other indirect effects on the community's infrastructure, resulting in an increased chance of police discrimination, fewer retail choices (and therefore healthy food options), and a lack of community spaces such as parks and playgrounds (Kawachi, 1999). Clinics may close down or operate on limited hours because of crime, which in turn limits prenatal visit opportunities. This is in addition to the mother being fearful of leaving her home (for a similar example in African Americans see Fong, 1995). Funding to health services may also be reduced if excessively high violence leads to the redistribution of Medicaid funds (Wright et al., 2004).

Indirect health problems resulting from elevated crime rates include child development problems (Leventhal & Brooks-Gunn, 2000; Sampson, Raudenbush, & Earls, 1997) and respiratory problems such as asthma (Wright & Steinbach, 2001), some of which can be explained by the increased psychological stress of living in a violent neighborhood (Breslau, Davis, Andreski, & Peterson, 1991; Wright, Rodriguez, & Cohen, 1998), but also because a violent neighborhood will force children to remain inside, exposing them to a series of poverty-related allergic reactions (Rauh, Chew, & Garfinkel, 2002) such as exposure to cockroach excretia (Rosenstreich et al., 1997). Stress can lead to other health problems (Boney-McCoy & Finkelhor, 1995; Martinez & Richters, 1993), including pregnancy problems (Zapata, Rebolledo, Atalah, Newman, & King, 1992). As was mentioned in the previous section, this stress can also be passed onto children in the residence, both in terms of the children reacting to adult stress manifestations, creating "stresses" of their own, but also physically, with changes in neuroendocrine during development (Liu, Diorio, Day, Francis, & Meaney, 2000; Vallee et al., 1997).

Unfortunately, children are also directly impacted by crime, (Fitzpatrick & Boldizar, 1993; Groves, Zuckerman, Marans, & Cohen, 1993), both in the neighborhood and in the home (Zuckerman, Augustyn, Groves, & Parker, 1995). For example, studies in Boston (Groves et al., 1993) found that from an inner-city sample, 10% of young children had witnessed a knifing or shooting, while in Chicago 42% of children between 7 and 13 had witnessed a shooting (K. Sheehan et al., 1997).

Crime is not the only threat to children — exposure to environmental hazards, the risk of sexual predators, and substance use are all of concern. One neighborhood threat which lends itself to a GIS analysis is that of pediatric traffic injuries. In a Floridian study, Hameed, Popkin, Cohn, Johnson, and the Miami Pediatric Traffic Injury Task Force (2004) found that the majority of children hit by automobiles were at grade school (53%), and that most of these injuries occurred close to their schools. African Americans comprised 60% of the accidents, which although the authors did not provide a racial profile of the study area, was stated as being disproportionately high. Also, and again of relevance to many inner-city areas, especially Baton Rouge, 69% of the children came from single parent homes.

Some of the physical reasons attributed to causing the accidents, which were identified by a case-by-case investigation, were view obstruction (causing 46%

of accidents), long stretches of road which allowed for the car to build up speed, and inappropriate crossing points. The lack of traffic lights in locations that had previously caused community concern was also found to be at fault. Other studies have found that high pediatric accident rates are associated with a lack of community facilities, such as parks and playgrounds (Durkin, Laraque, Lubman, & Barlow, 1999). In a situation that mirrors the overly simplistic "guilt" associated with smoking choice, behavioral factors of children leading to accidents should be modified with the lack of available neighborhood facilities (Roberts & Coggan, 1994). Obviously, children need somewhere to play (Mueller, Rivara, Lii, & Weiss, 1990). A lack of a political voice to affect change; a high proportion of single families, meaning children are left unattended in the home while the mother works; an impoverished neighborhood with broken garden fences, allowing for children to reach the road; and even a known lack of police presence, resulting in speeding, are all factors linked to pediatric accidents that are likely to display as hot spots on a map.

In terms of a GIS analysis, a first start would be to liaise with the police department for traffic accident data. In Baton Rouge, an online traffic report details all daily accidents, including type and injuries. If these data were address-matched into the GIS, simple queries, or using techniques such as the kernel density analysis to identify hot spots, could reveal patterns by sections of the city and for different times of day. Other information, such as neighborhood "perception" of dangerous intersections could also be overlaid in the GIS.

Environmental Risks

The idea of environmental pollution and health is well established within modern western society, with many examples in the scientific literature (Betts, 1997; Blatter, van der Star, & Roeleveld, 1994; Briggs & Elliott, 1995; Hardy et al., 1990), and popular media with *A Civil Action* (1999) and *Erin Brockovich* (2000) being two recent movie examples. Names such as Love Canal, Three Mile Island, and Chernobyl have certainly established themselves in the popular lexicon as examples of how modern western development can have serious health impacts. Black lung is one of the few medical conditions directly and unequivocally linked to the industrial complex, or at least one of its inputs (Derickson, 1998). Birth outcomes have also been linked to environmental pollution; for example, exposure to toxins (Huisman, Eerenstein et al., 1995; Huisman, Koopman-Esseboom et al., 1995; Infante-Rivard, Labuda, Krajinovic, & Sinnett, 1999; Infante-Rivard & Sinnett, 1999), or proximity to contaminated groundwater (Sloss, 1999). Both low birth weight and congenital malformations have been linked to landfill sites (Rushton, 2003).

In order to prove the linkage between an environmental hazard and a health outcome, the ideal approach is to identify a single source and tie health outcomes to it without other confounders. A typical point source could be a National Priorities List (NPL) site (Krautheim & Aldrich, 1997), or a hazardous waste site (Croen, Shaw, Sanbonmatsu, Selvin, & Buffler, 1997; Dolk, Vrijheid et al., 1998; Geschwind et al., 1992; Stallones, Nuckols, & Berry, 1992), though also linear exposures, such as living close to a busy roadway (Pearson, Wachtel, & Ebi, 2000), beneath overhead power lines, or near polluted water could also result in health problems. For example, it has been found that proximity to traffic-related pollutants presents more of a mortality risk than from background city pollution (American Heart Association, 2004).

More geographically widespread are airborne exposures, which can be directly inhaled (or absorbed through the skin) or indirectly ingested through the pollution of water bodies, soils, and food types. Common types of airborne exposures include carbon monoxide, sulfur dioxide, ozone, lead, secondhand smoking, and particulate matter. Examples of particulate matter most likely to impact inner-city neighborhoods include transport-related emissions (vehicle emission, tire byproducts, and minute road debris), industrial emissions (power generation, industrial process output), city infrastructure (construction and deconstruction debris), and within the domicile (fuel burning and molds).

The difficulty in all these studies is to control for the confounding influences. For example, it was only recently that the American Heart Association released a statement that many pollution studies have research design flaws resulting in a reluctance of associating long-term health effects to chronic pollution exposure, though the preponderance of studies suggest that heart and stroke deaths are related to chronic pollution exposure (American Heart Association, 2004). This report goes on to state that lower socioeconomic populations are especially at risk, a common finding in environmental risk assessment (Burke, 1993).

For example, asthma, as was discussed in the previous section, can result from individual and social conditions, such as in utero and background tobacco smoke, from the domicile, such as indoor allergens, and from outside air pollution (Wright et al., 2004). Other common environmental concerns for many impoverished inner-city neighborhoods are lead exposure and lead poisoning (Barker, 1994; Cramer, 1995; Dubay, Joyce, Kaestner, & Kenney, 2001; Hanchette, 1999; Reissman, Staley, Curtis, & Kaufmann, 2001; Sargent et al., 1995). Lead exposure can also result from multiple sources. Examples of neighborhood lead pollution sources include being proximate to busy roadways, polluted soil, food, water (running through lead pipes) and paint. Several studies have included the spatial distribution of lead exposure (Stretesky & Lynch, 2001). For example, Griffith, Doyle, Wheeler, and Johnson (1998) used GIS to identify hot spots of lead exposure in Syracuse and found associations with housing material. Spatially identifying neighborhoods mostly likely to be affected by lead exposure is a relatively easy task to accomplish with a GIS, as Centers for Disease Control and Prevention (CDC) guidelines have been laid down which often result in spatial footprints. The specific CDC recommendation is lead screening in neighborhoods where more than 27% of houses were built before 1950. (CDC, 1997; Griffith et al., 1998; Hanchette, 1999). These housing data are available through the census, and many Internet sites allow for data download. It would be relatively simple to overlay housing age onto other socio economic data to gain a first visual impression of lead exposure risk.

Interestingly enough, as with asthma, a link has been established between lead exposure and crime (Needleman, McFarland, Ness, Fienberg, & Tobin, 2002; Needleman, Riess, Tobin, Biesecker, & Greenhouse, 1996; Stretesky & Lynch, 2001, 2004). For example, Denno (1990) performed a longitudinal study on the relationship between lead exposure and crime in African Americans. Unlike asthma, the link between lead exposure and crime may be bidirectional, with some studies suggesting that up to 20% of crime is related to lead exposure (Needleman, 1990). For example, empirical studies have established a relationship between per capita consumption of leaded gasoline and violent crime (Nevin, 2000), while air-lead levels have been linked with homicide (Stretesky & Lynch, 2001). The irony is that residing in a violent neighborhood may result in lead exposure from indoor sources (for the same reasons as discussed for asthma), which eventually might result in the development of a violent individual.

Crime is not the only negative neighborhood and social outcome linked to lead exposure. The "neurotoxicity hypothesis" also relates poor school performance, attention difficulties, and hyperactivity (Stretesky & Lynch, 2004) to lead. Yet again, a poverty cycle is in effect, as impoverished neighborhoods tend to have high crime and poor infrastructure, both of which may lead to an increased lead exposure which in turn limits educational abilities and one of the routes out of poverty. Compounding this condition is the fact that a poor diet (again linked to social and individual poverty) may exacerbate lead poisoning (Reed, 1992)

Not all lead exposure is linked to the home. Airborne lead exposure can occur from industrial and manufacturing emissions. For example, the Cumulative Exposure Project determined county-level airborne pollutant exposure based on the output of a dispersion model, the Assessment System for Population Exposure Nationwide (ASPEN), for a total of 148 pollutants (input being from the Environmental Protection Agency's [EPA] toxic release inventory). Exposures were estimated based on release amounts and local climate data (Rosenbaum, Ligocki, & Wei, 1998). The study found a positive correlation between estimated exposure and Ohio children's blood-lead levels. Ohio was chosen for the analysis as it is a state acknowledged by the CDC for its accuracy in lead screening (Stretesky & Lynch, 2004). Although this study was again searching for a relationship between lead exposure and crime (both property and violent crime were positively associated, with the former having the strongest relationship), more analysis is needed on the effect atmospheric lead pollution (and actually all atmospheric pollution) has on birth outcomes. A simple research methodology to achieve this goal will be presented towards the end of this chapter.

Other common pollutants affecting urban populations include carbon monoxide (CO) and nitrogen dioxide (NO2). Carbon monoxide is generated by vehicle emissions and tobacco smoking, the two often working in tandem to increase carbon monoxide levels in the blood, which in turn reduces oxygenated blood and results in respiratory disease. Nitrogen dioxide is an indoor pollutant which can lead to ozone (O3), though outdoor exposure can result from vehicle emissions and proximity to power plants. Elevated nitrogen dioxide levels in cities with heavy traffic flows strongly suggest that those with respiratory or heart problems, and especially those suffering from asthma, should not live close to busy roadways (American Heart Association, 2004).

Industrial and manufacturing pollution-producing plants are found disproportionately in or proximate to poor, and often racially defined, neighborhoods (Hird & Reese, 1998). For example, the highest levels of cadmium, lead, and diphenyl-trichloroethane (DDT) in breast milk have been found in the poorest cohorts. This is not unexpected, as although the EPA monitors all toxic chemical releases, the majority are found in areas with high concentrations of minorities and low incomes (Neuman, Forman, & Rothlein, 1998). This situation can also be seen in Baton Rouge, where the industrial complexes are located around poor African American neighborhoods to the north of the city, the primary zip code being contained in the Baton Rouge Healthy Start program area due to its continually suffering from elevated negative birth outcomes. For neighborhoods such as these a further poverty trap is in effect, with poor families having little economic ability to make significant moves away from a pollution-affected neighborhood. In addition, being proximate to such emitters keeps housing stock depressed, resulting in other impoverished families moving in to the neighborhoods, which can reinforce existing problems (Ketkar, 1992; A. Nelson, Genereux, & Genereux, 1997). Just as importantly, families with income will not choose to relocate into these neighborhoods. As most of the families living in this poverty trap can be racially defined, it could be argued that polluting facilities help create segregated neighborhoods (Daniels & Friedman, 1999; Stretesky & Hogan, 1998).

This spatial injustice has been studied as an issue of social justice (Sheppard, Leitner, McMaster, & Tian, 1999), and its more environmental and spatial cousin, environmental justice (Stretesky & Hogan, 1998). In general, poor populations have less of a political voice to control the hazardous influences that may touch their communities. They also have fewer resources to react to these influences once established. For example, although minority and indigent neighborhoods are more likely to suffer lead exposure, these residents are the least likely to have the resources to be screened and treated for lead poisoning (Brody et al., 1994; Pirkle et al., 1994; Smedley, Stith, & Nelson, 2002; Williams & Collins, 1995). As was previously mentioned in this chapter, the effects of exposure also create temporal, as well as spatial, footprints of ill health. Whereas it has previously been stated that the health of the mother becomes the health of the child, with later life problems being linked to birth outcomes such as low birth weight, so too does pollution in the environment also have a longer lasting effect. Even if a child living in an impoverished neighborhood and experiencing chronic exposure to air pollution escapes in his or her twenties, it is possible that the reduction of lung capacity that resulted from the initial exposure could have a permanent reduction of lung capacity in later life, resulting in the possibility of respiratory illness and heart disease (Gauderman et al., 2004).

GIS Analyses of Environmental Risks

Health outcomes linked to an environmental risk should be identifiable within a GIS because they will present as nonrandom spatial distributions displaying a proximity relationship to a pollution source. This proximity relationship could be a simple distance decay effect — those individuals living closer to the pollution experience more effects, as was the case with Love Canal — or have the effect modified by other external factors, such as wind or water flow. In theory, the difference of proportions test to be described in the next chapter could be used to answer simple questions such as, are there really more negative outcomes in the neighborhood surrounding this plant as compared to a similar cohort taken from another part of the city? Alternatively, are there more illnesses within a linear buffer following overhead power lines (Valijus et al., 1995)? For example, one non-GIS study found that breast cancer mortality was linked to a toxic disposal site in New Jersey (Najem & Greer, 1985). This relationship is explicitly spatial and would have made for a good GIS investigation.

Release data, such as industrial emissions and infrastructure data (for example pre-1950s housing stock), are relatively accessible. The Toxic Release Inventory (TRI) is a spatially specific database containing hazardous emissions required by the Emergency Planning and Community Right-to-Know Act (EPCRA) to be reported to the EPA. These data are available online. In 1997, the EPA introduced the National Ambient Air Quality Standards (NAAQS), which is a mechanism for distributing information about daily pollutant exposure levels to the public, with fine particulate matter being added to the Air Quality Index in 2003. This system (available at www.epa.gov/airnow) provides daily alerts regarding pollution levels, along with mitigation strategies for the public on poor air quality days.

The Carcinogenic Potency Database (CPDB) is a list of 1,500 chemicals that have been investigated for carcinogenic properties. The sticking point, as usual, is the access to spatially precise health outcome data. With appropriate access to health outcome data there is considerable exploratory spatial analyses that can be performed. Currently, approximately 7% of high volume chemicals, which are chemicals either imported or produced in the United States at a volume greater than one million pounds per year, have received a full toxicological profile, and 43% have no associated hazard information. Further, there are 75,000 different chemicals in use. Although the EPA is addressing these problems, the process will be slow, leaving a degree of uncertainty as to what releases may have harmful effects on the population (EPA, 2000; Roe, Pease, Florini, & Silbergeld, 1997). For example, preliminary research at Louisiana State University has found that in zip code 70805, which is at the heart of the Baton Rouge Healthy Start Program and which will form the basis of mobility investigations in Chapter VII, 41 of 94 chemicals reported in the TRI are not listed in the CPDB (Cunningham, 2004).

Particulate pollution has been linked to mortality increases, with a 10 mg/m3 increase in particulate matter leading to a 1% increase in mortality, a 3.4% increase in respiratory mortality, and a 1.4% increase in cardiovascular mortality. These findings are echoed by the EPA, showing that death rates in U.S. cities rose by 0.5% with a 10 microgram per cubic meter increase of particles smaller than 10 micrometers in size. It has been estimated that in large U.S. cities as many as 60,000 deaths per year can be linked to particulate pollution (American Heart Association, 2004).

Possibly one of the more widely known associations (*A Civil Action* and *Erin Brockovich* being popular examples) is the exposure to a chemical release which either pollutes a local water source (Mallin, 1990) or the surrounding soil. Examples include childhood leukemia and polluted well water in Woburn, Massachusetts (Lagakos, Wesser, & Zelen, 1986), cancer mortality and an Illinois river polluted with agricultural and industrial pollutants (Osborne, Shy, & Kaplan, 1990), and bladder cancer that has been linked to polluted ground water (Mallin, 1990). The initial release is often as a result of a legal or illegal dumpsite (Najem, Louria, Lavenhar, & Feuerman, 1985; Najem, Thind, Lavenhar, & Louria, 1983).

It is, however, the proximity to nuclear facilities that has possibly generated the most spatial environmental risk and health research, both in the United States (K. Crump & R. Cuddihy, 1987; Crump & Cuddihy, 1987; Enstrom, 1983) and especially in Britain, where one of the first ever GIS analyses was applied to childhood leukemia. Elevated levels of lymphoid leukemia had been found in Seascale, near Sellafield, an English nuclear processing plant. A series of articles by Gardner and colleagues (Gardner, 1989; Gardner, Hall, Downes, Powell, & Terrell, 1990a, 1990b; Gardner, Hall, Downes, & Terrell, 1987; Gardner, Hall, Snee, Downes, & Terrell, 1987) linked paternal exposure to ionizing radiation whilst working at the plant. In his seminal GIS/Spatial Analysis investigation, Openshaw (Openshaw, 1989; Openshaw, Craft, Charlton, & Birch, 1988), intrigued by the possible presence of a spatial cluster of leukemia, applied his recently developed technique called the Geographic Analysis Machine (GAM) to a longitudinal cancer data set (1968 to 1985) and for a greater geographic area. Spatial clusters were indeed found, though it was acknowledged that the lack of clustering around a single causative location meant that further social investigation would be needed to determine origin. In other words, confounders could not be ruled out.

It is interesting to wonder how any of the above associations affect birth outcomes. Although it is accepted that at the biochemical level certain chemicals or toxins react with human cells to create an expected result — in other words, there is a biochemical reason for an exposure and health outcome — one does have to wonder what other unknown and previously uninvestigated pregnancy outcomes may result. This seems especially relevant given how few chemicals have been adequately profiled.

An excellent example of where GIS could have been used as an exploratory tool to identify negative birth outcomes in association with a chemical exposure, in this case Methyl Parathion (MP), is in Southern Mississippi. This exposure resulted from indigent families applying a field pesticide (MP) to control infestations within their homes (Rehner, Kolbo, Trump, Smith, & Reid, 2000). Chapter X will discuss this example again in reference to the vulnerability of poor populations during a disaster. However, could this exposure directly cause a negative birth outcome, such as a spontaneous abortion? If a pattern exists, is it directly caused by the chemical, or as a result of a secondary impact such as stress? There was a degree of psychological damage left by the exposure, the *knowing* about the presence of the chemical leading to chronic stress in the home. The problem was that if a family did not suffer a dangerously high level of exposure deserving the financial assistance that would enable relocation, the family was left no choice but to pay for the clean-up or live with the exposure. As many of these families were poor — indeed, the original exposure was caused because it was a cheap insect treatment option — they did not have the resources to clean their house or to relocate. Although all residents would likely suffer some degree of stress in this situation, it is likely females would suffer the most, partly because they spend longer periods within the home fulfilling their domestic roles, which are often more pronounced in poorer communities (Cutter, 1995; Steady, 1993), and partly because of their responsibility to care for the children in the house. This anxiety is often more pronounced during pregnancy due to the fear of the fetus' vulnerability. Although Rehner et al. (2000) did not explicitly consider pregnancy outcomes, as stress has been found to cause low-birth-weight deliveries, it would be interesting to see if exposed households suffered disproportionately from these negative birth outcomes.

This investigation could be easily performed using a GIS. Births in known exposure locations could be compared as a group to births occurring in non-affected properties for the same time period (while at the same time controlling for cohort effects). With longitudinal birth data, pre-and post-exposure births to actual families could also be compared. As the exposure was extensive, it is also possible that spatial associations could be found. For example, aggregations of space (Chapter XI will discuss some of the problems with working with rural aggregations of space) with a high proportion of affected properties could be compared for birth outcomes both before and after exposure. For both longitudinal studies, at the individual and aggregate space level, it would be useful to keep tracking birth outcomes in subsequent years to fully understand the stress/negative birth outcome echo. The reason for this type of analysis is to understand how environmental exposure, or stress related with any disaster, can affect birth outcomes. By understanding this relationship, adequate mitigation strategies can be developed for future exposures. As will be discussed in Chapter X, far too little is known about the vulnerability/disaster/birth outcome triangle.

GIS, Cancer, and Low Birth Weight Research in Louisiana

The Department of Geography and Anthropology at Louisiana State University contains a research focus on the spatial analysis of health outcome and disease data. Indeed, the department currently houses the World Health Organization's Collaborating Center for Remote Sensing and GIS for Public Health. One of the students graduating in this area considered spatial patterns of both cancer and low birth weight within East Baton Rouge Parish, the results of which have been summarized in Ozdenerol and Lam (2004). It is somewhat surprising for two reasons that more studies have not simultaneously investigated these health outcomes. Firstly, similar social, neighborhood, cultural, and medical access issues that have been found associated with negative birth outcomes are likely to affect cancer distribution. Louisiana does not have a disproportionately high cancer morbidity rate, but it ranks as one of the highest for cancer mortality. Louisiana exceeds the United States average mortality rate for every major type of cancer (for white and black populations). This is indicative of a large population without the resources (or education) to access preventative diagnoses centers, such as breast cancer screening. The second reason for conducting the research is as an exploratory analysis based around the premise that a chemical release may result in both cancer and pregnancy outcomes. For example, Mellemkjaer et al. (2003) found that both high birth weight (probably due to estrogens in utero) and low birth weight were linked to the early onset of breast cancer. The association with low birth weight, resulting in a 1.59 times greater risk, is less likely to be caused directly by the pregnancy, but both are expressions of other influences.

Cancer and Birth Outcome Co-Investigation Template

The following research template is based on an ongoing investigation of environmental causations of cancer and birth outcomes in Louisiana. The template is included as a guide for similar investigations in other areas. Required data for the analysis would include cancer morbidity and mortality data (the Louisiana study has concentrated on breast cancer) and linked birth and infant death certificate data. These data are preferably available at the individual level, though census block level aggregations would suffice. The analysis of choice is the spatial filter DMAP (Disease Mapping and Analysis Program) which will be described in detail in Chapter VI. This technique calculates a smooth rate surface calculated from a numerator (health outcome) and a denominator (reference population) expressed per 1,000 outcomes. A Monte Carlo simulation also allows for statistical tests of significance to be generated across this risk surface.

In the initial analysis, the numerator includes all women suffering (for the year under investigation) from breast cancer. The denominator includes the total female population (age 15 and above) by census block. A second approach limits the numerator and denominators to specific cohorts, both by race and age. For example, the numerator being breast cancer cases for African Americans calculated as rates for the entire African American female population. For all analyses, temporal hot spots can be identified (areas exceeding a given rate threshold or significance level). Spatial filter analysis is run for every year (for all morbidity, and separately for those being diagnosed for that year). Once these cancer rate surfaces have been generated they can be overlaid on similarly constructed maps of low birth weight and infant mortality. The same overall and cohort-specific birth outcome surfaces are generated as for the cancer investigation. Therefore, for example, African American hot spots of low-birth-weight deliveries and cancer morbidity are overlaid. By overlaying these analysis grids, areas of the city can be identified whereby:

A: Both low birth weight rates and breast cancer incidence is high.

B: Low birth weight rates are low but cancer incidence is high.

C: Both infant mortality rates and breast cancer rates are high.

D: Infant mortality rates are low but cancer rates are high.

In order to test the environmental causation theory, cancer and birth outcome hot spots can be overlaid on chemical exposure estimates as generated by an air dispersion model, which can be used for either particles or gases. These types of models can either estimate the geographically impacted area of release from a point source, or be used to calculate ambient pollution. Among the relevant model inputs are the horizontal wind flow, the degree of turbulence (which allows for movement within the vertical air column), and the degree to which dry and wet deposition removes the pollutant from the atmosphere. In addition, geographic variables such as the distance from the pollution emitter and the surrounding physical geography all have an impact. In addition, the physical and chemical properties of the release and the amount of background pollution concentrations will also impact the estimate.

There are several options available when determining the correct dispersion model, with some of the differences (in addition to those named above) being the extent and type of the geography under impact (ranging from city to nation), the temporal scale (ranging from hours to years of impact), and whether the emission source is stationary such as a plant, or mobile, such as a tanker. Dispersion models are used to estimate exposure where continual monitoring is unrealistic. In general, the best results are generated for long-term release and exposures over a simple topography. For example, the generally flat terrain and still wind conditions in Baton Rouge should mean that exposure estimates generated for the chemical works should produce reasonably accurate model output.

It is strongly recommended that an expert in particulate dispersion modeling be included in the investigation team, as with any simulation program erroneous data input will affect results. Similarly, exposure type is more complicated than a simple proximity-to-release scenario (which suggests breathing in the pollutant). Other exposure pathways may include food or water ingestion. These pathways may be a simple function of the amount of release, or the toxicity may build to a harmful threshold. Exposure may also result from a secondary interaction, such as eating fish caught from a polluted water source, again with a harmful toxicity level being reached.

The usual approach to dispersion modeling is to employ a two-stage modeling investigation, with the first "screening" being a less sophisticated model using worst case scenario inputs (such as worst case local meteorological conditions) to see if ambient concentrations exceed National Ambient Air Quality Standards (NAAQS).

The second stage requires a more complex model. Four general categories of model exist: Gaussian, numerical, statistical/empirical, and physical. The earlier mentioned dispersion model used in the analysis of Ohio lead exposure was a Gaussian model. In the EPA Guideline, the EPA lists 10 of its own dispersion models, while a further 20 models were included from private sources. The general situations for which these models have been developed, and which are of interest for Louisiana and Baton Rouge-type environments, are rural, urban industrial complex, mobile source, and long-range transport. The first step is to determine the appropriate dispersion model, which for Baton Rouge is an urban industrial complex model. The EPA suggests different ways to define whether the surrounding land (to the pollution site) is urban or rural, with one such measure being that more than 50% of the land within a 3-mile buffer surrounding the source fits within an urban land classification type. An alternative approach, and one easily implemented within the GIS, is that population density within the area under investigation must exceed 750 people/km squared.

According to the EPA Guideline, no one model outperforms all others for a simple-terrain stationary source, which basically means the top of the release is above the surrounding geography. However, "based on past use, public familiarity, and availability, ISC is the recommended model…" (Environmental Protection Agency, 2001, p. 394). The latest version of the ISC model is ISCST3 (Industrial Source Complex Model), which can estimate both air concentrations and deposition rates for all terrains. The ISCST3 estimates emission using a steady-state Gaussian model pollution impact from a single industrial complex up to a distance of 50 km.

Input data includes location, emission rate, physical stack height, stack gas exit velocity, stack inside diameter, and stack gas temperature. Other optional inputs include the particle size and settling velocities. The model is designed to work with continuous releases of toxic and hazardous pollutants. Other required input data include hourly weather data (in Louisiana supplied by the Southern Regional Climate Center), and coordinates of "receptors." Receptors are the locations for which emissions are being estimated. For the template study, these would include coordinates inside the cancer and infant mortality hot spots and controls taken from other areas of the state and city. For coordinates falling inside these hot spots, average concentration or deposition will be displayed (a simple procedure in the GIS of importing coordinates with attached attributes). Again, it is strongly advised that an expert in environmental toxicology be included to interpret the results of these estimated exposures.

This template research has been included because the procedure is relatively easy to perform within a GIS, and yet relatively little of this type of analysis has been performed. Hopefully, this will spur the reader to investigate the link between environmental exposure and cancer/birth outcomes in his or her own back yard.

Summarizing It All: The Relationship Between Risk and Stress

At a Fetal and Infant Mortality Review (FIMR) meeting a comment was made that there is a need to "put yourself in the shoes of the program participant," to understand her risks, her dangers at home, and the reasons why decisions are made, such as why she did not attend a clinic visit. Partly in an attempt to gain this holistic understanding, this chapter has presented a cross section of the threats facing a pregnant woman in an inner-city environment. Some of these risks have immediate pregnancy implications and others may have an indirect effect by contributing to the woman's chronic stress load (Turner & Lloyd, 1999; Turner & Marino, 1994). In addition to these chronic stresses, day-to-day stressors occur (Krieger, Williams, & Moss, 1997), which create their own physical and mental health outcomes (Almeida, Wethington, & Kessler, 2002; Grzywacz, Almeida, Neupert, & Ettner, 2004) that can magnify the effects of the chronic stresses (Caspi, Bolger, & Eckenrode, 1987; Evans, Lepore, Shejwal, & Palsane, 1998; Lepore, Evans, & Palsane, 1991).

As many of the program participants served by the Baton Rouge Healthy Start have young children at home, this stress load is elevated due to a disproportionate number of dangers facing the children in many inner-city neighborhoods. Although stress has been linked to a variety of negative health outcomes (Cohen & Herbert, 1996; Kelly, Hertzman, & Daniels, 1997; Theorell, 1982), it has also been directly associated with pregnancy outcomes, such as low-birth-weight deliveries (Collins et al., 1998).

The disproportionate stress burden placed on people in poverty has been described in the "life stress hypothesis," which basically means that people in these neighborhoods face more risks, while at the same time being more vulnerable to them (Baum, Garofalo, & Yali, 1999), a good example being lead exposure and screening, or the Southern Mississippi MP exposure. People from indigent neighborhoods are victims of different poverty traps, that self-fulfilling cycle of despair that comes from being poor. Poverty traps are often expressed as physical phenomena, for example the family who can only afford the house on the flood plain, where flood insurance is not available, knowing at some point the river will flood and they will lose everything. The same hopelessness can be found when considering poverty-related stress. Families in poor neighborhoods are faced by a disproportionate amount of risk, which causes stress, while at the same time they do not have the resources to move away or seek medical help to diagnose and treat the psychological symptoms that result from the stress (Kessler, 1979a, 1979b; Kessler & Cleary, 1980). If untreated, these stresses may result in a low-birth-weight delivery, which in turn can cause future health problems in the child and on into adulthood, making this particular poverty trap multigenerational.

So What Can Be Done?

This (long) chapter has only scratched the surface in terms of risks facing a pregnant woman. However, all the GIS analysis under the sun is useless unless something positive can result in terms of mitigation, response, or a general improvement in the quality of life. This book has preached a *spatially specific credo* — that is, risks present as spatial patterns, and that these patterns vary in different combinations across space. Therefore, in order for mitigation strategies to be effective, they have to be spatially specific. This belief is hardly a revelation.

For example:

> *Indeed, when the components are carefully selected and administered to pregnant women who exhibit specific, modifiable, risk characteristics (e.g., smoking, substance use, undernutrition, unmitigated pregnancy-reacted anxiety, lower genital tract infections) and perhaps to very young, poor, gravid teenagers, some programs can be effective. The problem is that when programs are applied indiscriminately at the population level to pregnant women, who are assumed to be at increased risk for low birth weight because they are poor or live in high-stress neighborhoods, they are ineffective.* (Stevens-Simon & Orleans, 1999, p. 189)

GIS provides an excellent tool for the spatially specific targeting of mitigation strategies. GIS can also be used as an information system to target individual mitigation strategies. For example, Morgan et al. (2004) found that home-based intervention could reduce asthma morbidity amongst inner-city children by identifying contributing factors, including allergens in the bed and on the bedroom floor (*Dermatophagoides farinae* and *D. pteronyssinus*), and cockroach allergens, again on the bedroom floor. The Healthy Start GIS could easily target which families need such an intervention. Additionally, as all community health programs have limited resources, education strategies could be targeted toward neighborhoods where the incidence of asthma was found to be highest. This last approach, targeting the neighborhood, would result from the ability of the GIS to manipulate, visualize, and analyze spatial risk data in the ways described in the first few chapters. Exploratory analysis can be used to identify spatial clusters of the risk (asthma) using techniques described in Chapters V and VI, while insights gained from the "risk" literature can be used to create testable hypotheses regarding the associations causing the cluster. By becoming more familiar with GIS and its techniques, and at the same time gaining a greater appreciation of space, community health units can start to explore their local situation (Chapters V and VII). Spatially stable "risks" can be identified (Chapter VII) and appropriate mitigation strategies developed.

In order to effectively use the GIS to understand risk we need to have good data, and be able to interpret these data correctly (Chapter II). Sometimes these interpretations come from sophisticated data collection methods and statistical analysis. Sometimes a little intuition and experience is needed. For example, consider Table 1, which shows birth data associated with six women who had three births in a 3-year period. This table is part of a larger investigation into maternal mobility, which will be discussed in Chapter VII.

Of the six women, two recorded three different addresses for the births. Now, consider mothers 1 and 5. Mother 1 would have been an extremely high-risk mother going into the second pregnancy, as she had received no prenatal care during the first. In addition, this first birth was preterm. During the second pregnancy, she only made one visit in the first trimester. However, by the third pregnancy, she received 10 visits starting in the fourth month. Luckily, for all births, the baby was a healthy size. Compare this outcome to mother 5, who lived at a different address for each birth. Her birth data for the first year was good, with 15 visits starting in the first trimester. Birth weight and gestation were also good. However, things started to go downhill with the second pregnancy. She made only five visits in the second year, and three visits in the third year. She also started making these visits later into the pregnancy, during the sixth month in both cases. For both the second and third year, the baby is also a short gestation delivery (though not quite making the low-birth weight-threshold). So what went wrong? Look at the smoking column. She had not smoked with the first pregnancy, but was smoking during the second and third. A superficial reading of this data would be that smoking was the cause, though as has been shown

Table 1. Risk patterns associated with mobile mothers

		Weight	Pren	Month	Gest	Tob	Alc	Edu
Mother 1								
	1998	3062	10	4	38	n	n	11
	1997	3459	1	3	38	n	n	
	1996	3317	0	0	36	n	n	
Mother 2								
	1998	2863	8	2	39	n	n	12
	1997	3175	9	4	39	n	n	
	1996	2552	10	5	37	n	n	
Mother 3								
	1998	2580	14	3	40	n	n	12
	1997	2863	10	4	38	n	n	
	1996	3175	8	2	37	n	n	
Mother 4								
	1998	2552	14	4	38	y	n	9
	1997	2466	2	3	36	y	n	
	1996	2693	12	1	38	n	n	
Mother 5								
	1998	2891	3	6	35	y	n	10
	1997	2637	5	6	36	y	n	
	1996	3430	15	3	38	n	n	
Mother 6								
	1998	3374	9	2	38	n	n	10
	1997	3288	9	2	38	n	n	
	1996	3345	5	4	38	n	n	

throughout this chapter, smoking is often part of a multi-risk situation. Was the movement an indication of a stressful time in her life, did she smoke because of this stress (in other words it was a coping measure), did the resulting prenatal visits drop because of the stress? Alternatively, is the admission to smoking a telltale of other substances uses, and is this woman now in a downward spiral (transient, substance user, continually pregnant, etc.)?

This provides an example of how a dataset and an information system, even working with the limited data fields of a birth certificate, can still be used to develop an understanding of a woman's risk surface. This information can be extremely useful if she becomes pregnant again, as prior pregnancy risks are related to current risks. However, a better approach would be to create an information system containing better data, with far more detail, far more risks being captured, far more quality control, and where risks are automatically flagged if a threshold is reached. The examples of the Baton Rouge Healthy Start caseworker data entry sheets presented throughout this chapter provide an example of such a system. Although the Baton Rouge Healthy Start GIS is not a fully fledged decision support system, it certainly possesses some of its qualities, one example being if a program participant reaches 9 to 13 on the Edinburgh Depression scale, an automatic referral is generated, and at 14 immediate action is required.

The Baton Rouge Healthy Start GIS allows for multiple investigations of the risks captured for its program participants. These risks can be queried in multiple combinations. Figures 7 and 8 show three such summary sheets that can be generated. Figure 7 displays a variety of risks for active infants and program participants, with particular emphasis on the success of intervention: how many women who had reported substance, alcohol, or tobacco use during pregnancy had stopped by the time of birth (as a result of either counseling or Healthy Start education). Figure 8 displays a series of medical risks for Healthy Start Program participants. This query could just have easily focused on active program participants, those who had left the program, or women in the interconception phase of the program.

Of course, GIS is the focus of this book, and although the creation of a management information system is highly beneficial, if we truly believe that many risks are spatial, and that they display as neighborhood footprints on the map, then we must also consider the spatial pattern of these risks. Figure 9 displays six neighborhood risks. Of a subset of 232 program participants enrolled in the Baton Rouge Healthy Start, the number in parentheses shows how many reported the risk in question. Figure 10 displays a sample of medical, individual, and domestic risks. Three points of interest can be gleaned from these maps. Firstly, in the map of chemical exposure (remember this is the program participant's perception), it is interesting that a possible cluster occurs in the upper top left, which is an area of the city proximate to the chemical plants. The

Figure 7. Summary of substance risks and outcomes of intervention

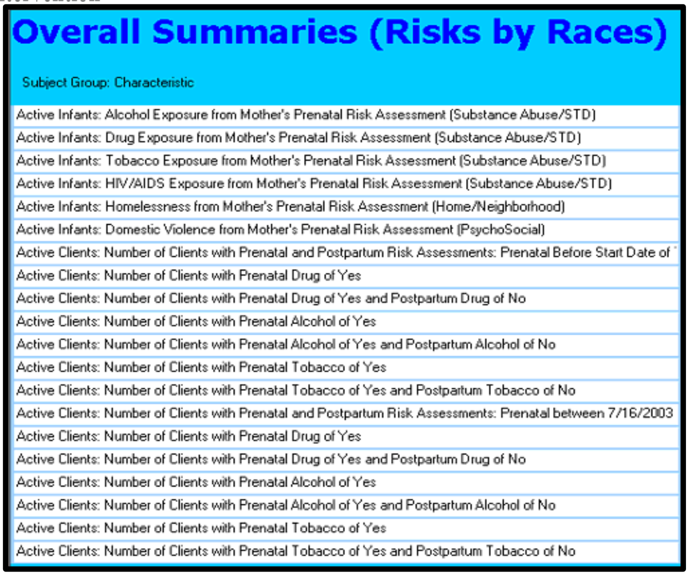
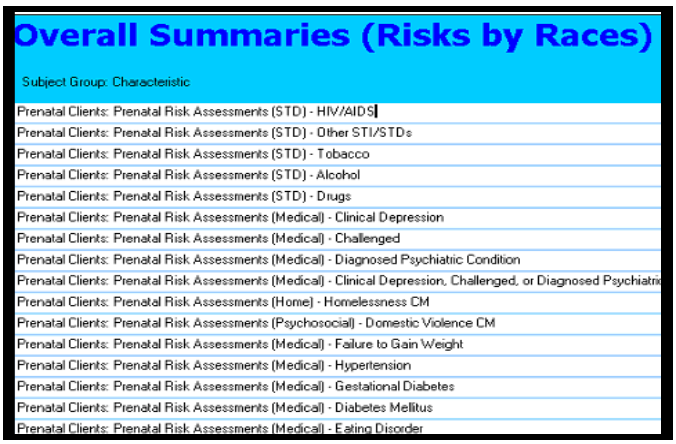
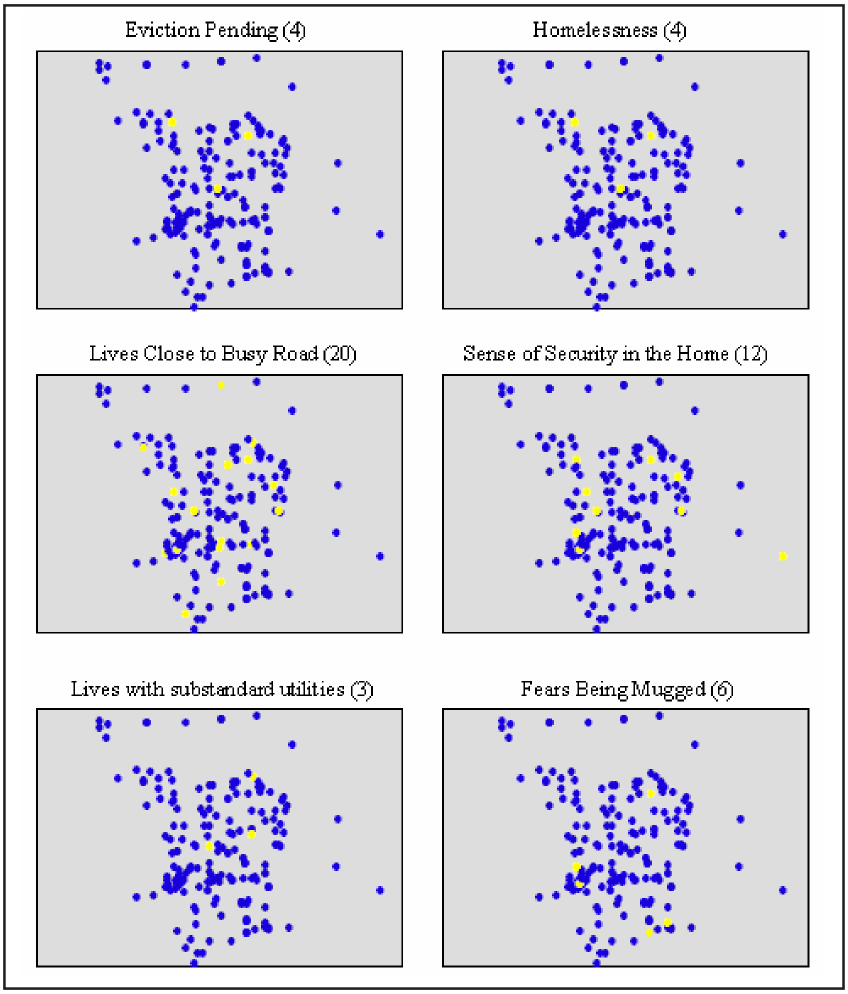


Figure 8. Risk summaries for current



second point of interest is the combination of "smokers" and "lives with a smoker," as these maps show how the category of smoking on the traditional birth certificate underestimates the problem, not only in terms of underreporting by the individual, but also the presence of secondhand smoke in the home. The third point of interest comes from the two maps displaying sexually transmitted disease (STD), as although it has already been mentioned that Baton Rouge has a serious HIV/AIDS problem, it also has an epidemic of STD in general.

Figure 9. Neighborhood/Environmental risks

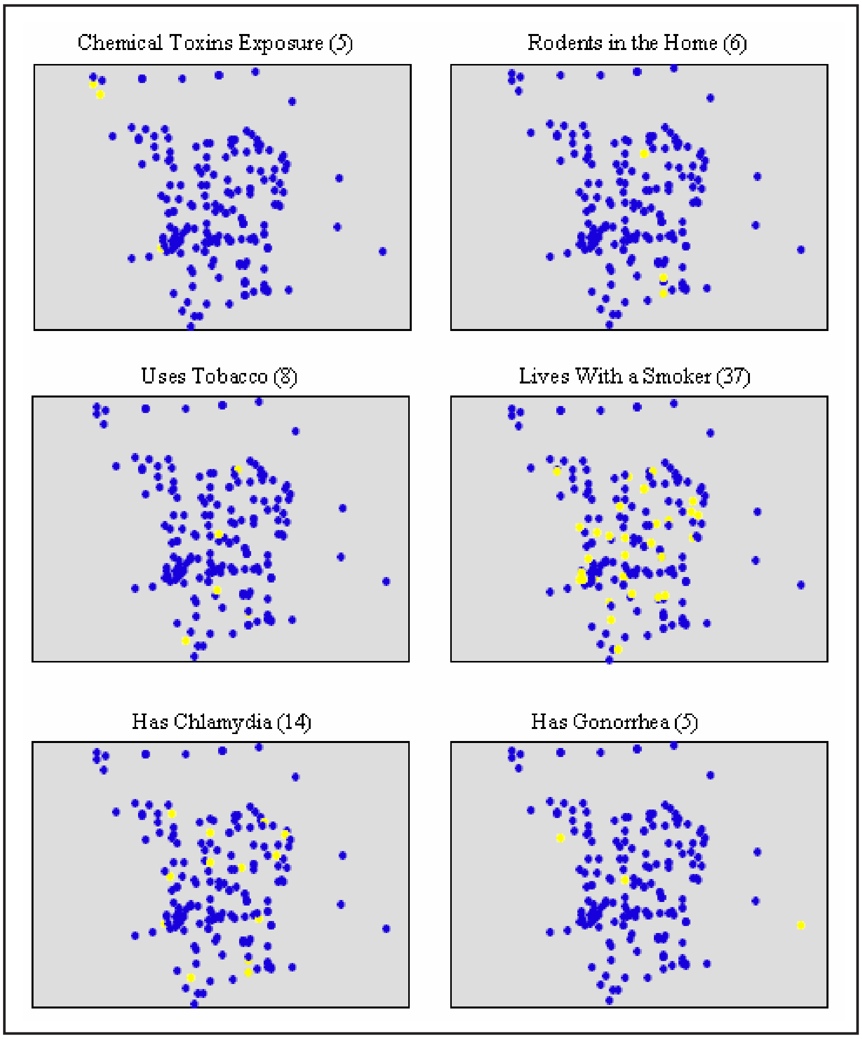


References

Adams, M. M., Berg, C. J., Rhodes, P. H., & McCarthy, B. J. (1991). Another look at the black-white gap in gestation-specific perinatal mortality. *International Journal of Epidemiology, 20*, 950-957.

Adams, M., Delaney, K., Stupp, P., McCarthy, B., & Rawlings, J. (1997). The relationship of interpregnancy interval to infant birthweight and length of gestation among low-risk women, Georgia [Suppl. 1]. *Pediatrics and Perinatal Epidemiology, 11*, S48-62.

Figure 10. Domicile/Behavioral/Medical risks



Adams, M. M., Elam-Evans, L. D., Wilson, H. G., & Gilbertz, D. A. (2000). Rates of and factors associated with recurrence of preterm delivery. *Journal of the American Medical Association, 283*(12), 1591-1596.

Addis, A., Moretti, M. E., Ahmed Syed, F., Einarson, T. R., & Koren, G. (2001). Fetal effects of cocaine: an updated meta-analysis. *Reproductive Toxicology, 15*(4), 341-369.

Ahern, J., Pickett, K. E., Selvin, S., & Abrams, B. (2003). Preterm birth among African American and white women: A multilevel analysis of socioeconomic characteristics and cigarette smoking. *Journal of Epidemiology and Community Health, 57*(8), 606-611.

Ahluwalia, I. B., Hogan, V. K., Grummer-Strawn, L., Colville, W. R., & Peterson, A. (1998). The effect of WIC participation on small-for-gestational-age births: Michigan, 1992. *American Journal of Public Health, 88*(9), 1374-1377.

Alexander, G. R., & Korenbrot, C. C. (1995). The role of prenatal care in preventing low birth weight. *Future Child, 5*(1), 103-120.

Almeida, D. M., Wethington, E., & Kessler, R. C. (2002). The daily inventory of stressful events: An interview-based approach for measuring daily stressors. *Assessment, 9*(1), 41-55.

American Heart Association. (2004). *Air Pollution, Heart Disease and Stroke*. Retrieved September 11, 2004, from http://www.americanheart.org/presenter.jhtml?identifier=4419

American Hollow. (1999). R. Kennedy (Director). In M. F. LLC (Producer): HBO.

Ancel, P., Saurel-Cubizolles, M., Di Renzo, G., Papiernik, E., & Breart, G. (1999). Very and moderate preterm births: Are the risk factors different? *British Journal of Obstetrics and Gynaecology, 106*, 1162-1170.

Anderson, O. W. (1952). *Infant mortality and cultural factors: Historical trends and current patterns*. Washington, DC: U.S. Department of Health, Education, and Welfare.

Anderson, R. T., Sorlie, P., Backlund, E., Johnson, N., & Kaplan, G. A. (1997). Mortality effects of community socioeconomic status. *Epidemiology, 8*(1), 42-47.

Bakewell, J., Stockbauer, J., & Schramm, W. (1997). Factors associated with repetition of low birthweight: Missouri longitudinal study [Suppl. 1]. *Paediatrics and Perinatal Epidemiology, 11*, S119-129.

Bakketeig, L., Hoffman, H., & Harley, E. (1979). The tendency to repeat gestational age and birth weight in successive births. *American Journal of Obstetrics and Gynecology, 135*, 1086-1103.

Barker, D. J. P. (Ed.). (1992). *Fetal & infant origins of adult disease*. London: British Medical Journal Publishing Group.

Barker, D. J. P. (Ed.). (1994). *Mothers, babies and disease in later life*. London: British Medical Journal Publishing Group.

Bartal, M. (2001). Health effects of tobacco use and exposure. *Monaldi Archives of Chest Disease, 56*(6), 545-554.

Baum, A., Garofalo, J. P., & Yali, A. M. (1999). Socioeconomic status and chronic stress: Does stress account for SES effects on health? *Annals of New York Academy Science, 896*, 131-144.

BBC News. (2002). *Brain haemorrhage in preterm babies*. Retrieved January 7, 2002, from http://news.bbc.co.uk

Beaudry, M., Dufour, R., & Marcoux, S. (1995). Relation between infant feeding and infections during the first six months of life. *Journal of Pediatrics, 126*(2), 191-197.

Berkowitz, G., & Papiernik, E. (1993). Epidemiology of preterm birth. *Epidemiologic Reviews, 15*, 414-443.

Betemps, E. J., & Boucher, C. R. (1993). Birthplace as a risk factor in motor neurone disease and Parkinsons disease, International. *International Journal of Epidemiology, 22*, 898-904.

Betts, K. S. (1997). Mapping the environment. *Environmental Health Perspectives, 105*(6), 594-596.

Blackmore, C. A., Ferre, C. D., Rowley, D. L., Hogue, C. J., Gaiter, J., & Atrash, H. (1993). Is race a risk factor or a risk marker for preterm delivery? *Ethnicity and Disease, 3*(4), 372-377.

Blankson, M., Cliver, S., Goldenberg, R., Hickey, C., & Dubard, M. (1993). Health behavior and outcomes in sequential pregnancies of black and white adolescents. *Journal of the American Medical Association, 269*, 1404-1403.

Blatter, B. M., van der Star, M., & Roeleveld, N. (1994). Review of neural tube defects: Risk factors in parental occupation and the environment. *Environmental Health Perspectives, 102*(2), 140-145.

Bode, M. M., O'Shea, T. M., Metzguer, K. R., & Stiles, A. D. (2001). Perinatal regionalization and neonatal mortality in North Carolina, 1968-1994. *American Journal of Obstetrics and Gynecology, 184*(6), 1302-1307.

Bohin, S., Draper, E. S., & Field, D. J. (1999). Health status of a population of infants born before 26 weeks gestation derived from routine data collected between 21 and 27 months post-delivery. *Early Human Development, 55*(1), 9-18.

Boney-McCoy, S., & Finkelhor, D. (1995). Psychosocial sequelae of violent victimization in a national youth sample. *Journal of Consulting and Clinical Psychology, 63*(5), 726-736.

Breslau, N., Davis, G. C., Andreski, P., & Peterson, E. (1991). Traumatic events and posttraumatic stress disorder in an urban population of young adults. *Archives of General Psychiatry, 48*(3), 216-222.

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.

Brody, D. J., & Bracken, M. B. (1987). Short interpregnancy interval: A risk factor for low birthweight. *American Journal of Perinatology, 4*(1), 50-54.

Brody, D. J., Pirkle, J. L., Kramer, R. A., Flegal, K. M., Matte, T. D., Gunter, E. W., et al. (1994). Blood lead levels in the US population: Phase 1 of the Third National Health and Nutrition Examination Survey (NHANES III, 1988 to 1991). *Journal of the American Medical Association, 272*(4), 277-283.

Brooke, O., Anderson, H., Bland, J., Peacock, J., & Stewart, C. (1989). Effects on birth weight of smoking, alcohol, caffeine, socioeconomic factors, and psychosocial stress. *British Medical Journal* [Clinical Research Edition]*, 298,* 795-801.

Buescher, P. A., & Horton, S. (2001). Prenatal WIC participation in relation to low birth weight and Medicaid infant costs in North Carolina — A 1997 update. *Journal of the American Dietetic Association, 101*(9), 997.

Buescher, P. A., Horton, S. J., Devaney, B. L., Roholt, S. J., Lenihan, A. J., Whitmire, J. T., et al. (2003). Child participation in WIC: Medicaid costs and use of health care services. *American Journal of Public Health, 93*(1), 145-150.

Buescher, P. A., Larson, L. C., Nelson, M. D., Jr., & Lenihan, A. J. (1993). Prenatal WIC participation can reduce low birth weight and newborn medical costs: A cost-benefit analysis of WIC participation in North Carolina. *Journal of the American Dietetic Association, 93*(2), 163-166.

Bukowski, J., Somers, G., & Bryanton, J. (2001). Agricultural contamination of groundwater as a possible risk factor for growth restriction or prematurity. *Journal of Occupational and Environmental Medicine, 43*(4), 377-383.

Burke, L. M. (1993, October). Race and environmental equity: A geographic analysis in Los Angeles. *Geographical Information Systems*, 44-50.

Camargo, C. A., Jr., Weiss, S. T., Zhang, S., Willett, W. C., & Speizer, F. E. (1999). Prospective study of body mass index, weight change, and risk of adult-onset asthma in women. *Archives of Internal Medicine, 159*(21), 2582-2588.

Caspi, A., Bolger, N., & Eckenrode, J. (1987). Linking person and context in the daily stress process. *Journal of Personality and Social Psychology, 52*(1), 184-195.

Catalano, R. F., Morrison, D. M., Wells, E. A., Gillmore, M. R., Iritani, B., & Hawkins, J. D. (1992). Ethnic differences in family factors related to early drug initiation. *Journal of Studies on Alcohol, 53*(3), 208-217.

Centers for Disease Control and Prevention (CDC). (1997). *Screening young children for lead poisoning: Guidance for state and local public health officials*. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service, CDC.

Chasnoff, I. (1991). Cocaine and pregnancy: Clinical and methodological issues. *Clinics in Perinatology, 18*, 113-123.

Chomitz, V. R., Cheung, L. W., & Lieberman, E. (1995). The role of lifestyle in preventing low birth weight. *Future Child, 5*(1), 121-138.

A Civil Action. (1999). S. Zaillian (Writer and Director).

Clark, P. I., Scarisbrick-Hauser, A., Gautam, S. P., & Wirk, S. J. (1999). Anti-tobacco socialization in homes of African-American and white parents, and smoking and nonsmoking parents. *Journal of Adolescent Health, 24*(5), 329-339.

Cnattingius, S., Granath, F., Petersson, G., & Harlow, B. L. (1999). The influence of gestational age and smoking habits on the risk of subsequent preterm deliveries. *New England Journal of Medicine, 341*(13), 943-948.

Cohen, S., & Herbert, T. B. (1996). Health psychology: Psychological factors and physical disease from the perspective of human psychoneuroimmunology. *Annual Review of Psychology, 47*, 113-142.

Collins, J. W., Jr., & David, R. J. (1990). The differential effect of traditional risk factors on infant birthweight among blacks and whites in Chicago. *American Journal of Public Health, 80*(6), 679-681.

Collins, J. W., Jr., & David, R. J. (1992). Differences in neonatal mortality by race, income, and prenatal care. *Ethnicity and Disease, 2*(1), 18-26.

Collins, J. W., Jr., & David, R. J. (1993). Race and birthweight in biracial infants. *American Journal of Public Health, 83*(8), 1125-1129.

Collins, J. W., Jr., David, R. J., Symons, R., Handler, A., Wall, S., & Andes, S. (1998). African-American mothers' perception of their residential environment, stressful life events, and very low birthweight. *Epidemiology, 9*(3), 286-289.

Committee on Substance Abuse. (2001). American academy of pediatrics: Tobacco's toll. *Pediatrics, 107*(4), 794-798.

Copper, R. L., Goldenberg, R. L., Das, A., Elder, N., Swain, M., Norman, G., et al. (1996). The preterm prediction study: Maternal stress is associated with spontaneous preterm birth at less than thirty-five weeks' gestation. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. *American Journal of Obstetrics and Gynecology, 175*(5), 1286-1292.

Cotch, M. F., Pastorek, J. G., II, Nugent, R. P., Hillier, S. L., Gibbs, R. S., Martin, D. H., et al. (1997). Trichomonas vaginalis associated with low birth weight and preterm delivery. The Vaginal Infections and Prematurity Study Group. *Sexually Transmitted Diseases, 24*(6), 353-360.

Covington, D. L., Hage, M., Hall, T., & Mathis, M. (2001). Preterm delivery and the severity of violence during pregnancy. *Journal of Reproductive Medicine, 46*(12), 1031-1039.

Cramer, J. C. (1995). Racial and ethnic differences in birthweight: The role of income and financial assistance. *Demography, 32*(2), 231-247.

Croen, L. A., Shaw, C. M., Sanbonmatsu, L., Selvin, S., & Buffler, P. A. (1997). Maternal residential proximity to hazardous waste sites and risk for selected congenital malformations. *Epidemiology, 8*(4), 347-354.

Crosse, E., Alder, R. J., Ostbye, T., & Campbell, M. K. (1997). Small area variation in low birthweight: Looking beyond socioeconomic predictors. *Canadian Journal of Public Health, 88*(1), 57-61.

Crump, K. S., & Cuddihy, R. G. (1987). Cancer incidence patterns in the Denver metropolitan area in relation to the Ricky Flats plant. *American Journal of Epidemiology, 126*, 127-135.

Cunningham, A. (2004). Personal communication. Department of Environmental Studies, Louisiana State University.

Cutter, S. (1995). The forgotten casualties: Women, children and environmental change. *Human and Policy Dimensions, 5*(3), 181-194.

Daniels, G., & Friedman, S. (1999). Spatial inequality and the distribution of industrial toxic releases: Evidence from the 1990 TRI. *Social Science Quarterly, 80*, 244-262.

David, R. J., & Collins, J. W., Jr. (1991). Bad outcomes in black babies: Race or racism? *Ethnicity and Disease, 1*(3), 236-244.

Davidson, L. L., Durkin, M. S., O'Connor, P., Barlow, B., & Heagarty, M. C. (1992). The epidemiology of severe injuries to children in northern Manhattan: Methods and incidence rates. *Paediatric and Perinatal Epidemiology, 6*(2), 153-165.

Davis, K., & Sandman, D. (1998). Economics and urban health. *Journal of Urban Health, 75*(2), 281-293.

Denno, D. (1990). *Biology and violence: From birth to adulthood*. Cambridge: Cambridge University Press.

Derickson, A. (1998). *Black lung: Anatomy of a public health disaster*. Ithaca, NY: Cornell University Press.

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., Kiefe, C. I., Jacobs, D. R., Jr., Haan, M., Jackson, S. A., Nieto, F. J., et al. (2001). Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies. *Annals of Epidemiology, 11*(6), 395-405.

Dole, N., Savitz, D. A., Siega-Riz, A. M., Hertz-Picciotto, I., McMahon, M. J., & Buekens, P. (2004). Psychosocial factors and preterm birth among African American and white women in central north Carolina. *American Journal of Public Health, 94*(8), 1358-1365.

Dolk, H. M., Vrijheid, M., et al. (1998). Risk of congenital anomalies near hazardous-waste landfill sites in Europe: The EUROHAZCON study. *Lancet, 352*(9216), 423-427.

Dorn, M., & Laws, G. (1994). Social theory, body politics, and medical geography: Extending Kearns's invitation. *Professional Geographer, 46*, 106-110.

Draper, E. S., Manktelow, B., Field, D. J., & James, D. (1999). Prediction of survival for preterm births by weight and gestational age: Retrospective population based study. *British Medical Journal, 319*(7217), 1093-1097.

Dubay, L., Joyce, T., Kaestner, R., & Kenney, G. M. (2001). Changes in prenatal care timing and low birth weight by race and socioeconomic status: Implications for the Medicaid expansions for pregnant women. *Health Services Research, 36*(2), 373-398.

Duncan, C., Jones, K., & Moon, G. (1999). Smoking and deprivation: Are there neighbourhood effects? *Social Science and Medicine, 48*(4), 497-505.

Durkin, M. S., Kuhn, L., Davidson, L. L., Laraque, D., & Barlow, B. (1996). Epidemiology and prevention of severe assault and gun injuries to children in an urban community. *Journal of Trauma, 41*(4), 667-673.

Durkin, M. S., Laraque, D., Lubman, I., & Barlow, B. (1999). Epidemiology and prevention of traffic injuries to urban children and adolescents. *Pediatrics, 103*(6), e74.

Ekwo, E. E., & Moawad, A. (2000). Maternal age and preterm births in a black population. *Paediatric and Perinatal Epidemiology, 14*(2), 145-151.

Elliott, P., Briggs, D., Morris, S., de Hoogh, C., Hurt, C., Jensen, T. K., et al. (2001). Risk of adverse birth outcomes in populations living near landfill sites. *British Medical Journal (clinical research edition), 323*(7309), 363-368.

Emery, E. S., III, Eaton, A., Grether, J. K., & Nelson, K. B. (1997). Assessment of gestational age using birth certificate data compared with medical record data. *Paediatric and Perinatal Epidemiology, 11*(3), 313-321.

England, L. J., Kendrick, J. S., Gargiullo, P. M., Zahniser, S. C., & Hannon, W. H. (2001). Measures of maternal tobacco exposure and infant birth weight at term. *American Journal of Epidemiology, 153*(10), 954-960.

England, L. J., Kendrick, J. S., Wilson, H. G., Merritt, R. K., Gargiullo, P. M., & Zahniser, S. C. (2001). Effects of smoking reduction during pregnancy on the birth weight of term infants. *American Journal of Epidemiology, 154*(8), 694-701.

Enstrom, J. E. (1983). Cancer mortality patterns around the Sanonofre nuclear power plant, 1960-1978. *American Journal of Public Health, 73*, 83-92.

Environmental Protection Agency (EPA). (2001). *Pt. 51, App. W 40 CFR Ch. 1*. Environmental Protection Agency. Retrieved from http://www.epa.gov/scrarn001/guidance/guide/appw_01.pdf

EPA. (2000). *Data collection and development on high production volume (HPV) chemicals* (No. Fed. Regist. 65). EPA. Retrieved from http://www.epa.gov/chemrtk/ts42213.pdf

Epstein, J. A., Botvin, G. J., & Diaz, T. (1998). Ethnic and gender differences in smoking prevalence among a longitudinal sample of inner-city adolescents. *Journal of Adolescent Health, 23*(3), 160-166.

Erin Brockovich. (2000). S. Grant (Writer), & S. Soderbergh (Director).

Evans, G. W., Lepore, S. J., Shejwal, B. R., & Palsane, M. N. (1998). Chronic residential crowding and children's well-being: An ecological perspective. *Child Development, 69*(6), 1514-1523.

Fang, J., Madhavan, S., & Alderman, M. H. (1999). Low birthweight: Race and maternal nativity — impact of community income. *Pediatrics, 103*(E5).

Ferraz, E. M., Gray, R. H., Fleming, P. L., & Maia, T. M. (1988). Interpregnancy interval and low birth weight: Findings from a case-control study. *American Journal of Epidemiology, 128*(5), 1111-1116.

Field, D., & Draper, E. S. (1999). Survival and place of delivery following preterm birth: 1994-96. *Archives of Disease in Childhood, 80*(2), F111-114.

Fitzpatrick, K. M., & Boldizar, J. P. (1993). The prevalence and consequences of exposure to violence among African-American youth. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*(2), 424-430.

Fong, R. L. (1995). Violence as a barrier to compliance for the hypertensive urban African American. *J Natl Med Assoc, 87*(3), 203-207.

Ford, R. P., Tappin, D. M., Schluter, P. J., & Wild, C. J. (1997). Smoking during pregnancy: How reliable are maternal self reports in New Zealand? *Journal of Epidemiology and Community Health, 51*(3), 246-251.

Gardner, M. J. (1989). Review of reported increases of childhood cancer rates in the vicinity of nuclear installations in the UK. *Journal of the Royal Statistical Society, Series A, 152*, 307-325.

Gardner, M. J., Hall, A. J., Downes, S., Powell, C. A., & Terrell, J. D. (1990a). Methods and basic data of a case-control study of leukemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. *British Medical Journal, 300*, 429-434.

Gardner, M. J., Hall, A. J., Downes, S., Powell, C. A., & Terrell, J. D. (1990b). Results of a case-control study of leukemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. *British Medical Journal, 300*, 423-429.

Gardner, M. J., Hall, A. J., Downes, S., & Terrell, J. D. (1987). Follow up study of children born to mothers resident in Seacastle, West Cumbria (birth cohort). *British Medical Journal, 295*, 819-822.

Gardner, M. J., Hall, A. J., Snee, M. P., Downes, S., & Terrell, J. D. (1987). Follow up study of children born to mothers resident in Seacastle, West Cumbria (birth cohort). *British Medical Journal, 295*, 822-827.

Gatrell, A. C. (2002). *Geographies of health*. Oxford: Blackwell.

Gauderman, W. J., Avol, E., Gilliland, F., Vora, H., Thomas, D., Berhane, K., McConnell, R., Kuenzli, N., Lurmann, F., Rappaport, E., Margolis, H., Bates, D., & Peters, J. (2004). The effect of air pollution on lung development from 10 to 18 years of age. *The New England Journal of Medicine, 351*(11), 1057-1067.

Gergen, P. J., Fowler, J. A., Maurer, K. R., Davis, W. W., & Overpeck, M. D. (1998). The burden of environmental tobacco smoke exposure on the respiratory health of children 2 months through 5 years of age in the United States: Third National Health and Nutrition Examination Survey, 1988 to 1994. *Pediatrics, 101*(2), E8.

Geronimus, A. T. (1986). The effects of race, residence, and prenatal care on the relationship of maternal age to neonatal mortality. *American Journal of Public Health, 76*(12), 1416-1421.

Geschwind, S. A., Stolwijk, J. A. J., Bracken, M., Fitzgerald, E., Stark, A., Olsen, C., et al. (1992). Risk of congenital malformations associated with proximity to hazardous waste sites. *American Journal of Epidemiology, 135*(11), 1197-1207.

Gesler, W. M., Bird, S. T., & Oljeski, S. A. (1997). Disease ecology and a reformist alternative: The case of infant mortality. *Social Science & Medicine, 44*(5), 657-671.

Gessner, B. D., & Muth, P. T. (2001). Perinatal care regionalization and low birth weight infant mortality rates in Alaska. *American Journal of Obstetrics and Gynecology, 185*(3), 623-628.

Goldenberg, R. L., Klebanoff, M. A., Nugent, R., Krohn, M. A., Hillier, S., & Andrews, W. W. (1996). Bacterial colonization of the vagina during pregnancy in four ethnic groups. Vaginal Infections and Prematurity Study Group. *American Journal of Obstetrics and Gynecology, 174*(5), 1618-1621.

Goldenberg, R. L., & Rouse, D. J. (1998). Prevention of premature birth. *New England Journal of Medicine, 339*(5), 313-320.

Goodman, D. C., Fisher, E. S., Little, G. A., Stukel, T. A., & Chang, C. H. (2001). Are neonatal intensive care resources located according to need? Regional variation in neonatologists, beds, and low birth weight newborns. *Pediatrics, 108*(2), 426-431.

Gorman, B. K. (1999). Racial and ethnic variation in low birthweight in the United States: Individual and contextual determinants. *Health Place, 5*(3), 195-207.

Griesler, P. C., & Kandel, D. B. (1998). Ethnic differences in correlates of adolescent cigarette smoking. *Journal of Adolescent Health, 23*(3), 167-180.

Griesler, P. C., Kandel, D. B., & Davies, M. (2002). Ethnic differences in predictors of initiation and persistence of adolescent cigarette smoking in the National Longitudinal Survey of Youth. *Nicotine and Tobacco Research, 4*(1), 79-93.

Griffith, D., Doyle, P., Wheeler, D., & Johnson, D. (1998). A tale of two swaths: Urban childhood blood lead levels across Syracuse, NY. *Annals Association of American Geographers, 88*, 640-665.

Groves, B. M., Zuckerman, B., Marans, S., & Cohen, D. J. (1993). Silent victims: Children who witness violence. *Journal of the American Medical Association, 269*(2), 262-264.

Grzywacz, J. G., Almeida, D. M., Neupert, S. D., & Ettner, S. L. (2004). Socioeconomic status and health: A micro-level analysis of exposure and vulnerability to daily stressors. *Journal of Health and Social Behavior, 45*(1), 1-16.

Hameed, S. M., Popkin, C. A., Cohn, S. M., Johnson, E. W., & the Miami Pediatric Traffic Injury Task Force. (2004). The epidemic of pediatric traffic injuries in South Florida: A review of the problem and initial results of a prospective surveillance strategy. *American Journal of Public Health, 94*(4), 554-556.

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.

Hardy, R. J., Schroder, G. D., Cooper, S. P., Buffler, P. A., Prichard, H. M., & Crane, M. (1990). A surveillance system for assessing health effects from hazardous exposures [Suppl. 1]. *American Journal of Epidemiology, 132*, S32-42.

Hauth, J. C., Goldenberg, R. L., Andrews, W. W., DuBard, M. B., & Copper, R. L. (1995). Reduced incidence of preterm delivery with metronidazole and erythromycin in women with bacterial vaginosis. *New England Journal of Medicine, 333*(26), 1732-1736.

Headen, S. W., Bauman, K. E., Deane, G. D., & Koch, G. G. (1991). Are the correlates of cigarette smoking initiation different for black and white adolescents? *American Journal of Public Health, 81*(7), 854-858.

Heimendinger, J., Laird, N., Austin, J. E., Timmer, P., & Gershoff, S. (1984). The effects of the WIC program on the growth of infants. *American Journal of Clinical Nutrition, 40*(6), 1250-1257.

Hillier, S. L., Krohn, M. A., Cassen, E., Easterling, T. R., Rabe, L. K., & Eschenbach, D. A. (1995). The role of bacterial vaginosis and vaginal bacteria in amniotic fluid infection in women in preterm labor with intact fetal membranes [Suppl. 2]. *Clinical Infectious Diseases, 20*, S276-278.

Hillier, S. L., Nugent, R. P., Eschenbach, D. A., Krohn, M. A., Gibbs, R. S., Martin, D. H., et al. (1995). Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. The Vaginal Infections and Prematurity Study Group. *New England Journal of Medicine, 333*(26), 1737-1742.

Hird, J. A., & Reese, M. (1998). The distribution of environmental quality: An empirical analysis. *Social Science Quarterly, 79*, 693-716.

Hoffman, H. J., & Bakketeig, L. S. (1984). Risk factors associated with the occurrence of preterm birth. *Clinical Obstetrics and Gynecology, 27*(3), 539-552.

Hogan, D., Hao, L., & Parish, W. (1990). Race, kin networks, and assistance to mother-headed families. *Social Forces, 68*, 797-812.

Howie, P. W., Forsyth, J. S., Ogston, S. A., Clark, A., & Florey, C. D. (1990). Protective effect of breast feeding against infection. *British Medical Journal, 300*(6716), 11-16.

Huisman, M., Eerenstein, S. E., Koopman-Esseboom, C., Brouwer, M., Fidler, V., Muskiet, F. A., et al. (1995). Perinatal exposure to polychlorinated biphenyls and dioxins through dietary intake. *Chemosphere, 31*(10), 4273-4287.

Huisman, M., Koopman-Esseboom, C., Fidler, V., Hadders-Algra, M., van der Paauw, C. G., Tuinstra, L. G., et al. (1995). Perinatal exposure to polychlorinated biphenyls and dioxins and its effect on neonatal neurological development. *Early Human Development, 41*(2), 111-127.

Hulsey, T. C., Levkoff, A. H., Alexander, G. R., & Tompkins, M. (1991). Differences in African-American and white infant birth weights: The role of maternal demographic factors and medical complications of pregnancy. *Southern Medical Journal, 84*, 443-446.

Infante-Rivard, C., Labuda, D., Krajinovic, M., & Sinnett, D. (1999). Risk of childhood leukemia associated with exposure to pesticides and with gene polymorphisms. *Epidemiology, 10*(5), 481-487.

Infante-Rivard, C., & Sinnett, D. (1999). Preconceptional paternal exposure to pesticides and increased risk of childhood leukaemia. *Lancet, 354*(9192), 1819.

Israel, B. A., Schulz, A. J., Parker, E. A., & Becker, A. B. (1998). Review of community-based research: Assessing partnership approaches to improve public health. *Annual Review of Public Health, 19*, 173-202.

Iyasu, S., Becerra, J. E., Rowley, D. L., & Hougue, C. J. (1992). Impact of very low birthweight on the black-white infant mortality gap. *American Journal of Preventative Medicine, 8*(5), 271-277.

Jaakola, J., Jaakkola, N., & Zahlsen, K. (2001). Fetal growth and length of gestation in relation to prenatal exposure to environmental tobacco smoke assessed by hair nicotine concentration. *Environmental Health Perspectives, 109*(6), 557-561.

Jeneks, C., & Peterson, P. (1991). *The urban underclass*. Washington DC: The Brookings Institution.

Joesoef, M. R., Hillier, S. L., Wiknjosastro, G., Sumampouw, H., Linnan, M., Norojono, W., et al. (1995). Intravaginal clindamycin treatment for bacterial vaginosis: Effects on preterm delivery and low birth weight. *American Journal of Obstetrics and Gynecology, 173*(5), 1527-1531.

Kawachi, I. (1999). Social capital and community effects on population and individual health. *Annals of New York Academy Science, 896*, 120-130.

Kelly, S., Hertzman, C., & Daniels, M. (1997). Searching for the biological pathways between stress and health. *Annu Rev Public Health, 18*, 437-462.

Kempe, A., Wise, P., Barkan, S., Sappenfield, W., Sachs, B., Gortmaker, S., et al. (1992). Clinical determinants of the racial disparity in very low birth weight. *New England Journal of Medicine, 327*, 969-973.

Kennedy, B. P., Kawachi, I., Prothrow-Stith, D., Lochner, K., & Gupta, V. (1998). Social capital, income inequality, and firearm violent crime. *Social Science and Medicine, 47*(1), 7-17.

Kennedy, E. T., & Kotelchuck, M. (1984). The effect of WIC supplemental feeding on birth weight: A case-control analysis. *American Journal of Clinical Nutrition, 40*(3), 579-585.

Kesmodel, U., & Olsen, S. F. (2001). Self reported alcohol intake in pregnancy: Comparison between four methods. *Journal of Epidemiology and Community Health, 55*(10), 738-745.

Kesmodel, U., Wisborg, K., Olsen, S. F., Henriksen, T. B., & Secher, N. J. (2002). Moderate alcohol intake during pregnancy and the risk of stillbirth and death in the first year of life. *American Journal of Epidemiology, 155*(4), 305-312.

Kessel, S., Kelinman, J., Koontz, A., Hogue, C., & Berendes, H. (1988). Racial differences in pregnancy outcomes. *Clinics in Perinatology, 15*, 745-754.

Kessler, R. C. (1979a). A strategy for studying differential vulnerability to the psychological consequences of stress. *Journal of Health and Social Behavior, 20*(2), 100-108.

Kessler, R. C. (1979b). Stress, social status, and psychological distress. *Journal of Health and Social Behavior, 20*(3), 259-272.

Kessler, R. C., & Cleary, P. D. (1980). Social class and psychological distress. *American Sociological Review, 45*(3), 463-478.

Ketkar, K. (1992). Hazardous waste sites and property values in the State of New Jersey. *Applied Economics, 24*, 647-659.

Kiely, J. L., & Kleinman, J. C. (1993). Birth-weight-adjusted infant mortality in evaluations of perinatal care: Towards a useful summary measure. *Statistics in Medicine, 12*(3-4), 377-392.

Kirby, D., Coyle, K., & Gould, J. B. (2001). Manifestations of poverty and birthrates among young teenagers in California zip code areas. *Family Planning Perspectives, 33*(2), 63-69.

Klebanoff, M. A., Mednick, B. R., Schulsinger, C., Secher, N. J., & Shiono, P. H. (1998). Father's effect on infant birth weight. *American Journal Obstetrics and Gynecology, 178*, 1022-1026.

Kleinman, J., & Kessel, S. (1987). Racial differences in low birth weight — Trends and risk factors. *New England Journal of Medicine, 317*, 31-39.

Kleinman, J. C., & Madans, J. H. (1985). The effects of maternal smoking, physical stature, and educational attainment on the incidence of low birth weight. *American Journal of Epidemiology, 121*(6), 843-855.

Kleinschmidt, I., Hills, M., & Elliott, P. (1995). Smoking behaviour can be predicted by neighbourhood deprivation measures [Suppl. 2]. *Journal of Epidemiology and Community Health, 49*, S72-77.

Klesges, L. M., Johnson, K. C., Ward, K. D., & Barnard, M. (2001). Smoking cessation in pregnant women. *Obstetrics and Gynecology Clinics of North America, 28*(2), 269-282.

Kone, A., Sullivan, M., Senturia, K. D., Chrisman, N. J., Ciske, S. J., & Krieger, J. W. (2000). Improving collaboration between researchers and communities. *Public Health Report, 115*(2-3), 243-248.

Kotelchuck, M., Schwartz, J. B., Anderka, M. T., & Finison, K. S. (1984). WIC participation and pregnancy outcomes: Massachusetts Statewide Evaluation Project. *American Journal of Public Health, 74*(10), 1086-1092.

Kramer, M., McLean, F., Eason, E., & Usher, R. H. (1992). Maternal nutrition and spontaneous preterm birth. *American Journal of Epidemiology, 136*, 574-583.

Krautheim, K. R., & Aldrich, T. E. (1997). Geographic Information System (GIS) Studies of Cancer around NPL Sites. *Toxicology and Industrial Health, 13*(2/3), 357-362.

Krieger, N. (1991). Women and social class: A methodological study comparing individual, household, and census measures as predictors of black/white differences in reproductive history. *Journal of Epidemiology and Community Health, 45*(1), 35-42.

Krieger, N. (1993). Analyzing socioeconomic and racial/ethnic patterns in health and health care. *American Journal of Public Health, 83*(8), 1086-1087.

Krieger, N., Rowley, D. L., Herman, A. A., Avery, B., & Phillips, M. T. (1993). Racism, sexism, and social class: Implications for studies of health, disease, and well-being [Suppl.]. *American Journal of Preventative Medicine, 9*(6), S82-122.

Krieger, N., Williams, D. R., & Moss, N. E. (1997). Measuring social class in US public health research: Concepts, methodologies, and guidelines. *Annual Review of Public Health, 18*, 341-378.

Krohn, M. A., Hillier, S. L., Lee, M. L., Rabe, L. K., & Eschenbach, D. A. (1991). Vaginal bacteroides species are associated with an increased rate of preterm delivery among women in preterm labor. *Journal of Infectious Diseases, 164*(1), 88-93.

Lagakos, S. W., Wesser, B. J., & Zelen, M. (1986). An analysis of contaminated well water and health effects in Woburn, Massachusetts. *Journal of the American Statistical Association, 81*, 583-614.

LaLee's Kin: The Legacy of Cotton. (2001). S. Froemke, D. Dickson, & A. Maysles (Directors-Screenwriters). In HBO (Producer): HBO.

Lane, S. D., Cibula, D. A., Milano, L. P., Shaw, M., Bourgeois, B., Schweitzer, F., et al. (2001). Racial and ethnic disparities in infant mortality: Risk in social context. *Journal of Public Health Practices, 7*(3), 30-46.

Lang, J., Lieberman, E., Ryan, K., & Monson, R. (1990). Interpregnancy interval and risk of preterm labor. *American Journal of Epidemiology, 132*, 304-309.

LaVeist, T. A. (1993). Segregation, poverty, and empowerment: Health consequences for African Americans. *Milbank Quarterly, 71*(1), 41-64.

LaVeist, T. A., Rolley, N. C., & Diala, C. (2003). Prevalence and patterns of discrimination among U.S. health care consumers. *International Journal of Health Services, 33*(2), 331-344.

LaVeist, T. A., Sellers, R., & Neighbors, H. W. (2001). Perceived racism and self and system blame attribution: Consequences for longevity. *Ethnicity and Disease, 11*(4), 711-721.

LeClere, F. B., Rogers, R. G., & Peters, K. (1998). Neighborhood social context and racial differences in women's heart disease mortality. *Journal of Health and Social Behavior, 39*(2), 91-107.

Lee, T., & Silver, H. (2001). Etiology and epidemiology of preterm premature rupture of the membranes. *Clinics in Perinatology, 28*(4), 721-734.

Lepore, S. J., Evans, G. W., & Palsane, M. N. (1991). Social hassles and psychological health in the context of chronic crowding. *J Health Soc Behav, 32*(4), 357-367.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126*(2), 309-337.

Lillie-Blanton, M., & Laveist, T. (1996). Race/ethnicity, the social environment, and health. *Social Science and Medicine, 43*(1), 83-91.

Lillie-Blanton, M., Parsons, P. E., Gayle, H., & Dievler, A. (1996). Racial differences in health: Not just black and white, but shades of gray. *Annual Review of Public Health, 17*, 411-448.

Litva, A., & Eyles, J. (1995). Coming out: Exposing social theory in medical geography. *Health and Place, 1*, 5-14.

Liu, D., Diorio, J., Day, J. C., Francis, D. D., & Meaney, M. J. (2000). Maternal care, hippocampal synaptogenesis and cognitive development in rats. *Nature Neuroscience, 3*(8), 799-806.

Lochner, K. A., Kawachi, I., Brennan, R. T., & Buka, S. L. (2003). Social capital and neighborhood mortality rates in Chicago. *Social Science and Medicine, 56*(8), 1797-1805.

Lochner, K., Kawachi, I., & Kennedy, B. P. (1999). Social capital: A guide to its measurement. *Health Place, 5*(4), 259-270.

Luginaah, I. N., Lee, K. S., Abernathy, T. J., Sheehan, D., & Webster, G. (1999). Trends and variations in perinatal mortality and low birthweight: The contribution of socio-economic factors. *Canadian Journal of Public Health, 90*(6), 377-381.

Luke, B., Williams, C., Minogue, J., & Keith, L. (1993). The changing pattern of infant mortality in the US: The role of prenatal factors and their obstetrical implications. *International Journal of Gynaecology and Obstetrics, 40*, 199-212.

Lumley, J. (1993). Unhappy babies, unhappy mothers. *Journal of Paediatrics and Child Health, 29*(4), 264-265.

Macintyre, S., MacIver, S., & Sooman, A. (1993). Area, class and health: Should we be focusing on places or people? *Journal of Social Policy, 22*, 213-234.

Madianos, P. N., Lieff, S., Murtha, A. P., Boggess, K. A., Auten, R. L., Jr., Beck, J. D., et al. (2001). Maternal periodontitis and prematurity, Part II: Maternal infection and fetal exposure. *Annals Periodontology, 6*(1), 175-182.

Mallin, K. (1990). Investigation of a bladder cancer cluster in northwestern Illinois. *American Journal of Epidemiology, 132*, S96-106.

Mangold, W. D., & Powell-Griner, E. (1991). Race of parents and infant birthweight in the United States. *Social Biology, 38*(1-2), 13-27.

Martinez, P., & Richters, J. E. (1993). The NIMH community violence project II: Children's distress symptoms associated with violence exposure. *Psychiatry, 56*(1), 22-35.

Mathews, T. J. (1998). Smoking during pregnancy, 1990-96. *National Vital Statistics Reports, 47*(10), 1-12.

Mathews, T. J. (2001). Smoking during pregnancy in the 1990s. *National Vital Statistics Reports, 49*(7), 1-14.

Mayer, J. D., & Meade, M. (1994). A reformed medical geography reconsidered. *Professional Geographer, 4*, 103-106.

McCauley, J., Kern, D. E., Kolodner, K., Derogatis, L. R., & Bass, E. B. (1998). Relation of low-severity violence to women's health. *Journal of General Internal Medicine, 13*(10), 687-691.

McCauley, J., Kern, D. E., Kolodner, K., Dill, L., Schroeder, A. F., DeChant, H. K., et al. (1995). The "battering syndrome": Prevalence and clinical characteristics of domestic violence in primary care internal medicine practices. *Annals of Internal Medicine, 123*(10), 737-746.

McCormick, M. (1983). The contribution of low birth weight to infant mortality and childhood morbidity. *New England Journal of Medicine, 312*, 82-90.

McGrady, G. A., Sung, J. F. C., Rowley, D. L., & Hogue, C. J. (1992). Preterm delivery and low birth weight among first-born infants of black and white college graduates. *American Journal of Epidemiology, 136*(3), 266-276.

Meade, M. S., & Earickson, R. (2000). *Medical geography* (2nd ed.). New York: Guildford Press.

Mellemkjaer, L., Olsen, M. L., Sorensen, H. T., Thulstrup, A. M., Olsen, J., & Olsen, J. H. (2003). Birth weight and risk of early-onset breast cancer (Denmark). *Cancer Causes Control, 14*, 61-64.

Mercer, B. M., Goldenberg, R. L., Moawad, A. H., Meis, P. J., Iams, J. D., Das, A. F., et al. (1999). The preterm prediction study: Effect of gestational age and cause of preterm birth on subsequent obstetric outcome. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network (Pt. 1). *American Journal of Obstetrics and Gynecology, 181*(5), 1216-1221.

Morgan, W. J., Crain, E. F., Gruchalla, R. S., O'Connor, G. T., Kattan, M., Evans, III, R., Stout, J., Malindzak, G., Smartt, E., Plaut, M., Walter, M., Vaughn, B., & Mitchell, H. (2004). Results of a home-based environmental intervention among urban children with asthma. *The New England Journal of Medicine, 351*(11), 1068-1080.

Moss, N., & Carver, K. (1998). The effect of WIC and Medicaid on infant mortality in the United States. *American Journal of Public Health, 88*(9), 1354-1361.

Mueller, B. A., Rivara, F. P., Lii, S. M., & Weiss, N. S. (1990). Environmental factors and the risk for childhood pedestrian-motor vehicle collision occurrence. *American Journal of Epidemiology, 132*(3), 550-560.

Murphy, C. C., Schei, B., Myhr, T. L., & Du Mont, J. (2001). Abuse: A risk factor for low birth weight? A systematic review and meta-analysis. *Canadian Medical Association Journal, 164*(11), 1567-1572.

Najem, G. R., & Greer, T. W. (1985). Female reproductive organs and breast cancer mortality in New Jersey counties and the relationship with certain environmental variables. *Preventative Medicine, 14*, 620-635.

Najem, G. R., Louria, D. B., Lavenhar, M. A., & Feuerman, M. (1985). Clusters of cancer mortality in New Jersey municipalities: With special reference to chemical toxic waste disposal sites and per capita income. *International Journal of Epidemiology, 14*, 528-537.

Najem, G. R., Thind, I. S., Lavenhar, M. A., & Louria, D. B. (1983). Gastrointestinal cancer mortality in New Jersey counties, and the relationship with environmental variables. *International Journal of Epidemiology, 12*, 276-289.

National Institute on Drug Abuse (NIDA) Info Facts. (1997, Wednesday, June 25, 2003.). *Pregnancy and Drug Use Trends*. Retrieved April 9, 2004, from http://www.nida.nih.gov/Infofax/pregnancytrends.html

Needleman, H. L. (1990). The future challenge of lead toxicity. *Environmental Health Perspectives, 89*, 85-89.

Needleman, H. L., McFarland, C., Ness, R. B., Fienberg, S. E., & Tobin, M. J. (2002). Bone lead levels in adjudicated delinquents: A case control study. *Neurotoxicology and Teratology, 24*(6), 711-717.

Needleman, H. L., Riess, J. A., Tobin, M. J., Biesecker, G. E., & Greenhouse, J. B. (1996). Bone lead levels and delinquent behavior. *Journal of the American Medical Association, 275*(5), 363-369.

Nelson, A., Genereux, J., & Genereux, M. (1997). Price effects of landfills on different house value strata. *Journal of Urban Planning and Development, 123*, 59-67.

Nelson, D. E., Giovino, G. A., Shopland, D. R., Mowery, P. D., Mills, S. L., & Eriksen, M. P. (1995). Trends in cigarette smoking among US adolescents, 1974 through 1991. *American Journal of Public Health, 85*(1), 34-40.

Neuman, C., Forman, D., & Rothlein, J. (1998). Hazard screening of chemical releases and environmental equity analysis of populations proximate to toxic release inventory facilities in Oregon. *Environmental Health Perspectives, 106*, 217-226.

Nevin, R. (2000). How lead exposure relates to temporal changes in IQ, violent crime, and unwed pregnancy. *Environmental Research Section A, 83*, 1-22.

Nordstrom, M. L., & Cnattinguis, S. (1996). Effect on birth weights of maternal education, socio-economic status, and work-related characteristics. *Scandinavian Journal Social Medicine, 1*, 55-61.

O'Campo, P., Xue, X., Wang, M., & Caughy, M. (1997). Neighborhood risk factors for low birthweight in Baltimore: A multilevel analysis. *American Journal of Public Health, 87*, 1113-1118.

Offenbacher, S., Lieff, S., Boggess, K. A., Murtha, A. P., Madianos, P. N., Champagne, C. M., et al. (2001). Maternal periodontitis and prematurity, Part I: Obstetric outcome of prematurity and growth restriction. *Annals Periodontology, 6*(1), 164-174.

Openshaw, S. (1989). Automating the Search for Cancer Clusters. *Professional Statistician, 8*, 7-8.

Openshaw, S., Craft, A. W., Charlton, M. G., & Birch, J. M. (1988, February 6). Investigation of Leukemia Clusters by use of a Geographic Analysis Machine. *Lancet*, 272-273.

Osborne, J. S., Shy, C. M., & Kaplan, B. H. (1990). Epidemiologic analysis of a reported cancer cluster in a small rural population. *American Journal of Epidemiology, 132*, S87-S95.

Ozdenerol, E., & Lam, N. S.-N. (2004). Detecting spatial clusters of cancer mortality in East Baton Rouge parish, Louisiana. In K. Hansen (Ed.), *World minds: Geographical perspectives on 100 problems* (pp. 75-79). Washington, DC: Kluwer Academic Publishers.

Paneth, N. S. (1995). The problem of low birth weight. *The Future of Children, 5*(1), 19-34.

Pappas, G. (1994). Elucidating the relationships between race, socioeconomic status, and health. *American Journal of Public Health, 84*(6), 892-893.

Pattenden, S., Dolk, H., & Vrijheid, M. (1999). Inequalities in low birth weight: Parental social class, area deprivation, and "lone mother" status. *Journal of Epidemiology and Community Health, 53*(6), 355-358.

Peacock, J. L., Bland, J. M., & Anderson, H. R. (1995). Preterm delivery: Effects of socioeconomic factors, psychological stress, smoking, alcohol, and caffeine. *British Medical Journal, 311*(7004), 531-535.

Pearl, M., Braveman, P., & Abrams, B. (2001). The relationship of neighborhood socioeconomic characteristics to birthweight among 5 ethnic groups in California. *American Journal of Public Health, 91*(11), 1808-1814.

Pearson, R. L., Wachtel, H., & Ebi, K. L. (2000). Distance-weighted traffic density in proximity to a home is a risk factor for leukemia and other childhood cancers. *Air and Waste Management Association, 50*, 175-180.

Pickett, K., & Pearl, M. (2001). Multilevel analyses of neighborhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health, 55*, 111-122.

Pickett, K. E., Collins, J. W., Jr., Masi, C. M., & Wilkinson, R. G. (2005). The effects of racial density and income incongruity on pregnancy outcomes. *Social Science and Medicine, 60*(10), 2229-2238.

Pirkle, J. L., Brody, D. J., Gunter, E. W., Kramer, R. A., Paschal, D. C., Flegal, K. M., et al. (1994). The decline in blood lead levels in the United States, The National Health and Nutrition Examination Surveys (NHANES). *Journal of the American Medical Association, 272*(4), 284-291.

Polednak, A. P. (1991). Black-white differences in infant mortality in 38 standard metropolitan statistical areas. *American Journal of Public Health, 81*(11), 1480-1482.

Politzer, R. M., Yoon, J., Sh, I. L., Hughes, R. G., Regan, J., & Gaston, M. H. (2001). Inequality in America: The contribution of health centers in reducing and eliminating disparities in access to care. *Medical Care Research and Review, 58*(2), 234-248.

Quinn, S. (2004). Protecting human subjects: The role of community advisory boards. *American Journal of Public Health, 94*(6), 918-922.

Rauh, V. A., Andrews, H. F., & Garfinkel, R. S. (2001). The contribution of maternal age to racial disparities in birthweight: A multilevel perspective. *American Journal of Public Health, 91*(11), 1815-1824.

Rauh, V. A., Chew, G. R., & Garfinkel, R. S. (2002). Deteriorated housing contributes to high cockroach allergen levels in inner-city households [Suppl. 2]. *Environmental Health Perspectives, 110*, 323-327.

Reagan, P. B., & Salsberry, P. J. (2005). Race and ethnic differences in determinants of preterm birth in the USA: Broadening the social context. *Social Science and Medicine, 60*(10), 2217-2228.

Reed, W. (1992). Lead poisoning: A modern plague among African-American children. In S. Taylor (Ed.), *Health issues in the black community* (pp. 178-191). San Francisco: Jossey-Bass.

Rehner, T., Kolbo, J., Trump, R., Smith, C., & Reid, D. (2000). Depression among victims of South Mississippi's methyl parathion disaster. *National Association of Social Workers, 25*(1), 33-40.

Reissman, D. B., Staley, F., Curtis, G. B., & Kaufmann, R. B. (2001). Use of geographic information system technology to aid Health Department decision-making about childhood lead poisoning prevention activities. *Environmental Health Perspectives, 109*(1), 89-94.

Rich-Edwards, J., Krieger, N., Majzoub, J., Zierler, S., Lieberman, E., & Gillman, M. (2001). Maternal experiences of racism and violence as predictors of preterm birth: Rationale and study design. *Paediatric and Perinatal Epidemiology, 15*, 124-135.

Rini, C. K., Dunkel-Schetter, C., Wadhwa, P. D., & Sandman, C. A. (1999). Psychological adaptation and birth outcomes: The role of personal resources, stress, and sociocultural context in pregnancy. *Health Psychology, 18*(4), 333-345.

Robert, S. A. (1999). Socioeconomic position and health: The independent contribution of community socioeconomic context. *Annual Review of Sociology, 25*, 489-516.

Roberts, E. M. (1997). Neighborhood social environments and the distribution of low birth weight in Chicago. *American Journal of Public Health, 87*, 597-603.

Roberts, I., & Coggan, C. (1994). Blaming children for child pedestrian injuries. *Social Science and Medicine, 38*(5), 749-753.

Roe, D., Pease, W., Florini, K., & Silbergeld, E. T. (1997). *Toxic Ignorance*. Retrieved from http://www.environmentaldefense.org/pubs/Reports/ToxicIgnorance/

Rolett, A., & Kiely, J. L. (2000). Maternal sociodemographic characteristics as risk factors for preterm birth in twins versus singletons. *Paediatric and Perinatal Epidemiology, 14*(3), 211-218.

Rosenbaum, A. S., Ligocki, M., & Wei, Y. (1998). *Modeling cumulative outdoor concentrations of hazardous air pollutants*. San Rafael: ICF Kaiser, Systems Applications International Division.

Rosenstreich, D. L., Eggleston, P., Kattan, M., Baker, D., Slavin, R. G., Gergen, P., et al. (1997). The role of cockroach allergy and exposure to cockroach allergen in causing morbidity among inner-city children with asthma. *New England Journal of Medicine, 336*(19), 1356-1363.

Rowley, D. L. (1994). Research issues in the study of very low birth weight and preterm delivery among African-American women. *Journal of the National Medical Association, 86*(10), 761-764.

Rushton, L. (2003). Health hazards and waste management. *British Medical Bulletin, 68*, 183-197.

Salas, M., Hofman, A., & Stricker, B. H. (1999). Confounding by indication: An example of variation in the use of epidemiologic terminology. *American Journal of Epidemiology, 149*(11), 981-983.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*(5328), 918-924.

Sargent, J. D., Brown, M. J., Freeman, J. L., Bailey, A., Goodman, D., & Freeman, D. H., Jr. (1995). Childhood lead poisoning in Massachusetts communities: Its association with sociodemographic and housing characteristics. *American Journal of Public Health, 85*(4), 528-534.

Schieve, L. A., Cogswell, M. E., Scanlon, K. S., Perry, G., Ferre, C., Blackmore-Prince, C., et al. (2000). Prepregnancy body mass index and pregnancy weight gain: Associations with preterm delivery. The NMIHS Collaborative Study Group. *Obstetrics and Gynecology, 96*(2), 194-200.

Schieve, L. A., & Handler, A. (1996). Preterm delivery and perinatal death among black and white infants in a Chicago perinatal registry. *Obstetrics and Gynecology, 88*, 356-363.

Schoendorf, K. C., Hogue, C. J., Kleinman, J. C., & Rowley, D. (1992). Mortality among infants of black as compared with white college-educated parents. *New England Journal of Medicine, 326*(23), 1522-1526.

Shah, N. R., & Bracken, M. B. (2000). A systematic review and meta-analysis of prospective studies on the association between maternal cigarette smoking and preterm delivery. *American Journal of Obstetrics and Gynecology, 182*(2), 465-472.

Sheehan, K., DiCara, J. A., LeBailly, S., & Christoffel, K. K. (1997). Children's exposure to violence in an urban setting. *Archives of Pediatrics & Adolescent Medicine, 151*(5), 502-504.

Sheehan, T. J. (1998). Stress and low birth weight: A structural modeling approach using real life stressors. *Social Science and Medicine, 47*(10), 1503-1512.

Sheppard, E., Leitner, H., McMaster, R. B., & Tian, H. (1999). GIS-based measures of environmental equity: Exploring their sensitivity and significance. *Journal of Exposure Analysis and Environmental Epidemiology, 9*, 18-28.

Shiono, P., Klebanoff, M., & Rhoads, G. (1986). Smoking and drinking during pregnancy: Their effects on preterm birth. *Journal of the American Medical Association, 255*, 82-84.

Shiono, P. H., & Klebanoff, M. A. (1986). Ethnic differences in preterm and very preterm delivery. *American Journal of Public Health, 76*(11), 1317-1321.

Shiono, P. H., Rauh, V. A., Park, M., Lederman, S. A., & Zuskar, D. (1997). Ethnic differences in birthweight: The role of life style and other factors. *American Journal of Public Health, 87*, 787-793.

Shu, X. O., Linet, M. S., Steinbuch, M., Wen, W. Q., Buckley, J. D., Neglia, J. P., et al. (1999). Breast-feeding and risk of childhood acute leukemia. *Journal of the National Cancer Institute, 91*(20), 1765-1772.

Shults, R., Arndt, V., Olshan, A., Martin, C., & Royce, R. (1999). Effects of short interpreganancy intervals on small-for-gestational age and preterm births. *Epidemiology, 10*, 250-254.

Singer, L. T., Hawkins, S., Huang, J., Davillier, M., & Baley J. (2001). Development outcomes and environmental correlates of very low birth, cocaine-exposed infants. *Early Human Development, 64*(2), 91-103.

Sloss, E. M. (1999). *Groundwater recharge with reclaimed water*. Santa Monica, CA: Rand Publications.

Smedley, B., Stith, A., & Nelson, A. (2002). *Confronting racial and ethnic disparities in health care*. Washington DC: National Academy Press.

Sondergaard, C., Henriksen, T. B., Obel, C., & Wisborg, K. (2001). Smoking during pregnancy and infantile colic. *Pediatrics, 108*(2), 342-346.

Sooman, A., & Macintyre, A. (1995). Health and perceptions of the local environment in socially contrasting neighborhoods in Glasgow. *Health and Place, 1*, 15-26.

Sparks, G., Graven, M. A., & Worth, C. (1994). Understanding differences between high and low childhood accident rates: The importance of qualitiative data. *Journal of Public Health Medicine, 17*, 193-199.

Spurlock, M. (Writer/Director). (2004). *Super size me* [Motion Picture]. United States: Samuel Goldwyn Films.

Stack, C. (1974). *All our kin: Strategies for survival in a black community*. New York: Harper & Row.

Stallones, L., Nuckols, J. R., & Berry, J. K. (1992). Surveillance around hazardous waste sites: Geographic information systems and reproductive outcomes. *Environmental Research, 59*, 81-92.

Stancil, T. R., Hertz-Picciotto, I., Schramm, M., & Watt-Morse, M. (2000). Stress and pregnancy among African-American women. *Paediatric and Perinatal Epidemiology, 14*(2), 127-135.

Steady, F. (Ed.). (1993). *Women and children first: Environment, poverty and sustainable development*. Rochester: Schenkman Books.

Stein, C. E., Fall, C. H., Kumaran, K., Osmond, C., Cox, V., & Barker, D. J. (1996). Fetal growth and coronary heart disease in South India. *The Lancet, 348*, 1269-1273.

Stenius-Aarniala, B., Poussa, T., Kvarnstrom, J., Gronlund, E. L., Ylikahri, M., & Mustajoki, P. (2000). Immediate and long term effects of weight reduction in obese people with asthma: Randomised controlled study. *British Medical Journal, 320*(7238), 827-832.

Stevens-Simon, C., & Orleans, M. (1999). Low-birthweight prevention programs: The enigma of failure. *Birth, 26*(3), 184-191.

Stretesky, P., & Hogan, M. (1998). Environmental justice: An analysis of superfund sites in Florida. *Social Problems, 45*, 268-287.

Stretesky, P. B., & Lynch, M. J. (2001). The relationship between lead exposure and homicide. *Archives of Pediatrics & Adolescent Medicine, 155*(5), 579-582.

Stretesky, P. B., & Lynch, M. J. (2004). The relationship between lead and crime. *Journal of Health and Social Behavior, 45*(2), 214-229.

Subramanian, S. V., Acevedo-Garcia, D., & Osypuk, T. L. (2005). Racial residential segregation and geographic heterogeneity in black/white disparity in poor self-rated health in the US: A multilevel statistical analysis. *Social Science and Medicine, 60*(8), 1667-1679.

Theorell, T. G. (1982). Review of research on life events and cardiovascular illness. *Advances in Cardiology, 29*, 140-149.

Tilden, V. P., Schmidt, T. A., Limandri, B. J., Chiodo, G. T., Garland, M. J., & Loveless, P. A. (1994). Factors that influence clinicians' assessment and management of family violence. *American Journal of Public Health, 84*(4), 628-633.

Tin, W., Wariyar, U., & Hey, E. (1997a). Changing prognosis for babies of less than 28 weeks' gestation in the north of England between 1983 and 1994. Northern Neonatal Network. *British Medical Journal, 314*(7074), 107-111.

Tin, W., Wariyar, U., & Hey, E. (1997b). Selection biases invalidate current low birthweight weight-for-gestation standards. The Northern Neonatal Network. *British Journal of Obstetrics and Gynaecology, 104*(2), 180-185.

Tucker, J. M., Goldenberg, R. L., Davis, R. O., Copper, R. L., Winkler, C. L., & Hauth, J. C. (1991). Etiologies of preterm birth in an indigent population:

Is prevention a logical expectation? *Obstetrics and Gynecology, 77*(3), 343-347.

Turner, R. J., & Lloyd, D. A. (1999). The stress process and the social distribution of depression. *Journal of Health and Social Behavior, 40*(4), 374-404.

Turner, R. J., & Marino, F. (1994). Social support and social structure: A descriptive epidemiology. *Journal of Health and Social Behavior, 35*(3), 193-212.

Turrell, G., & Mengersen, K. (2000). Socioeconomic status and infant mortality in Australia: A national study of small urban areas, 1985-89. *Social Science and Medicine, 50*, 1209-1225.

Valijus, J., Hongisto, M., Verkasalo, P., Jarvinen, P., Heikkila, K., & Koskenvuo, M. (1995). Residential exposure to magnetic fields generated by 100-400 kv power lines in Finland. *Bioelectromagnetics, 16*, 365-376.

Vallee, M., Mayo, W., Dellu, F., Le Moal, M., Simon, H., & Maccari, S. (1997). Prenatal stress induces high anxiety and postnatal handling induces low anxiety in adult offspring: Correlation with stress-induced corticosterone secretion. *Journal of Neuroscience, 17*(7), 2626-2636.

Verkerk, P. (1998). Alcohol, pregnancy and the Holy Bible. *Archives of Disease in Childhood, 78*(2), F158-159.

Verkerk, P. H., van Noord-Zaadstra, B. M., Florey, C. D., de Jonge, G. A., & Verloove-Vanhorick, S. P. (1993). The effect of moderate maternal alcohol consumption on birth weight and gestational age in a low risk population. *Early Human Development, 32*(2-3), 121-129.

Wadhwa, P. D., Culhane, J. F., Rauh, V., & Barve, S. S. (2001). Stress and preterm birth: Neuroendocrine, immune/inflammatory, and vascular mechanisms. *Maternal and Child Health Journal, 5*(2), 119-125.

Wadhwa, P. D., Culhane, J. F., Rauh, V., Barve, S. S., Hogan, V., Sandman, C. A., et al. (2001). Stress, infection and preterm birth: A biobehavioural perspective [Suppl. 2]. *Paediatric and Perinatal Epidemiology, 15*, S17-29.

Walsh, R. A., Lowe, J. B., & Hopkins, P. J. (2001). Quitting smoking in pregnancy. *Medical Journal of Australia, 175*(6), 320-323.

Watts, D. H., Krohn, M. A., Hillier, S. L., & Eschenbach, D. A. (1992). The association of occult amniotic fluid infection with gestational age and neonatal outcome among women in preterm labor. *Obstetrics and Gynecology, 79*(3), 351-357.

Watts, D. H., Krohn, M. A., Hillier, S. L., Wener, M. H., Kiviat, N. B., & Eschenbach, D. A. (1993). Characteristics of women in preterm labor associated with elevated C-reactive protein levels (Pt. 1). *Obstetrics and Gynecology, 82*(4), 509-514.

Weijer, C. (1999). Protecting communities in research: Philosophical and pragmatic challenges. *Cambridge Quarterly of Healthcare Ethics, 8*(4), 501-513.

Weijer, C., & Emanuel, E. J. (2000). Ethics: Protecting communities in biomedical research. *Science, 289*(5482), 1142-1144.

Wilcox, A., & Skjoerven, R. (1992). Birth weight and perinatal mortality: The effect of gestational age. *American Journal of Public Health, 82*, 378-382.

Wilcox, A. J. (2001). On the importance — and the unimportance — of birthweight. *International Journal of Epidemiology, 30*, 1233-1241.

Williams, D., & Collins, C. (1995). US socioeconomic and racial differences in health: Patterns and explanations. *Annual Review of Sociology, 21*, 349-386.

Winkvist, A., Mogren, I., & Hogberg, U. (1998). Familial patterns in birth characteristics: Impact on individual and population risks. *International Journal of Epidemiology, 27*, 248-254.

Wisborg, K., Kesmodel, U., Bech, B. H., Hedegaard, M., & Henriksen, T. B. (2003). Maternal consumption of coffee during pregnancy and stillbirth and infant death in first year of life: Prospective study. *British Medical Journal, 326*(7386), 420.

Wisborg, K., Kesmodel, U., Henriksen, T. B., Olsen, S. F., & Secher, N. J. (2001). Exposure to tobacco smoke in utero and the risk of stillbirth and death in the first year of life. *American Journal of Epidemiology, 154*(4), 322-327.

Witter, F. (1993). Perinatal mortality and intrauterine growth retardation. *Current Opinion in Obstetrics and Gynecology, 5*, 56-59.

Wright, R. J., Mitchell, H., Visness, C. M., Cohen, S., Stout, J., Evans, R., et al. (2004). Community violence and asthma morbidity: The inner-city asthma study. *American Journal of Public Health, 94*(4), 625-632.

Wright, R. J., Rodriguez, M., & Cohen, S. (1998). Review of psychosocial stress and asthma: An integrated biopsychosocial approach. *Thorax, 53*(12), 1066-1074.

Wright, R. J., & Steinbach, S. F. (2001). Violence: An unrecognized environmental exposure that may contribute to greater asthma morbidity in high risk inner-city populations. *Environmental Health Perspectives, 109*(10), 1085-1089.

Zapata, B. C., Rebolledo, A., Atalah, E., Newman, B., & King, M. C. (1992). The influence of social and political violence on the risk of pregnancy complications. *American Journal of Public Health, 82*(5), 685-690.

Zuckerman, B., Augustyn, M., Groves, B. M., & Parker, S. (1995). Silent victims revisited: The special case of domestic violence (Pt. 1). *Pediatrics, 96*(3), 511-513.

Chapter V

GIS and Spatial Analysis: Keeping It Simple

In the opening chapters, GIS was broken into four general components, one of which was the spatial analysis of data. This is probably the least utilized of all GIS functions outside of an academic environment. A point that is often missed when discussing GIS is that the technology often exceeds the capabilities of the user. This is especially true if the user has not received any academic training in spatial data and GIS use. In Chapter VI a more sophisticated overview will be presented of the latest spatial analysis techniques along with examples of their implementation. Although the number of "spatially" trained scientists continues to grow, there is still a gap between the number of available skilled GIS modelers and the community programs needing GIS analysis. This chapter is designed to provide a stopgap approach, using more simple spatial statistical approaches that can be applied to gain a reasonable first insight into a birth outcome surface.

Exploratory Analysis vs. Hypothesis Testing

GIS can be considered as a tool to map data and then perform pattern explorations upon that mapped data. Initial explorations are needed before more science-based inquiry can be applied. This raises an interesting discussion point. At previous academic presentations, I have been questioned as to a theoretical question or underpinning to the investigation. Chapter IV discussed the ongoing debate within medical geography, and for that matter geography in general, as to the validity of using GIS. In my own department comments have been made as to "GIS is not geography," "we are not a vocational school," and "GIS faculty should not mentor PhD students," Readers interested in following this path further are directed to an exchange in the flagship journal of our discipline, and the question of what is GIS (Pickles, 1997; Wright, Goodchild, & Proctor, 1997). I am sensitive to some of these criticisms, and hopefully the reader will have gleaned that I have tried to include a social context. However, when faced with these questions my usual response is an excuse of both time and resource limitation. The best must be made of what is available to make immediate improvements in the health of our communities.

Maybe the best way to consider this debate is to use the analogy of a battlefield hospital. When the wounded are coming in, triage is needed. At hospitals stateside, more holistic approaches to the healing of patients can be developed because of the safety and sanity of the peacetime environment. Advances made in these locations will probably work their way through to the front lines, but these developments are not expected to come from those front line doctors. That is not to say that occasionally a breakthrough occurs under fire. This may seem a dramatic analogy but it is true of much investigative epidemiology, especially if dealing with a disease of unknown etiology. Consider again John Snow — although his work certainly generated further research and theory, the initial aim was to stop the spread of disease and save lives. This was achieved by recognizing a spatial pattern.

A similar situation exists when working with infant mortality in Baton Rouge. Until I began to work on the problem, very little had been performed (if anything) in terms of spatial or GIS analysis. The question was where to start? Birth and death data needed to be applied for, and once obtained, their addresses needed cleaning. After this preparation stage, the analysis began. The first step was to identify those neighborhoods with the highest infant mortality, followed by low birth weight, and finally high numbers of women receiving inadequate prenatal care. More recently, neighborhoods have been identified with high alcohol and tobacco use during pregnancy (using the 4 P's Plus screening questionnaire), as well as breast cancer hot spots and areas where women are underserved by breast-screening programs. These investigations were driven by a need to direct limited resources in terms of outreach. When questions are asked such as "haven't you considered the impact of environmental toxins?" the answer is "no, not yet" (though in actuality that is also being addressed in the breast cancer research). It is an important question, and one that would naturally follow the first series of investigations based on an exploratory analysis. As a renowned epidemiologist once told me, we often do not know what we are looking for, or have the luxury of time to develop the scientific method. When faced with limited time and limited resources, the researcher must perform triage and identify where he or she can make the most impact.

This is not to dismiss all hypothesis testing in GIS science. The researcher can look at a revealed mortality hot spot, and maybe through overlaying other spatial data, such as a drug arrest surface (a surrogate for neighborhood stress), develop a working hypothesis. This hypothesis could be tested across the entire city or against other sample locations displaying similar socioeconomic and physical characteristics. One of the most commonly used approaches in the social sciences, multiple regression analysis, could be used to test the hypothesis, with further exploration of the residuals revealing even more detail in the spatial variation of the model. Chapter VI will show how this particular approach, in the form of Geographically Weighted Regression, has evolved to take into account local characteristics in the data, while at the same time solving the perennially tricky problem of dependence in the observations.

This last point nicely sets us up for the next section, for although insight can be gained from applying even simple statistical approaches to our data, it is important to know and not violate statistical basics. It is also important to have a good working knowledge of spatial design, adequate sample size, and the effect spatial aggregation can have on the data. Although a whole textbook could be written about these issues (and several have), each of these topics will be briefly discussed in this chapter because of their importance in the analysis of spatial data.

Spatial Design

When most people consider the spatial design of a project, they are usually thinking about what sample strategy to choose. However, the design should also include how best to attack the spatial problem in hand. I will return with examples of spatial sampling in a moment, but let me first explain this "attacking a spatial problem" statement. A researcher once came to LSU interested in performing an analysis of alcohol use around the campus. Previous work had involved the heads-up digitizing of all alcohol selling locations (bars, restaurants, gas stations, etc.). No attempt had been made to weight these locations by the amount of alcohol sold. The study in question was going to concentrate on bars. Most spatial scientists will tell you that although the spatial pattern is important, the weighting of those points will often reveal a completely different relationship. Although alcohol volume sold is collected for taxation purposes, getting that data is extremely hard. However, why not use the number of bar staff (working behind the bar and servers on the floor) as a proxy for total consumption? Bar managers generally understand the needs of their clientele and they will have adequate staff on duty to satisfy demand. A quick count, at different times of the day, and for different days of the week, will give a more informed amount of alcohol flow. A density function, such as a kernel density analysis, could then be run with the addition of this weight revealing hot spots of drinking.

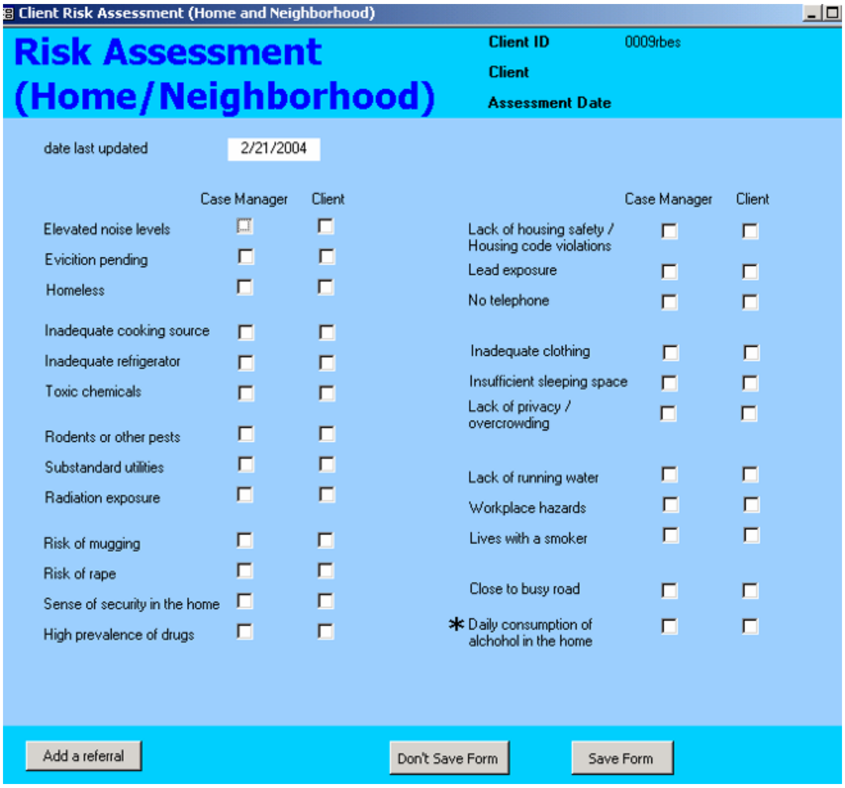
Spatial Sampling

There are several reasons to use a sample instead of a complete population. The first is an inability to reach the entire population because of logistical reasons, usually cost. Although this might not seem an appropriate reason to "compromise" research, the fact of the matter is that even the largest most diverse population can be reached, eventually, if you have enough money. For an amusing debate on this issue in regards to the U.S. census see the episode "Mr. Willis of Ohio" in the first season of the *West Wing* (n.d.). Most researchers are content to apply sampling theory and identify the correct number of respondents needed so that an extrapolation can be made to the entire population. The various formulas that are necessary to identify the correct sample size needed can be found in most elementary spatial statistics books (McGrew & Monroe, 2000).

A second reason for sampling is to control the quality of the sample collection. A smaller number of data collectors, better armed with appropriate skills, knowing the correct time (and location) to collect data, and carrying a longer and more involved set of questions create a far more robust data set. The analysis then benefits from a data-rich sample of superior quality. Although not designed as a sample, the program participants currently enrolled in the Baton Rouge Healthy Start can be considered as a sample and provide an excellent source of information about their environments, living conditions, social connections, psychosocial state, etc. Consider the information collected in the data entry form presented in Figure 1. It would be impossible to ask these questions of all pregnant women; resources are simply too limited. However, we could use these responses as samples for either pregnant women originating from the study region (a five zip code area), or from a single zip code, and eventually when numbers get high enough, for a single neighborhood. These data could also be used to extrapolate to the general community, as many of these responses are not limited to pregnant women, but to all women ("fear of rape"), or all residents ("lack of running water").

When constructing a sample, it is important to avoid creating any bias that will affect the results. For example, if you wanted to gain an insight into the

Figure 1.



environmental pressures faced by mothers across the city, a phone survey would not be appropriate. Not everyone owns a phone, and it is possible you would be missing the population most impacted by environmental problems — the poor who cannot afford a telephone. It is important to try to understand where these biases may exist. For example, the Baton Rouge Healthy Start program participants are a biased sample if generalized to the entire parish. The program participant base tends to be women close to the poverty level, and usually African American, so using these data for insights into the general pregnancy population, which would include suburban middle-class women, would obviously be limited. Therefore, any findings from an analysis of the program participant data should only be extrapolated to similar communities.

In terms of the actual mechanics, spatial sampling is similar to ordinary sampling, except that the location of the sample becomes important. There are three general spatial sampling approaches, which actually follow the spatial data types of a vector GIS: point, line, and area samples. The best way to describe these is with an example: Imagine a team is being sent to survey hazards in the home space. This survey will include exterior dangers, such as broken garden fences (these allow children to run into the road) or broken windows, and interior dangers such as lead paint or evidence of rodents. The researchers want to use these results in an analysis of environmental linkages to low-birth-weight deliveries.

For a point sampling scheme, a random number generator is used to calculate both an X and Y coordinate. These coordinates could be standard latitude and longitude, or a Universal Transverse Mercator coordinate if a Digital Orthophoto Quarter Quadrangle (DOQQ) is used, or even a simple Cartesian coordinate from a superimposed grid found on many city road maps. This process is repeated for the number of required sample locations. The advantage of this sample scheme is *ceteris paribus*, the resulting sample should not contain any bias. The disadvantage is that in a fragmented urban environment, such as Baton Rouge, it is entirely possible that all neighborhoods in poverty could be missed. Therefore, this approach would not be particularly useful in the analysis, as sample locations may not coincide with low-birth-weight clusters.

For a linear sample, otherwise known as a transect or traverse, a random number generator is again used to pick a coordinate on two desired axes. A line is drawn between these coordinates. A further random number generator would pick distances along each of these transects to provide sample locations. The advantage of using a linear sample scheme is that even in a fragmented urban environment, samples from all types of neighborhoods should be captured. In terms of the described analysis, this approach could be used, as it is likely that the transects would include both neighborhoods with high and low proportions of low-birth-weight deliveries.

For an area sample, the same approach is applied as for the point sample, with the randomly drawn coordinate being either the centroid or the corner of a sample neighborhood. All houses falling inside this neighborhood could be surveyed, or a further sample scheme is applied whereby a selection of houses inside the area is surveyed. The advantage of this sample scheme is that costs are reduced, as multiple locations can be sampled in close proximity. Secondly, the fear of picking the unrepresentative household from a neighborhood is eliminated.

Alternatives to a random sample are spatially systematic and proportionally stratified samples. A systematic sampling scheme would ensure that all of the study space was covered, usually by a lattice or grid. Therefore, if the city was 5 square miles, a sample scheme might result in a grid being created with a 0.5-mile graticule, with a survey being administered at each intersection point. A proportionally stratified sample takes into account the underlying global pattern in the data. For example, census tracts could be classified according to their median income. A proportionally stratified sample would ensure that the sample locations selected using any of the previously described approaches would sample within census tracts in proportion to the amount of census tracts falling into each classification. If 10 tracts were low-income, six tracts were medium-income, and four tracts high-income, a sample scheme selecting residences for survey requiring 100 residences would pick 50 from low-income, 30 from medium-income, and 20 from high-income census tracts. This approach could be further modified to also include other variables, such as race of the householder.

A disproportional stratified sample would take into account the above spatial structure, but then over sample according to an expected outcome. For example, a research statement presented with the environmental hazard analysis states that low-birth-weight clusters are expected to be found more frequently in low-income neighborhoods. Therefore, as this is the relationship under investigation, instead of 50 residences being selected from low-income census tracts, 75 would be surveyed, leaving 15 for middle-income and 10 for high-income census tracts. This approach probably provides the best sample scheme for the above stated research question.

These sample schemes by no means cover all approaches, as many hybrids can be constructed. For further reference see McGrew and Monroe (2000). It is also possible that a research design does not fit into any traditional approach, and the investigator needs to be flexible and a little creative (while still being mindful of creating bias). For example, one of my PhD students came to me needing a sample scheme involving tagging sharks, at different depths of water, around different underwater features, with attention being paid to currents and whether the sample location was on the gulf or land side of an island. I guarantee the answer to that one does not appear in any text.

One of the best examples of using a sampling approach in the field was applied in Chechnya in 1995. This sample scheme was originally designed by the World Health Organization's Expanded Programme on Immunization for determining the effectiveness of smallpox vaccine coverage (Bennet, Woods, Liyanage, & Smith, 1991), then modified during Hurricane Andrew to determine needs assessment (Hlady et al., 1994) and finally applied in Chechnya by a British non-governmental organization (NGO) again for needs assessment (Drysdale, Howarth, Powell, & Healing, 2000). In all of the above approaches a grid is placed over the study area. In Grozny, Chechnya, the goal of the NGO was to find whether different sections of the city had adequate water, medical supplies, and so forth. The relief team did not have the resources to knock on every door, so the city was split into 240 grids. In every other grid, seven houses were sampled in order to determine a needs assessment. In this way a relatively small team could cover a large geographic area and perform a needs assessment.

Aggregation Effects

The final note of caution comes with a consideration of numerators and denominators, and the effect aggregation plays on data analysis and visualization. An old epidemiologist once told me, numerators are of limited importance, it is the denominator that counts. Most of the time this makes a great deal of sense. What does a cluster of infant deaths really mean? Ten infant deaths located within 0.25 miles is obviously a hot spot, or is it? Not if a million births also occur in the same area. Some way is needed to identify whether this anomaly is really a cluster. The simplest way is by creating a rate surface. Benchmark rates for infant mortality are known (such as those stated in the Healthy People 2010 objectives); therefore, if a neighborhood exceeds what is expected, it could be considered a cluster. Further insight can be gained if the resulting map is classified by standard deviations, as was explained in Chapter III. This approach will identify areas of the city, whether they are zip codes or census tracts, that exceed the city-wide average rate by two, three, or more standard deviations. Presuming the distribution is normal, a probability of achieving the rate found in each of these classifications can also be calculated. For example, less than 3% (2.28 to be precise) of neighborhoods should be above a rate which is two standard deviations beyond the mean. Such neighborhoods could be targeted for outreach with a reasonable amount of "confidence."

However, a problem can arise with the way the geography of the city is carved up (remember Figure 6 in Chapter II). In the next chapter a technique will be presented (the spatial filter) that removes the effect of political boundaries and instead creates a smooth rate surface.

It should also be remembered, as briefly mentioned in Chapter II, that a spatial subset of the ecological fallacy known as the Modifiable Area Unit Problem (MAUP) can impact analysis results. Simply put, the outcome of a spatial investigation may depend on the way the data was initially aggregated and input into the model (Harris & Longley, 2004; Longley & Harris, 1999). This aggregation effect can occur because of different patterns being revealed due to the scale (the different aggregations of space as shown in Figure 6 in Chapter II) or the zone (the way those aggregations are carved up, such as thin strips or as squares). In order to counter this problem, it is always useful to perform analyses with more than one aggregation of space (such as by census tract and block group), though changing the physical shape of the aggregated area is far harder. These manipulations are often limited by the number of observations falling inside each unit, and the nature of the variable under investigation. For example, an analysis of infant deaths would probably have to be performed at the census tract level, as relatively few blocks would contain any deaths.

Without a large enough denominator value, it is hard to make a judgment about the significance of a cluster (as was mentioned in Chapter I, this being the law of small numbers). Many techniques, such as the spatial filter, allow a minimum denominator value to be included (for example, 40 cases). In this way, even if a mortality cluster is found, if fewer than 40 births occur in the specified neighborhood, no rate is calculated. The thinking behind this is as follows: If a house has a single birth, and that baby dies, the resulting rate would be 1,000/1,000 (notice how this law of small numbers is also an issue of space as well as denominators — the smaller the neighborhood, the fewer denominators). As another example, if a house burns down, killing an infant, and if only five births had occurred within the immediate area, little extrapolation can be made as to the safety of living conditions in the neighborhood.

However, having said that, I always instruct my students to look at their data both before and after analysis. Do not take analysis results as gospel, but instead think about what the analysis might miss.

Three Simple Techniques: Overlay, Density, and a Difference of Proportions Test

Overlay as Analysis

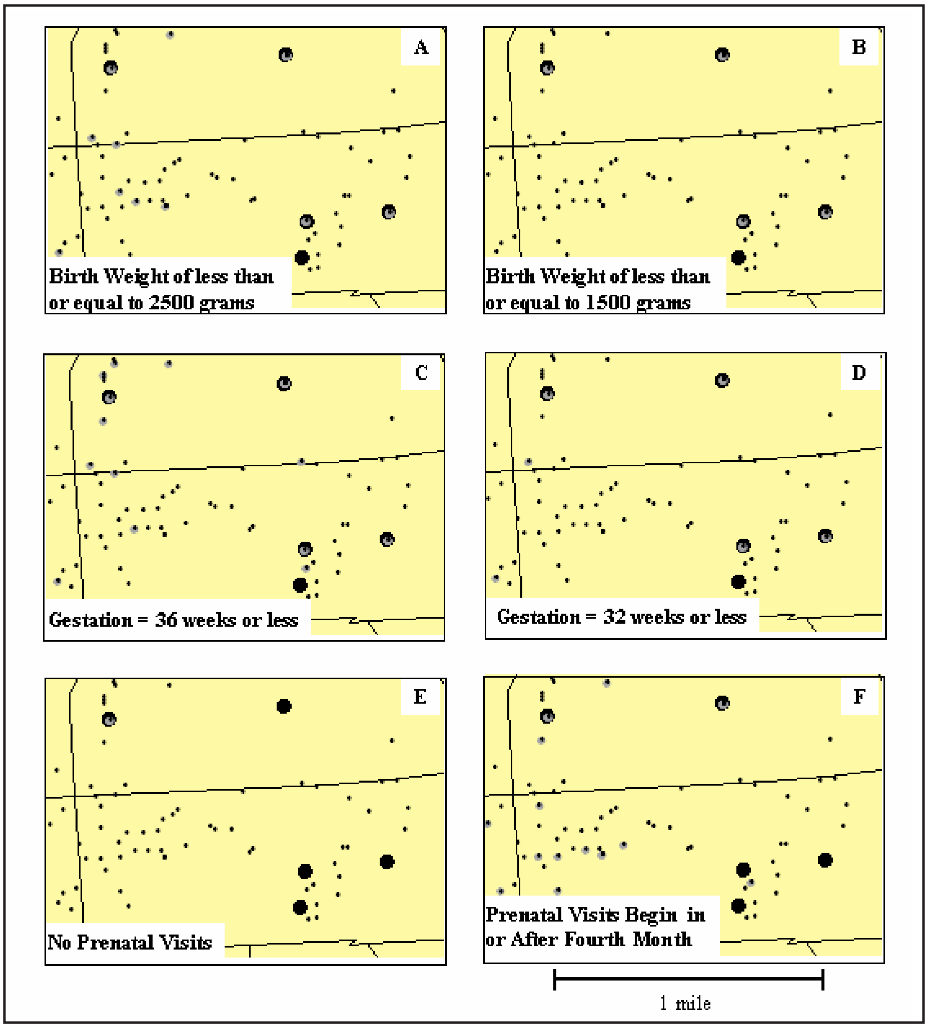
Let us return to the work of John Snow, who used a spatial overlay to identify a spatial pattern, in his case cholera cases and well locations. This approach is now made immeasurably simpler by a GIS. Spatial layers can be added as a preexisting GIS coverage, input as coordinates, or heads-up digitized. Further spatial information can be scanned in and georegistered. Once in the GIS, the layers can be compared against each other, as spatial associations between the data sets are identified. The data is overlaid onto each other, with space being the connection.

A similar approach uses a single data set, but by using queries multiple spatial layers are extracted from its attributes table, in effect turning a single coverage into multiple maps. For example, a birth layer could generate maps of birth weight, mother's age, and APGAR scores. These maps can be overlaid and compared for similarities. Consider the maps displayed in Figure 2, which show five infant death locations and all infant births for one area of the city. It is possible to gain a visual appreciation for pattern by simply querying each of these death locations for known risk factors, such as preterm (and very preterm) delivery or short periods between pregnancies, and displaying these on the maps. This form of overlay can be achieved as follows. The infant death layer is copied. The size of the circle representing the infant death is increased and darkened. The second copied layer is now queried according to one of the risk factors. If an infant death was also a preterm birth, a smaller gray circle will appear in the center. If the death was a preterm delivery and there was a short period between pregnancies, a bull's-eye effect will result.

Figure 2a and b displays this association for both low-birth-weight (less than 2,500 grams) and very low-birth-weight (less than or equal to 1,500 grams) deliveries. From these maps it is possible to see that four of the five infant deaths were very premature. If we consider all births in the area displayed, while every very low-birth-weight baby died in its first year of life, seven low-birth-weight babies survived (Figure 2b). A variable highly correlated to low birth weight, unsurprisingly, is preterm (36 weeks or less) and very preterm delivery (32 weeks or less). The association between premature births and infant mortality is also well documented and can be seen in Figures 2c and 2d. Inadequate or no prenatal care is also considered a risk factor for future infant health. Figures 2e (no prenatal visits) and 2f (prenatal visits commencing in the fourth month of pregnancy or later) do not show such a strong visual association with infant deaths as the previous maps. Only one of the five deaths was to a mother who received no prenatal care (the only mother to receive no care in this study region). Two of the infant deaths were to mothers who received prenatal care starting in the second trimester. This map (Figure 2e) also shows a possible spatial pattern with mothers in the bottom left of the study region starting prenatal care at a later than advised date. One interpretation of these maps would be that a mother residing in this study area who received no prenatal care should be monitored closely after giving birth.

By looking at these maps we can also see an apparent cluster of infant deaths in the bottom right corner of the map. Is this a significant cluster — that is, did it appear by chance, or is there something special in this area? If we can be certain there is indeed a cluster, then we can search for a causation and solution. These are the types of questions that can be asked and answered by linking and analyzing data spatially. Any of the other data fields found on the birth certificate (such as birth weight, race of mother, etc.) can be added to the map in the same way. In addition, other primary (personal interviews) and secondary data (locations of drug arrests, census collected information) can be "layered" onto the map. This visual "layering" of data can help suggest other areas of investigation. This simple GIS approach certainly meets the mission statement of the Maternal and Child Health Bureau (MCHB) directive, paraphrasing the continuing goals which are to improve information collection and analysis, to help effective problem solving, and identify areas of disparity.

Figure 2. Risk factors traditionally associated with infant mortality



Mapping approaches such as these can also be used to display multiple queries (show all births that were very low-birth-weight, and to mothers who had a previous infant death), though the investigator should be careful, as this approach could "lose" a mother if she did not meet both criteria in the query. An alternative approach would be to compare each map as part of a series and to see which death locations continually queried as positive for a risk factor. By taking this approach, four immediate patterns can be seen from the maps. Of the five deaths in the study regions, one was to a mother who was identified as having eight risk factors (very low birth weight, very premature, late starting prenatal visits, single, below 18 years of age, substance use, with a previous birth within 2 years, and a previous infant death). The one mother receiving no prenatal care also had an infant death.

A Cautionary Tale

An error has been included in the previous set of maps to illustrate how even this simple use of GIS must be applied with caution. It is possible that more than one birth occurred at a residence. This is the case for the mother exhibiting eight risk factors. In actuality, this is an apartment building where 10 births occurred in the year under investigation. These 10 births generated the risk factors, not to one single mother, but risks occurring in combination for all mothers. In other words, when displaying any one risk factor, at least one mother residing at this address met the query criteria. This type of mistake is easily corrected for in the GIS by clicking on any birth location to see how many underlying birth certificate records are generated. This mistake was left for two reasons. First, to illustrate a potential error that can be made. Second, the mistake, once corrected for, resulted in a different flag being raised. Of the 10 births born at this location, two died in the first year of life. The GIS has now helped identify a location within a neighborhood that warrants further investigation due to the presence of two deaths (to different mothers) in 10 births.

And so we return to the issue of denominator size. In earlier discussions in this chapter, the point was raised that a pattern of numerator events (infant deaths) made little sense unless compared to denominator events (total births). In order to remove the problem of small numbers, many techniques require a minimum denominator number. Yet such a requirement would have missed this apartment building. Indeed, if the building was relatively isolated with no other births occurring close by, the deaths may have been removed from the analysis. Are two deaths a significant cluster? Do we need such a test of significance to be alarmed by this event?

A solution is to investigate the data both before and after an analysis. One suggestion is to take the data and sort by address. This would extract out multiple birth addresses, such as apartment buildings. These could be investigated separately for worryingly high risks or death occurrences within their limited populations before being replaced in a global analysis. A second approach would be to perform a density analysis based only on the numerators.

Density Analysis

Most GIS packages allow for a density surface to be generated for point data (such as infant death locations). This approach helps “smooth” out individual data to reveal areal patterns (Harris, 2003). The resulting surface is usually color coded to display areas of low events and “hot spots”. The traditional form of this calculation is the total number of events (such as infant deaths), within a window (usually a circle), divided by the area of that window, and then expressed for a standardized area. For example, one death occurring in a 100-meter window would be calculated in the following way:

Density = n / area of the search circle

Therefore 1 / (3.142 * 10,000) = 0.0000318

If this density is expressed as square kilometers, the final density equals 31.83.

The interpretation of this value is that for every square kilometer, one would expect almost 32 deaths. For most GIS packages the density surface is created by turning the “map” into a grid, with each cell of this grid having a calculated density. The resulting raster surface can be queried for any single cell by simply clicking on it. Figure 3 shows a (kernel) density surface of infant mortality deaths in Baton Rouge for three consecutive years. The kernel window is 2 miles in diameter. The surface has been classified by standard deviations, with those areas being three standard deviations or higher colored yellow. By comparing the three maps, an area of high infant mortality is evident within the inner city area of Baton Rouge. This central area of consistent high infant mortality coincides with the program region for the Baton Rouge Healthy Start. Other “hot spots” are more transitory, though a southern area of concern has developed over the last 2 years of the analysis.

The size of the window in the density analysis plays an important role in the final pattern. At one extreme, if the window size is equivalent to the entire study area, the density calculation will be for all data (number of deaths/total study area) and the resulting surface will be flat. At the other extreme, if the window only has a radius of 10 meters, the resulting surface will be polka dot, a series of light intensity colors occurring at the death locations sitting on a darker background. The trick is to have a window which is large enough to reveal underlying density patterns. The correct size of these windows, or “kernels,” in spatial analysis has generated a great deal of debate in the literature (Harris & Longley, 2004; Martin, 2002). The underlying urban structure will always play a role in determining this window size. Window sizes in northeastern inner-city areas will be smaller than southern suburban areas, as the population is more densely packed together. It is prudent to perform all analyses with multiple window sizes so as not to be drawn in by a pattern that might be tied to a specific window size.

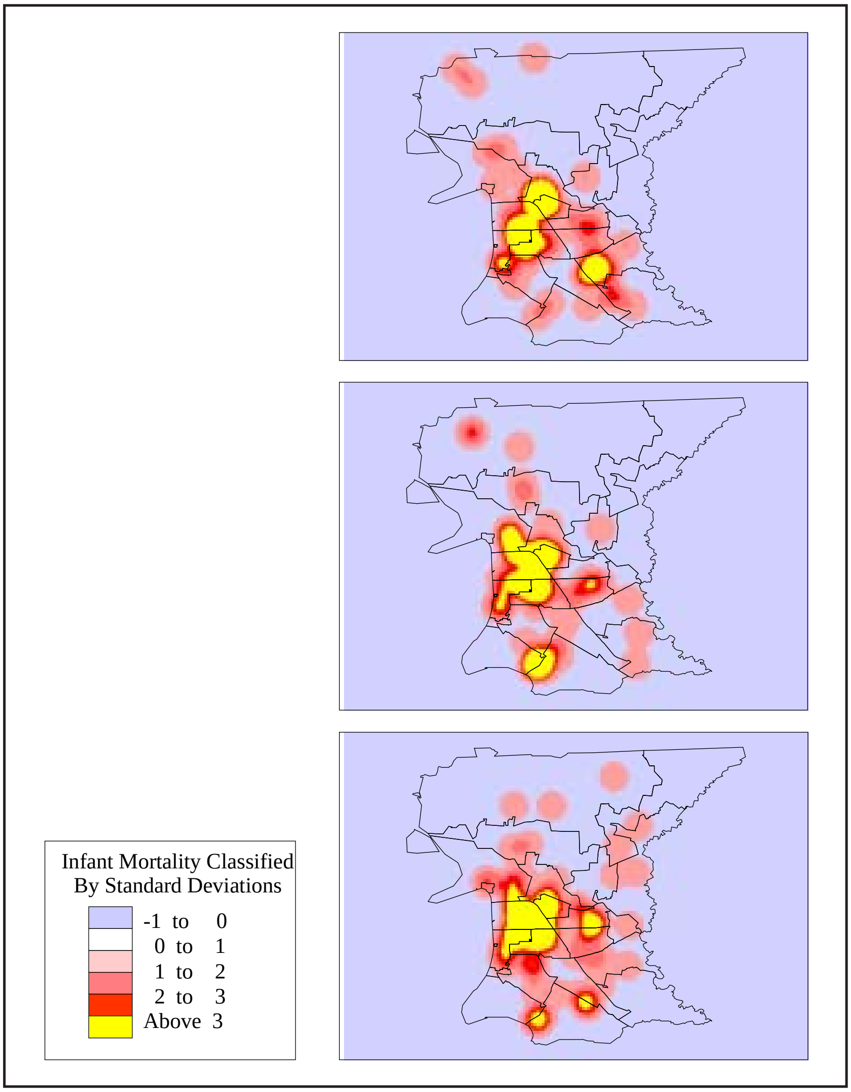
Density calculations can also be "weighted," meaning the single value occurring in space (the death residence) can be increased by an associated attribute. The simplest form of weight would be the number of deaths; in the previous example for the apartment building with two deaths, the weight would be "two."

The density surface shown in Figure 3 is in effect a raster coverage; the GIS screen becomes a matrix where each cell has had a density calculated. Two further manipulations are needed to transform these findings into a useful, interpretable GIS screen. A problem with this type of analysis is, again, that (usually) only numerator events are present. Therefore, it is useful to compare the densities against the entire mortality surface. This is achieved by classifying the density surface into standard deviations. Ideally, the classification should be normally distributed; therefore, those areas of the map that are three standard deviations or higher could be considered to be "areas of concern" (it is better not to refer to them as "significant" at this point). The density surface can be contoured, also an option available in most GIS packages, with the interval between contours being set as one that can be used to approximate the standard deviation classifications. For example, if one standard deviation equals 15, the contour interval should be also set at this. The resulting contour map can be queried out to identify those sections of the map being three standard deviations or higher (at or above 45). These contour "hot spots" can be transferred to a city map, such as a DOQQ. Alternatively, the density analysis can be performed on centroids of neighborhoods (or of census tracts, for example) where a rate has already been calculated. In this way, both numerator and denominator are included in the density calculation.

A second type of density calculation, often available in a GIS, is a kernel density analysis. This type of analysis will be described in more detail in Chapter VI; however, it can be applied in the same way as the simple density calculation presented here. Indeed, Figure 3 is actually a kernel density surface. The major difference between the two calculations is that in a simple density all points, irrespective of where they fall inside the window, have the same weight in the calculation (the weight here does not refer to multiple events occurring at the same location, just the impact a coordinate location has on the final calculation). In the kernel density, the concept of distance decay is introduced, that although all events occurring within the 2-mile window will be included in the calculation, those occurring closer to the center of the window will have more impact on the calculation than those towards the outer 2-mile bound.

Of course, this approach to finding a hot spot is crude and better techniques are available, some of which will be described in Chapter VI, though these usually require additional software and expertise. The density calculation described here is available in most GIS packages and can be easily implemented to create a reasonable first-time risk surface.

Figure 3. Kernel densities of infant mortality for 3 consecutive years in East Baton Rouge Parish



Difference of Proportions Test

This chapter has so far presented two techniques that can be used to gain a spatial impression of point data. This last section allows for areas or subsets of a population to be analyzed based on a simple t-test. Any elementary statistics text will explain how t-tests are constructed, and most importantly, what violations the investigator should be wary of. The simple explanation of these tests is for a subset population to be tested against the whole population in order to see if it differs *significantly*. For example, are births surrounding one clinic better or worse off than births for the whole city in general? Similarly, are mothers under the age of 17 more vulnerable to a particular birth risk, such as premature delivery, when compared to the whole birth surface? A GIS allows us to combine these queries to ask both spatial and aspatial questions. Are births to teen mothers around one clinic different from the whole birth population? When using these tests, consideration must be paid as to what is the reference population. For example, if we are comparing births to teen mothers around one clinic, should the reference population be the entire birth population of the state, the county, the city, just teen mothers, or if the investigated population is known to be indigent, to births from similar economic backgrounds? Obviously, the choice of this reference population will affect results. For most analyses in East Baton Rouge Parish, the entire parish is used as the reference population, though sometimes the reference population is limited to the five zip code region served by the Baton Rouge Healthy Start. This population is considered to be "at risk," so any sub-group found to be significantly different within this area truly presents a cohort of concern. The formulation of the test will be presented momentarily, though it is first useful to describe the type of question that can be posed and answered using this approach.

As was discussed in Chapter IV, there is an array of personal and neighborhood risks that can exhibit a spatial footprint on a map. One might think mapping economically disadvantaged neighborhoods in a city is the first step in reducing prenatal risk. Unfortunately, such a shotgun approach does not work for two reasons. The first is that pregnancy intervention programs do not have the finances to tackle all possible risks; therefore, the risks need to be prioritized. The second reason is more general, in that the risk surface is spatially complex with apparently similar economic status neighborhoods experiencing different combinations of risks; Neighborhood X may differ from Neighborhood Y in terms of risk experience, even if they are economically and racially similar. In other words, each neighborhood will experience its own unique risk pattern. It is important for any community-based health unit to understand the particular needs of the neighborhoods it serves so that the correct outreach strategies can be developed.

The t-test provides a simple means to test these differences. Again, the reader is referred to introductory statistics texts for the background of this test. The most common version compares means between the sample and the population. However, this is often not the best variant for comparing birth neighborhoods, because this measure often masks data thresholds of concern. For example, what does the average birth weight tell us for a neighborhood? Not much—we need to know what proportion of births is below a known risk level. It is for this reason that the difference of proportions test is applied. In effect, this test compares the proportion of births beneath a set threshold (1,500 grams for very low-birth-weight deliveries) for both the sample and the population. In simple terms, the test will tell us whether the sample area has a significantly higher proportion of low-birth-weight deliveries than the reference population.

The following equation can be used for the calculation of this test:

Equation: $$Z = \frac{P - p}{\sigma_p}$$

where $$Z = \frac{P - p}{\sqrt{P(1- P) / n}}$$

Z = sample statistic

P = sample proportion (for example proportion of low-birth-weight births in buffer range)

p = population proportion (for example proportion of low-birth-weight births in East Baton Rouge Parish)

σ_p = standard error of the proportion

The input data can be extracted from the GIS by a query and then exported to a spreadsheet or database in which the equation can be easily constructed. As a sample analysis to show how this technique can be applied, birth and death data for three time periods and two different risk populations (all births and mothers of younger than 18 years of age), for neighborhoods falling under the sphere-of-influence of three Early Head Start centers in Baton Rouge, were chosen for analysis.

The following data were included in the analysis:

1: Number of births

2: Number of deaths

3: Infant mortality rate (IMR)

4: Proportion of women not naming a father on the certificate (seen as a measure of a future lack of male parental involvement) (Men = 0)

5: Proportion of African American mothers (Black)

6: Proportion of mothers being 18 or younger at time of birth (U18)

7: Proportion of unmarried mothers (Single)

8: Proportion of babies weighing less than 1,500 grams (classified as a very low-birth-weight baby) (less 1,500)

9: Proportion of babies weighing less than 2,500 grams (classified as being low-birth-weight) (less 2,500)

10: Proportion of mothers making seven or less prenatal visits (less 7 visits)

11: Proportion of mothers receiving less than an 11th grade education (less 11th)

12: Proportion of mothers making a first prenatal visit in the fourth month or later (After 4th)

13: Proportion of mothers giving birth in the 32nd week of gestation or less (classified as a very premature birth) (less 32)

14: Proportion of mothers giving birth in the 36th week of gestation or less (classified as being a premature birth) (less 36)

15: Proportion of mothers who had previously had a live birth (Prev moth)

16: Proportion of mothers who had had a child die since birth (Prev death)

17: Proportion of mothers having previously given birth within the preceding 2 years (Within 2)

18: Proportion of mothers who self-reported smoking during pregnancy (Smoke)

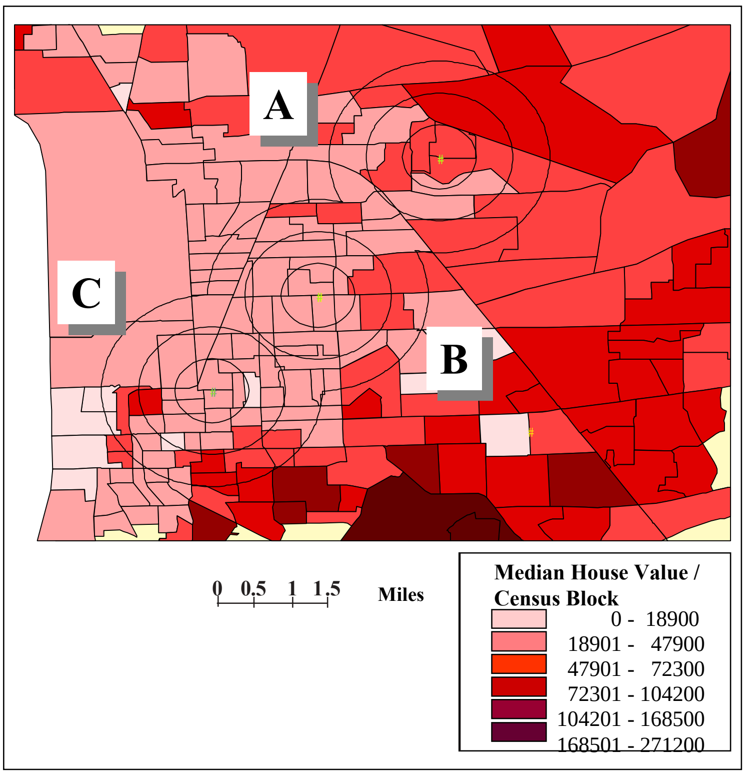
19: Proportion of mothers who self-reported alcohol use during pregnancy (Drink)

Hopefully the reader is now becoming familiar as to the role space plays in most calculations and investigations performed using the GIS. The size of the sample population will play a role in the difference of proportions test. If the driving question behind this analysis is the effectiveness of the Early Head Start Center, the first question is determining the “neighborhood” of that center (remember back to the discussions in Chapter III as to how size can affect results). In this case the neighborhood is defined by how far a woman might travel for a visit. Is it 0.5 miles, 1 mile, or more? A further consideration is that the sample population must have a minimum number of subjects, which traditionally is set at being no less than thirty. Therefore, this sphere-of-influence will be forced outwards depending on the surrounding environment. An inner-city environment will probably capture the required number of births within 0.5-miles, but this distance will have to be extended for suburban and rural areas (which makes sense due to families probably owning their transport in these areas).

For the following Baton Rouge example, three spheres-of-influence were used with radii of 0.5, 1 and 1.5 miles extending out from each center, each of which is located in the "at-risk" Healthy Start zip code area of the city. The sphere-of-influence was created by a buffer, which for this analysis was a concentric circle being drawn around each center at distances of 0.5 miles, 1 mile, and 1.5 miles. All birth and death locations falling inside the buffer were extracted for analysis. Figure 4 displays the three Early Head Start centers. The map also displays the median building value by census tract. This value can be used as a proxy for the relative wealth of the neighborhoods within range of each center. From this figure it is evident that the three centers are proximate to one another, as the 1.5-mile radius buffer bleeds together for centers A and B, and B and C. The median house value also shows that centers B and C have similar housing value surfaces, with neighborhoods around center A being slightly more prosperous.

Tables 1 to 3 display the actual counts of prenatal risks for each of the Early Head Start centers at each of the buffer ranges. The values in bold are statistically significant at the 0.05 level when compared to all births for East Baton Rouge Parish for that year.

Figure 4. The location of the three Early Head Start Centers



Results for Year One (Table 1)

Although Center A contains neighborhoods with higher median household values, it also has the higher infant mortality rate (IMR) for year one. Indeed, at all three ranges (0.5 miles, 1 mile, and 1.5 miles) the IMR does not drop below 30 per 1,000. Center B only exceeds the parish level IMR at the 0.5-mile range. Center C exceeds it for both 1.5 and 1 miles, yet this drops to 0 for the 0.5 miles range due to the relatively low number of births. The areas around each of the three centers are largely African American. The proportion of pregnancies to unmarried women is significant for all centers at all ranges except 1.5 miles around Center B. Center C at all ranges, and Center A at the 1.5- and 1-mile ranges, also have a significant proportion of women not recording information about the father on the birth certificate. There is a significant proportion of U18 mothers at all centers for the 1.5-mile range, and additionally at the 1-mile range for Center B. Center C has a significantly high proportion of women having seven or fewer prenatal visits within the 1.5-mile range. All three centers have a significantly high proportion of women having their first prenatal visit in the fourth month or later in at least one of the buffer ranges. The same pattern is found with the proportion of mothers receiving an 11th grade education or less.

Results for Year Two (Table 2)

The IMR for Center A is again considerably higher than the parish total at all three scales. The IMR for Center B is also higher than the parish level at the 1.5- and 1-mile ranges, though no deaths occur within the 0.5-mile range. The IMR only exceeds the parish level at the 1.5-mile range for Center C. As with year one, for all centers there is a significantly high proportion of unmarried African Americans at all ranges. There is a slight difference from year one in the proportion of women not recording information about the father on the birth certificate with both Centers A and B having significant proportions at the 1.5- and 1-mile ranges, while Center C has significant proportions at all ranges. The pattern for U18 mothers almost mirrors the pattern found for year one, with Centers B and C having significant proportions at the 1.5-mile range. Temporal stability is also found in terms of the proportion of women receiving education up to or including the 11th grade, and proportion of women commencing prenatal care in the fourth month or later. For both of these "risks," the distribution across centers and ranges is the same for both years, with the exception of Center C also becoming significant at the 0.5-mile range for education. The proportion of prenatal visits being seven or less rises in year two, with Center B and C being significant at the 1.5- and 1-mile ranges, whereas significance was only found for Center C at 1.5 miles in year one. Two related risk variables emerge as

Table 1. Statistically significant prenatal risks at Three Early Head Start Centers (year one)

	Births	Deaths	IMR	Men=0	Black	U18	Single
A 1.5	344	11	31.98	**145**	**328**	**68**	**238**
A 1.0	231	7	30.30	**90**	**225**	39	**154**
A 0.5	84	4	47.62	36	**82**	19	**61**
B 1.5	630	5	7.94	335	**596**	**127**	476
B 1.0	287	3	10.45	148	**266**	**70**	**214**
B 0.5	89	1	11.24	39	**86**	23	**60**
C 1.5	522	7	13.41	**282**	**454**	**95**	**393**
C 1.0	242	3	12.40	**145**	**232**	38	**193**
C 0.5	57	0	0.00	**35**	**56**	9	**47**
Parish	6142	65	10.58	1593	3008	658	2590
	less 1500	less 2500	less 7 visits	less 11th	After 4th	less 32	less 36
A 1.5	17	49	62	**104**	**93**	20	53
A 1.0	11	26	37	61	64	12	29
A 0.5	5	12	11	27	24	6	15
B 1.5	28	103	133	**216**	**195**	42	109
B 1.0	8	43	58	94	86	14	49
B 0.5	0	7	14	25	21	2	11
C 1.5	23	75	**115**	**182**	**168**	23	79
C 1.0	8	36	54	**90**	**76**	9	39
C 0.5	0	10	12	23	15	1	11
Parish	129	624	765	1225	1113	206	672
	Prev moth	Prev death	Within 2	Smoke	Drink		
A 1.5	194	6	86	32	5		
A 1.0	131	2	58	19	2		
A 0.5	40	1	21	4	1		
B 1.5	360	19	154	38	9		
B 1.0	159	11	65	16	3		
B 0.5	46	2	19	3	1		
C 1.5	296	6	139	46	12		
C 1.0	147	1	73	21	8		
C 0.5	34	1	19	8	1		
Parish	3534	113	1405	420	67		

Distances for centers = 0.5, 1, and 1.5 miles

Statistically significant prenatal risks at the 0.05 level are displayed in bold.

significant proportions for the first time in year two. There is a significant proportion of women having low-birth-weight babies at the 1.5-mile range for both Center A and B. The related variable of short gestation is also significant at the same range for Center B (and interestingly also at the 1-mile range) but not for Center A.

Table 2. Statistically significant prenatal risks at three Early Head Start Centers (year two)

	Births	Deaths	IMR	Men=0	Black	U18	Single
A 1.5	337	11	32.64	**142**	**320**	52	**251**
A 1.0	233	9	38.63	**98**	**228**	31	**175**
A 0.5	82	2	24.39	29	**81**	11	**60**
B 1.5	595	10	16.81	**283**	**565**	**123**	**461**
B 1.0	296	5	16.89	**128**	**276**	54	**223**
B 0.5	80	0	0.00	32	**70**	9	**56**
C 1.5	481	6	12.47	**230**	**417**	**92**	**357**
C 1.0	231	2	8.66	**121**	**225**	43	**183**
C 0.5	61	0	0.00	**33**	**60**	13	**48**
Parish	6072	67	11.03	1452	3045	602	2596
	less 1500	less 2500	less 7 visits	less 11th	After 4th	less 32	less 36
A 1.5	24	**68**	61	**93**	**97**	29	69
A 1.0	20	50	44	54	**68**	23	52
A 0.5	4	20	17	12	22	5	18
B 1.5	26	**106**	**146**	**215**	**186**	34	**126**
B 1.0	6	46	**72**	**103**	**94**	13	**67**
B 0.5	0	15	18	25	25	2	21
C 1.5	9	73	**114**	**170**	**147**	21	75
C 1.0	5	40	**63**	**85**	**85**	14	37
C 0.5	2	9	16	**27**	22	3	9
Parish	140	658	742	1107	1079	220	726
	Prev moth	Prev death	Within 2	Smoke	Drink		
A 1.5	205	13	82	18	8		
A 1.0	141	11	52	9	5		
A 0.5	44	5	22	2	0		
B 1.5	340	13	152	51	10		
B 1.0	179	8	91	21	2		
B 0.5	52	0	24	8	2		
C 1.5	274	16	123	39	7		
C 1.0	143	9	63	20	3		
C 0.5	39	3	16	9	1		
Parish	3530	136	1450	327	47		

Distances for centers = 0.5, 1, and 1.5 miles

Statistically significant prenatal risks at the 0.05 level are displayed in bold.

Results for Year Three (Table 3)

The IMR for Center A continues to be considerably above the parish level in year three, though this difference is not as pronounced as in the previous years. Of concern is that the IMR for Center B and C is more elevated in year three, especially for Center C where the IMR exceeds the parish level at all ranges.

Similar patterns to the previous 2 years are found in the significant proportions of African American, unmarried, women of 18 years and under, low education, and receiving the first prenatal visit in the fourth month or later. The pattern of women receiving seven or fewer prenatal visits is also similar to the preceding 2 years, though for the first time a significant proportion is also found at Center A at the 1.5-mile range. Birth weight and period of gestation do not show as a significant category for any center at any range in year three.

There are drawbacks in using cut-off significance levels to drive any community-based action. For example, is 0.05 an appropriate level of significance? Should it be more or less? Is the population (East Baton Rouge Parish) the correct reference population used in the difference of proportions test? Should the reference population be only those from the at-risk area? It is therefore also important to consider the raw data and the risk proportions within each category as the goal should be to reduce these levels whenever possible. The statistical level of significance does, however, allow for a prioritization of risk. Of course, this assumes all risks are a) of equal impact, and b) can be reduced and are not systematic of a larger structural issue (for example, how do you promote marriage?).

Outreach initiatives might be better targeted on risks that are temporally stable; that is, they appear in the neighborhoods year after year (this issue of temporal stability will be revisited in Chapter VII). For example, temporal stability is achieved in both the age of the mother, and the level of education for Centers B and C. This suggests that a high proportion of girls are becoming pregnant in these areas, and therefore programs might be targeted on abstinence, early pregnancy, and the benefits of staying in school. Temporal stability also occurs with the number of prenatal visits for all centers, and month prenatal visits began for Center A. This could be systematic of poor education (the women not understanding the need for prenatal care), poor infrastructure (the lack of access to prenatal centers, though we are analyzing an Early Head Start Center), a poor prenatal care interaction experience (a perception of racism at the clinic), or in terms of late entry, the prolonged application procedure associated with Medicaid. Results from this analysis would need follow-up interviews to ascertain which reason held so measures could be implemented (such as Presumptive Eligibility, which is a Medicaid fast track that provides a temporary card in 7 to 10 days). Interestingly enough, Center A, which appears to have fewer economically disadvantaged neighborhoods in its range, does have fewer statistically significant categories of risk than the other two centers, and yet it also has the highest IMR for all 3 years.

Table 3. Statistically significant prenatal risks at three Early Head Start Centers (year three)

	Births	Deaths	IMR	Men=0	Black	U18	Single
A 1.5	335	7	20.90	**129**	**311**	53	**239**
A 1.0	215	5	23.26	71	**211**	31	**155**
A 0.5	73	2	27.40	24	**72**	15	**56**
B 1.5	637	15	23.55	**282**	**614**	**132**	**491**
B 1.0	290	5	17.24	**114**	**281**	**60**	**217**
B 0.5	85	1	11.76	**39**	**84**	15	**65**
C 1.5	512	10	19.53	**224**	**451**	**107**	**386**
C 1.0	242	4	16.53	**115**	**232**	50	**196**
C 0.5	59	1	16.95	26	**58**	10	**47**
Parish	6058	81	13.37	1350	3067	639	2575
	less 1500	less 2500	less 7 visits	less 11th	less 4th	less 32	less 36
A 1.5	13	55	**71**	**110**	**96**	17	62
A 1.0	11	40	45	58	**66**	12	42
A 0.5	2	10	13	19	27	2	10
B 1.5	25	96	**154**	**250**	**207**	32	110
B 1.0	5	41	**72**	**99**	**88**	9	44
B 0.5	2	13	22	30	22	4	13
C 1.5	24	83	**120**	**206**	**157**	33	92
C 1.0	11	38	57	**111**	**73**	15	40
C 0.5	3	7	13	23	20	3	8
Parish	170	645	726	1166	1076	223	695
	Prev moth	Prev death	Within 2	Smoke	Drink		
A 1.5	211	11	104	22	3		
A 1.0	138	8	68	10	2		
A 0.5	40	4	19	1	0		
B 1.5	383	13	170	44	5		
B 1.0	171	5	75	17	1		
B 0.5	49	2	24	7	0		
C 1.5	300	8	150	45	6		
C 1.0	155	4	75	24	3		
C 0.5	36	1	14	6	0		
Parish	3545	115	1435	357	39		

Distances for centers = 0.5, 1, and 1.5 miles

Statistically significant prenatal risks at the 0.05 level are displayed in bold.

Under-18 Pregnancies (Table 4)

As was previously mentioned, both changing the sample under investigation, and the reference population can modify the difference of proportions test. For example, an aspatial query can be added to the sphere-of-influence analysis so that only births to teenage mothers are analyzed. This new sample is then compared to a new population, teenage births for the entire parish. Unfortunately, this selection results in a far smaller number of observations, so the sphere-of-influence can only include the 1.5-mile buffer. Due to the fact that the comparison population would traditionally be considered "at-risk," none of the centers produced any significant differences at the 0.05 level. This can be interpreted that for this vulnerable cohort, the area surrounding the centers produced no elevated areas of concern, at least at this level of significance, for this cohort. For an analysis such as this where the investigated cohort has been so reduced, it is often better to just consider the raw numbers contained in Table 4.

The most striking aspect from these data is the IMR for year three. In year one none of the centers exceed the parish IMR for births to U18 women. In year two, only Center C exceeds the parish IMR, and then barely. However, in year three, all centers exceed the parish level, both Center B and C considerably so. Indeed, of the 12 deaths that occurred in the parish, five occurred within 1.5 miles of Center C. For such a short study period these spikes might be random; however, it might also be prudent to further investigate the death certificates for cause of mortality.

For all three time periods, Center B has a higher percentage of U18 women delivering very low-birth-weight babies than for the same cohort in the parish. This elevated level is also seen for the same center with low-birth-weight babies, with years two and three again considerably exceeding the parish level. Potentially more worrying is the considerable increase in the percentage of U18 women giving birth to low-birth-weight babies in Center C's area: 25% in year two, and over year three in 1998. Obviously, this risk would have to be watched closely to see if a trend were emerging. Short gestational pregnancies generally mirror the pattern of low birth weight (as would be expected), though over 27% of U18 women around Center C experienced short gestation pregnancies in year three.

Both Center B (all 3 years) and C (years two and three) have higher percentages of women receiving seven or fewer prenatal visits than the same cohort in the rest of the parish. For Center B in years two and three, around 30% of pregnant women in this cohort received only seven prenatal visits or fewer. Similar percentages were found for Center C. The point at which prenatal care was initiated provides an even more worrying finding, with percentages edging toward 50% for Centers B and C. Although each of the three centers had years when 25% or more of the U18 mothers had previously given birth, Center A achieved the highest overall percentage in this category, reaching over 30% in year three.

The results from this analysis have shown how spatially complex prenatal risk patterns are for different neighborhoods (in this instance, the neighborhood being defined as the sphere-of-influence surrounding a clinic). Not only do risks vary between centers, but also in the range used to determine the sphere-of-influence

Table 4. Statistically significant prenatal risks of U18 women at three Early Head Start Center in Baton Rouge

	Births	Deaths	IMR	less 1500	%	less 2500	%	
A Year 1	68	1	14.71	3	4.41	8	11.76	
A Year 2	52	0	0	1	1.92	9	17.31	
A Year 3	53	1	18.87	1	1.89	8	15.09	
B Year 1	127	0	0	6	4.72	16	12.60	
B Year 2	123	1	8.13	6	4.88	27	21.95	
B Year 3	131	4	30.53	7	5.34	23	17.56	
C Year 1	107	1	9.35	2	1.87	9	8.41	
C Year 2	92	2	21.74	2	2.17	23	25.00	
C Year 3	107	5	46.73	7	6.54	25	23.36	
Parish 1	658	10	15.2	19	2.89	80	12.16	
Parish 2	602	11	18.27	20	3.32	95	15.78	
Parish 3	639	12	18.78	21	3.29	102	15.96	
	less 7 visits	%	After 4th	%	less 32	%	less 36	%
A Year 1	12	17.65	23	33.82	3	4.41	9	13.24
A Year 2	8	15.38	16	30.77	2	3.85	8	15.38
A Year 3	11	20.75	19	35.85	2	3.77	11	20.75
B Year 1	32	25.20	52	40.94	8	6.30	18	14.17
B Year 2	39	31.71	53	43.09	8	6.50	29	23.58
B Year 3	37	28.24	58	44.27	7	5.34	28	21.37
C Year 1	21	19.63	51	47.66	3	2.80	9	8.41
C Year 2	28	30.43	30	32.61	5	5.43	19	20.65
C Year 3	29	27.10	51	47.66	10	9.35	29	27.10
Parish 1	151	22.95	255	38.75	29	4.41	91	13.83
Parish 2	147	24.42	219	36.38	34	5.65	100	16.61
Parish 3	146	22.85	241	37.72	27	4.23	109	17.06
	Prev moth	%	Prev death	%	Within 2	%	Smoke	
A Year 1	17	25.00	0	0.00	16	23.53	4	5.88
A Year 2	8	15.38	0	0.00	6	11.54	2	3.85
A Year 3	16	30.19	2	3.77	13	24.53	1	1.89
B Year 1	19	14.96	0	0.00	17	13.39	2	1.57
B Year 2	31	25.20	2	1.63	26	21.14	5	4.07
B Year 3	31	23.66	1	0.76	26	19.85	5	3.82
C Year 1	17	15.89	0	0.00	16	14.95	4	3.74
C Year 2	23	25.00	4	4.35	17	18.48	1	1.09
C Year 3	25	23.36	1	0.93	20	18.69	4	3.74
Parish 1	115	17.48	0	0.00	96	14.59	36	5.47
Parish 2	124	20.60	7	1.16	102	16.94	33	5.48
Parish 3	121	18.94	6	0.94	101	15.81	38	5.95

(from 0.5 to 1.5 miles). Some risks are common to all centers (such as single mothers), some risks are common across all ranges for a single center (such as education in year two at Center C), and some risks are stable, meaning they occur at significant levels year after year (such as month of prenatal entry at Center B). Hopefully these limited results show how "shotgun" approaches to risk reduction are not likely to be effective when applied to large geographic areas, such as an entire city. It could even (and rightly) be argued that further spatial break down is needed to identify the true neighborhoods within these spherical buffers.

However, insight has been gained, even if it is to verify suspected prenatal risk patterns. Other questions have also been raised, such as of the 97 births to U18 women at Center C in year two, four had previously had a live birth die. For women as young as this, this is quite an alarming proportion and further analysis would be needed to ascertain similarities among causes of death.

The difference of proportions t-test will be revisited again in future chapters, where it will be used in an investigation of indigent mother mobility. This technique, along with overlay and density analysis, can be used in combination to investigate a birth surface without needing to rely on the progressively more powerful (and complex) techniques presented in Chapter VI. Of course, if the investigator has spatial analysis skills then read on, though there is always something to be said for keeping it simple.

References

Bennet, S., Woods, T., Liyanage, W. M., & Smith, D. L. (1991). A simplified general method for cluster sample surveys of health in developing countries. *World Health Statistics Quarterly, 44*, 98-106.

Drysdale, S., Howarth, J., Powell, V., & Healing, T. (2000). The use of cluster sampling to determine aid needs in Grozny, Chechnya in 1995. *Disasters, 24*(3), 217-227.

Harris, R. (2003). Population mapping by geodemographics and digital imagery. In V. Mesev (Ed.), *Remotely-sensed cities* (pp. 223-241). London: Taylor and Francis.

Harris, R., & Longley, P. (2004). Targeting clusters of deprivation within cities. In J. Stillwell, & G. Clarke (Eds.), *Applied GIS and spatial analysis* (pp. 89-110). John Wiley & Sons.

Hlady, W. G., Quenemoen, L. E., Armenia-Cope, R. R., Hurt, K. J., Malilay, J., Noji, E. K., & Wurm, G. (1994). Use of a modified cluster sampling method to perform rapid needs assessment after Hurricane Andrew. *Annals of Emergency Medicine, 23*, 719-725.

Longley, P. A., & Harris, R. J. (1999). Towards a new digital data infrastrucure for urban analysis and modelling. *Environment and Planning B, 26*, 855-878.

Martin, D. J. (2002). Census population surfaces. In P. Rees, D. J. Martin, & P. Williamson (Eds.), *The census data system* (pp. 139-148). Chichester, UK: Wiley.

McGrew, J. C., & Monroe, C. B. (2000). *An introduction to statistical problem solving in geography* (2nd ed.). McGraw-Hill Higher Education.

Pickles, J. (1997). Tool or science? GIS, technoscience, and the theoretical turn. *Annals of the Association of American Geographers, 87*(2), 363-372.

West Wing. (n.d.). Mr. Willis of Ohio (Season One, Episode Six). Written by Aaron Sorkin, Directed by Christopher Misiano.

Wright, D., Goodchild, M., & Proctor, J. (1997). GIS: Tool or science? Demystifying the persistent ambiguity of GIS as "tool" versus "science." *Annals Association of American Geographers, 87*(2), 346-362.

Chapter VI

Advanced Spatial Analysis

The last chapter presented several ideas of how to perform relatively simple forms of spatial analysis. Many of these approaches, though insightful, have been superceded by more advanced analytical techniques. This chapter will present a few of these approaches, namely methods of clustering, interpolation, and spatial association. Other concepts will also be addressed, such as spatial autocorrelation and the measures that can be used to find spatial clusters of significantly high (hot spots) or low (cold spots) values. Two additional cluster methods will also be discussed, these being nearest neighbor hierarchical clustering and the spatial filter. Kernel density interpolation will be introduced as the interpolation method for discrete incident locations. A discussion about spatial regression analysis will conclude this chapter. The analyses and examples shown in this chapter will again be based upon linked birth and death certificate data for East Baton Rouge Parish.

Spatial Autocorrelation

Spatial autocorrelation (SA) is an important concept in spatial analysis. It is directly related to Tobler's First Law of Geography, which has been described in colloquial terms previously but is actually defined as, "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, 1979). As such, spatial autocorrelation describes the similarity of nearby observations, and is used to measure spatial dependency or spatial association between attribute values at a location, and attribute values in the location's neighborhood (Worboys, 1995). Positive SA is evident when similar values of the same attribute lie close to each other within the study area. If similarly high values are close to each other, then this is referred to as a local hot spot or cluster. In contrast, cold spots are regions within the study area where similarly low values cluster together. Negative SA is detected when high attribute values are close to low attribute values, or vice versa (Fotheringham et al., 2002). No SA exists when the attribute values are randomly distributed across the study area. In this case, attribute values are independent of each other (Longley et al., 2001). Such a random distribution of attribute values across space is also referred to as Complete Spatial Randomness (CSR) (Boots & Getis, 1988). Past research has indicated that most human phenomena show positive SA.

Global Spatial Autocorrelation

Spatial autocorrelation can be measured at the global or local level. Global SA identifies spatial patterns for the entire study area, whereas local SA identifies relationships between neighborhood-like attribute values. Two classical measures for global SA exist: Moran's I (Moran, 1950) and Geary's C (Geary, 1954). Both statistics can be calculated for point and area data. For both statistics, either a contiguity or a distance-based matrix (weight) need to be defined. The matrix enters as a weight variable into the SA formula and indicates the spatial pattern in the data. A binary contiguity matrix only consists of zeros and ones. There are different ways binary contiguity can be defined. In the case of the rook contiguity only common boundaries are considered. For example, two census tracts are considered to be neighbors if they share a common boundary and a one is entered into the matrix. If both census tracts do not share a common boundary, a zero is entered. The queen contiguity is an extension of the rook case and identifies neighbors by common boundaries and common nodes (corners). Distance-based spatial weights matrices identify neighbors based on distances between the centroids of the spatial units. In the binary case, spatial units are considered to be neighbors if their centroids fall within a specified distance. In the inverse distance case, the relationship between the spatial units is defined by inverse distance weights (such as 1/distance). In other words, the closer two census tracts are, the higher the weight and vice versa. In general, distance-based matrices are a better and more realistic way to capture the spatial pattern in the data than contiguity-based matrices. But binary, distance-based matrices are less useful when the spatial units vary considerably in size, because when spatial units are small, the distances between their centroids are short. But these distances are too short to connect centroids between large spatial units. This results in isolated (unconnected) observations or islands. For this reason, longer distances have to be chosen to ensure that larger census tracts are connected and have at least one neighbor. By doing so, smaller census tracts have many other smaller census tracts as neighbors.

The Moran's I statistic lies between -1 and +1. If neighboring attribute values are similar (positive SA), then Moran's I lies between 0 and +1 — the more similar the values are, the closer the statistic is to +1. On the other hand, a negative SA is measured with a Moran's I between slightly less than 0 and -1, indicative of neighboring attribute values that are dissimilar. The closer the statistic is to -1, the more dissimilar neighboring attribute values are. A spatial distribution with attribute values independent of each other is expressed with a Moran's I value that is negative but very close to 0. In contrast, Geary's C varies between 0 and 2, with values between 0 and 1 indicating positive SA and values between 1 and 2, negative SA. A Geary's C value of 1 indicates complete spatial randomness. The reader should consult any introductory textbook on (spatial) statistics for a more in-depth discussion on global SA. Most of the spatial analysis software mentioned in Chapter III can calculate both classic measures of SA automatically.

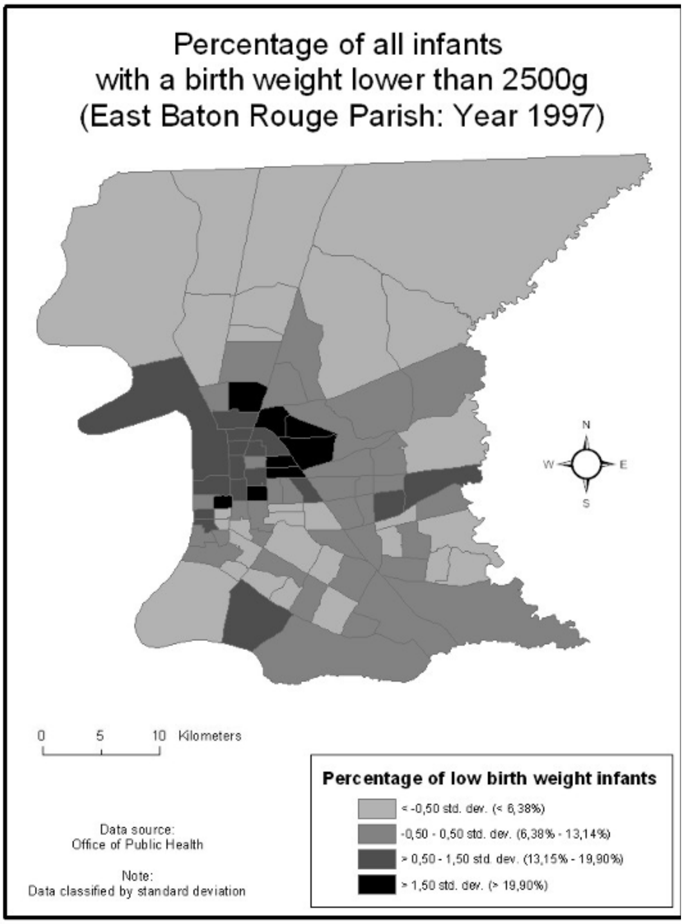
For illustrative purposes, the Moran's I was calculated for the percentages of low-birth-weight deliveries (birth weight less than 2,500 grams) for the East Baton Rouge Parish from 1990 to 2001 (Table 1). The input data was the number of low-birth-weight deliveries, divided by the total number of births and multiplied by 100 for each census tract falling within the parish boundaries. The rook contiguity was used. The proportion of low-birth-weight deliveries for the entire parish is relatively stable and varies slightly between 10.04% and 11.37%. Figure 1 shows the spatial distribution of the percentages of low-birth-weight deliveries for the year 1997 by census tracts. This year was selected because it measured the highest positive SA of each of the 12 years (Table 1). In Figure 1, the darker the gray tone, the higher the percentages of low-birth-weight deliveries. The highest proportions are clearly visible around the Baton Rouge downtown area, especially to the east and northeast. Above average percentages are also concentrated in downtown and in individual census tracts in the south and in the east of the parish. The lowest percentages of low-birth-weight deliveries are found in the north and in smaller areas in the east and in the south of the study area.

As expected, a positive SA is evident for each of the 12 years. The Moran's I statistic is highest in 1997, with a value of 0.4355, and lowest in 1999, with a value of 0.0258. The latter value comes close to a distribution that can be described as complete spatial randomness.

Table 1. Results of SA statistics for percentages of low-birth-weight deliveries in East Baton Rouge Parish from 1990 to 2001

Year	Moran's I
1990	0.3026
1991	0.1437
1992	0.3567
1993	0.0915
1994	0.2538
1995	0.3042
1996	0.3271
1997	0.4355
1998	0.1970
1999	0.0258
2000	0.3139
2001	0.0611

Figure 1. Spatial distribution of the percentages of low-birth-weight deliveries in East Baton Rouge Parish in 1997



Local Spatial Autocorrelation

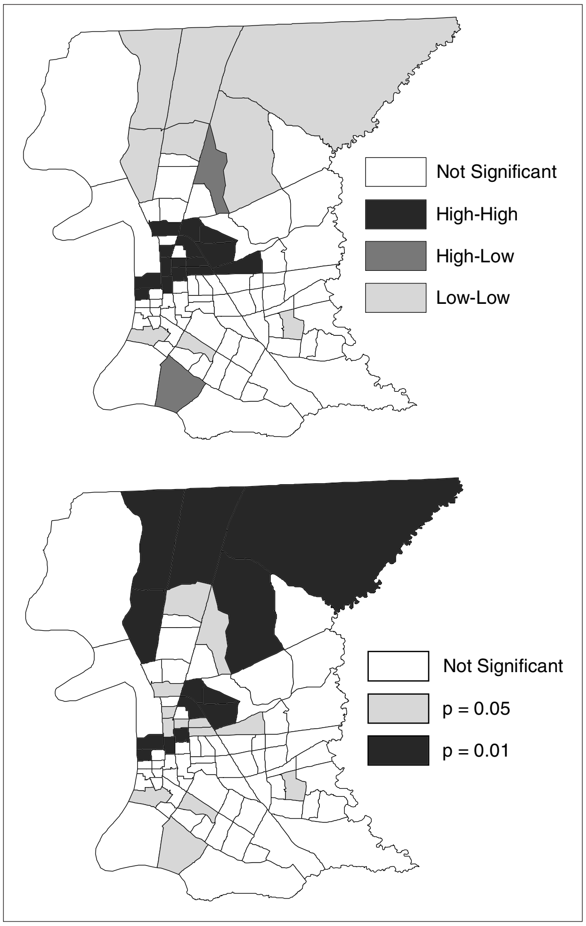
In case of positive global SA, we would expect similarly high values close to each other somewhere in the study area, but global indicators of SA do not reveal where those hot spots are located. Similarly, these indicators do not reveal where similarly low values, or cold spots, are clustered. In case of negative global SA, we would expect low attribute values next to high attribute values and vice versa. An arrangement whereby a low value is surrounded by high values and a high value is surrounded by low values is referred to as a spatial outlier. Such outliers are very interesting to explore further. They could be a data input error, or a region that is somehow unique amongst its neighbors. But what makes this region so unique? Again, global indicators fail to identify the exact location of spatial outliers.

Local SA statistics identify the neighborhood relationship between attribute values and are based on the concept of local indicators of spatial association (LISA) (Getis, 1991; Getis & Ord, 1996; Anselin, 1995). LISA is an indicator of the extent to which the value of an observation (such as census tract) is similar to or different from its neighboring observations. Figure 2 (top map) shows a LISA map for percentages of low-birth-weight deliveries in East Baton Rouge Parish in 1997. The results were calculated with the local Moran's statistic (Anselin, 1995), which is the local counterpart of the global Moran's I. This statistic requires attribute data to be aggregated to areas and the definition of a neighborhood in the form of a contiguity matrix. In this example, rook contiguity was used. For a detailed discussion of the local Moran's statistic, the reader is referred to Anselin (1995).

In Figure 2 (top map), areas in dark gray indicate census tracts with high percentages of low-birth-weight deliveries surrounded by census tracts with similar high percentages. These areas can be described as hot spots. In contrast, areas in light gray indicate census tracts with low percentages of low-birth-weight deliveries surrounded by similar low percentage census tracts. An area like that would be indicative of a cold spot. Two census tracts filled with medium gray in Figure 2 (top map) can be described as spatial outliers. Both spatial outliers are census tracts with a high percentage of low-birth-weight deliveries surrounded by census tracts with low percentages of low-birth-weight deliveries.

The bottom map in Figure 2 is referred to as a significance map and measures at what level of significance the value of an observation (such as census tract) is similar to or different from its neighboring observations. The darker the gray in the map is, the higher the level of significance. For example, the census tracts filled with dark gray forming the cluster in the downtown and surrounding areas of Baton Rouge in the top map have significantly and similar high values as their

Figure 2. Local indicator of spatial association (LISA) map for percentages of low-birth-weight deliveries in East Baton Rouge Parish in 1997 (top) and corresponding significance map (bottom)



neighbors. In other words, there is a high confidence in categorizing this region as a hot spot. Similarly, the cluster of census tracts in the north and northeast of the parish, filled with light gray (in the top map), can be confidently described as a cold spot.

Cluster Analysis

A cluster, or hot spot, is defined as a concentration of incidents within a limited geographical area that appear over time. Literally dozens of methods to identify clusters exist and results may differ depending on the method used. Therefore, it is good practice to subject your data to more than one cluster method and look for commonalities in the results. Areas where clusters from different methods overlap are locations where one can be reasonably confident that real clusters exist. Areas where only one method identifies a cluster, but others fail to do so, should obviously be treated with suspicion.

Cluster methods can find absolute and relative clusters in your data. Absolute clusters are based on a single incident distribution. For example, in a map showing all residences where a low birth weight baby was born, clusters of residences with low-birth-weight babies can be identified. One would expect the number of low-birth-weight babies to be higher in areas with many total births and lower in areas with few total births. For example, one area has 200 births and 20 of those are low-birth-weight. In a second area with only 100 births, 15 possess a low birth weight. If only the number of low-birth-weight deliveries is mapped, a cluster method may identify the area with the 20 as an absolute hot spot, but may fail to do so for the area with the 15 low-birth-weight deliveries. On the other hand, the same cluster method may identify a relative cluster for the area where the chance of having a low-birth-weight birth is 15%, but may not find a cluster for the area where this chance is only 10%. This simple example shows that absolute clusters can be misleading, and relative or standardized clusters should be calculated whenever possible. This requires, however, collecting both numerator (such as low-birth-weight births) and denominator data (such as all births), since relative clusters are calculated from two different incident distributions. The denominator data are also referred to as the base or at-risk "population." It should also be noted that all cluster methods explore the data set for high concentrations of incidents, but do not explain the underlying processes that are responsible for the clustering. Additional methods (such as spatial association techniques) need to be employed to uncover such causes.

According to Everett (1974) and Can and Megbolugbe (1996), the following groups of cluster analysis methods can be distinguished:

1. Point locations: The spatial mode and the spatial fuzzy mode are two simple measures of clusters, defined as a high number of incidents being concentrated at a single point location (spatial mode) or within a user-specified distance of a single point location (spatial fuzzy mode). In other words, the spatial mode represents the highest number of incidents for any location in the study area. Similarly, the spatial fuzzy mode represents the highest number of incidents within a user-specified distance of any location in the study area. For example, multiple incidents of low-birth-weight deliveries occur in the same apartment complex with the same street address. The number of births at this location is higher than at any other location

(address) in the study area. In statistical terms, this location would be referred to as a spatial mode. In addition to counting the number of births at the apartment complex, the spatial fuzzy mode would also consider all births within a user-defined radius (e.g., 300 m) of that apartment location.

2. Hierarchical techniques: Original incident locations are first grouped into clusters on the basis of some criteria (e.g., nearest neighbor distance). These clusters are referred to as first-order clusters, and they are subsequently grouped into second-order clusters based on the same criteria (e.g., nearest neighbor distances between first-order clusters). This process of grouping clusters into higher-order clusters continues until either all incidents fall into a single cluster or else the grouping criterion fails. The result is a hierarchy of clusters that can be displayed with a dendrogram, which looks like an inverted tree diagram. The nearest neighbor hierarchical clustering method is one important representative from this group and it will be discussed in more detail later.
3. Partitioning techniques: These methods partition incidents into a specified number of groupings, usually defined by the investigator. Thus, all incidents are assigned to one, and only one, group. This method is less useful if the goal is to find true clusters of high concentrations of incidents. One representative is the K-means partitioning technique.
4. Density techniques: These methods interpolate discrete incident locations to a density surface and then identify clusters as the highest densities. Since only one incident distribution is interpolated, only absolute clusters are found. One representative method from this group is the single kernel density method, which will be discussed in detail later.
5. Clumping techniques: These methods involve the partitioning of incidents into clusters, but allow overlapping membership.
6. Risk-based techniques: These methods identify relative or standardized clusters using two sets of incident locations, one of which (such as denominator data) is the "at-risk population," Two representative methods are the risk-adjusted nearest neighbor hierarchical clustering routine and the dual kernel density method. The latter method will be discussed in more detail in subsequent paragraphs.
7. Miscellaneous techniques: These methods are applied when data have been aggregated to areas or zones. Among those methods are local measures of spatial autocorrelation (Anselin, 1995).

Cluster Techniques

Spatial Filtering (DMAP)

Chapter III explained how the predominant use of GIS in terms of identifying spatial patterns is as a visualization tool. In other words, it is used to make maps, and the most common map type is a choropleth map. A typical map showing high rates of an infant risk (such as infant mortality) would vary from light shades (a low IMR) to dark shades (a high IMR). Unfortunately, and as mentioned in Chapter III, there are pitfalls with using this type of mapping approach (Monmonier, 1996; Rushton & Armstrong, 1997). Results can vary according to the classification scheme and number of classes chosen. Results can also vary according to the unit of aggregation, whether zip code, census tract, census block group, or census block. Irrespective of the variation that occurs due to the spatial aggregation chosen, the underlying premise of a political boundary providing a data cutoff is fundamentally wrong. Two sides of the same street may fall in different census tracts; births occurring to residences on either side of the street are more likely to have similarities than to other births 500 meters away but still housed in the same census tract.

A cluster approach that allows for the creation of smooth rate surfaces and is not reliant on any political boundary (as long as point data is available), and one that offers a statistical test of significance is the spatial filter. This analytical method was originally developed to identify infant mortality hot spots, and has been used frequently in association with the Healthy Start program in Baton Rouge. Examples of the output contours can be seen in the chapter detailing GIS in the Baton Rouge Healthy Start (specifically Figures 6, 7, and 8 in Chapter IX). Spatial filter analysis works as follows: Using a mother's residence at time of birth (and mothers residence at time of death in the case of infant mortality) a point surface is split into numerators (the variable under investigation, such as low birth weight), and denominators (total births). A grid is overlaid onto the map, with a circle or "filter" extending from each grid point. The distance between grid points and size of the filter are variable, though the basic premise is to create an overlapping surface of filters that capture trends across the city. The larger the filter, the smoother the surface becomes, with less neighborhood detail being revealed. If, however, the filter is too small, the resulting surface displays no neighborhood pattern. At the center of each filter (the grid point), a rate is calculated whereby numerator events (for example, low-birth-weight births) are divided by denominator events (all births) and multiplied by 1,000. The resulting surface of rates assigned to each grid location can be contoured using one of the interpolation routines found in most GIS packages. In addition, a Monte Carlo simulation can also be applied to create a simulation distribution against which the actual events are compared. Each denominator event is stochastically allowed to become a numerator event. In effect, a new numerator surface is created, which should approximate the actual number of numerator events, though in varied spatial locations. This simulation is repeated "n" times, usually 1,000 runs. The actual rate calculated for a grid location is compared to the range of rates calculated in the simulation distribution. If the actual rate at that grid point exceeds 950 of the 1,000 simulations, it can be said that there is a 95% confidence of that rate being higher than one would expect by chance alone. By again contouring these rates based on the percentage of confidence, neighborhoods of risk can be identified.

The rate calculations and Monte Carlo simulation can all be performed in DMAP, with the grid being exported into a GIS for interpolation (Rushton & Armstrong, 1997). This software is available online through the University of Iowa Department of Geography (for an example of this technique see Rushton & Lolonois, 1996).

Nearest Neighbor Hierarchical Clustering (NNHC)

This hierarchical technique is also one of the oldest clustering methods. Different clustering criteria have been used to group incident locations, including the nearest neighbor method (Johnson, 1967; D'Andrade, 1978), farthest neighbor, the centroid method (King, 1967), median clusters (Gowers, 1967), group averages (Sokal & Michener, 1958), and minimum error (Ward, 1963). The example that follows (see Figure 3) uses the nearest neighbor method that identifies groups of incidents that are spatially close. This method defines a threshold distance and compares this threshold to the distances for all pairs of points. Only points that are closer to one or more other points than the threshold distance are selected for clustering. The threshold distance is a probability level for selecting any two points (a pair) on the basis of a chance distribution (complete spatial random distribution). In order to get statistically significant clusters, low probability levels should be selected, such as 0.01, 0.05 or 0.1. In addition, the user can specify a minimum number of points to be included in a cluster. This guarantees that many small clusters with only a few incident locations are not created. Only points that fit both criteria closer than the threshold and belonging to a group with the minimum number of points — are clustered at the first level (first-order clusters). This process is repeated, applying the same two criteria to all first-order clusters (instead of the original incident locations) to create second-order clusters. Second-order clusters, in turn, are aggregated to third-order clusters and this reclustering process is continued until no more clustering is possible, either because all clusters converge into a single cluster or, more likely, the clustering criteria fail. A relationship exists between both criteria (threshold distance and minimum number of points) and the number of clusters being calculated. The higher the number of minimum points and the shorter the threshold distances, the fewer clusters will be calculated. Accordingly, a higher number of clusters can be achieved with longer threshold distances and a lower number of minimum points. From a practical point of view, the investigator should run this cluster method with different combinations of both criteria and then select the combination that best fits all results. It is important that the final combination of threshold distance and minimum number of points should always be reported.

One nice feature about this cluster technique is that hot spots can be detected at different spatial scales. For example, if the study area is a city, then first-order clusters would indicate local hot spots extending over several blocks, second-order clusters would show hot spots at the neighborhood-level, and third-order clusters would spread over large areas of the entire city. Finally, clusters are displayed in the form of so-called standard deviational ellipses. Ellipses can vary in size depending on the number of standard deviations chosen to display them. For example, the major (minor) axis for a two-standard deviational ellipse is twice as long as for a one-standard deviational ellipse.

Figure 3a shows the results of a NNHC method applied to the residences of all mothers who had a low-birth-weight birth in the city of Baton Rouge in 2001. From a total number of 5,228 births, 5,181 were successfully geocoded (99.10% success rate) to the street network of the city of Baton Rouge; 569 (10.98%) of all births possessed a low-birth-weight and the clusters in Figure 3a are calculated from this number. The results show thirteen first-order clusters (small ellipses) and one second-order cluster (large ellipse) using a minimum number of six points per cluster and a 0.05 level of significance. All clusters are displayed as two-standard deviational ellipses. Since only one data set is used in this example, all clusters can be described as absolute clusters. In Figure 3a, the background information includes the census tracts of the city of Baton Rouge.

Kernel Density Estimation

The kernel density estimation is an interpolation technique that generalizes individual point locations or events to an entire area, and provides density estimates at any location within the study region (Silverman, 1986; Härdle, 1991; Bailey & Gatrell, 1995; Burt & Barber, 1996; Bowman & Azalini, 1997). Density estimates are derived by placing a symmetrical surface, called the kernel function, over each event and summing the value of all surfaces onto a regular reference grid superimposed over the study region. Typically, a symmetrical kernel function falls off with distance from each event at a rate that is dependent on the shape of the kernel function and the chosen bandwidth. A number of different kernel functions have been used, including normal, triangular, quartic, negative exponential, and uniform. The bandwidth determines the amount of smoothing, and for the limited distance functions (triangular, quartic, negative exponential, and uniform), the size of the kernel's search area. In case of the normal kernel function, the bandwidth is the length of the standard deviation of the normal distribution. The normal kernel function produces a density estimate over the entire region (that is why it is referred to as an unlimited distance function), whereas the other four functions produce estimates only for the circumscribed bandwidth radius. Kernel density estimates can also be applied to two different sets of point patterns or events at the same time, and this is referred to as dual kernel density estimates. Additionally, density calculations can be carried out for events that are weighted or events that are not weighted.

The choice of the cell size of the regular reference grid influences the resolution of the final density surface map. Larger cell sizes result in more generalized density surfaces having a coarser resolution. Accordingly, smaller cell sizes visualize the density surface at a finer resolution. In general, a higher number of events per unit area warrant smaller cell sizes, and vice versa. Chorley and Haggett (1968) suggest calculating the cell size, Q, as being dependent on the study region size and the number of events:

$$Q = 2x^2 / n \qquad (1)$$

where x is the maximum areal extent of the point pattern map, and n is the size of the spatial point pattern. A possible extension to this formula, especially for point patterns with rapidly changing point densities, could be the inclusion of different cell sizes dependent on the point densities across the study region.

Although a number of different kernel functions do exist, the normal is the most commonly used (Kelsall and Diggle, 1995). Kernel functions differ with respect to the shape of their kernels and the size of their search areas. For the only unlimited distance function — the normal kernel function — the entire study region is searched when calculating density estimations for each individual cell. In contrast, all other kernel functions use a limited, circular search area around each event location, when calculating cell density estimations.

Selecting an appropriate bandwidth is a critical step in kernel estimation, and its choice affects the results much more than the choice of cell size or the type of kernel function. In general, the bandwidth determines the amount of smoothing of the point pattern. An increasing bandwidth expands the kernel at the cell center and results in a smoothed and generalized map with low-density values. In contrast, a small bandwidth will result in less smoothing, producing a map that depicts local variations in point densities. A very small bandwidth almost reproduces the original point pattern and is spiky in appearance.

There is no agreement in the literature on how wide a particular bandwidth should be. The Spatial Analyst Extension of ArcView 3.2 (ESRI, 1999) uses a bandwidth that is calculated as

$$b = Minimum(x, y) / 30 \tag{2}$$

where x and y are the extent of the point pattern in x- or y-direction.

In order to calculate b, a rectangle is drawn around the point pattern and the shorter of the two sides of the rectangle is selected and divided by 30. This is a crude measurement of the bandwidth and does not take into account the size of the point pattern (i.e., the number of events) and the distribution of the points within the study area.

Diggle (1981) accounts for the size of the point pattern and suggests calculating

$$b = 0.68(n)^{-0.2}\sqrt{\Re} \tag{3}$$

where n is the size of the point pattern and $\Re$ is the size of the study area. Again, this approach fails to include the spatial distribution of the points. One way to achieve this is to base the bandwidth on average distances among points (Williamson et al., 1998) or to use different bandwidths in different parts of the study area, an approach known as adaptive kernel estimation (Bailey & Gatrell, 1995).

In the k-nearest neighbor approach, Williamson et al. suggest calculating the bandwidth as

$$b = \left(\sum_{i=1}^{n} \sum_{j=1}^{k} d_{ij} \right) 1/\ kn \tag{4}$$

where d_{ij} is the distance between point i and its j-th neighbor, k is the nearest neighbor order, and n is the size of the point pattern.

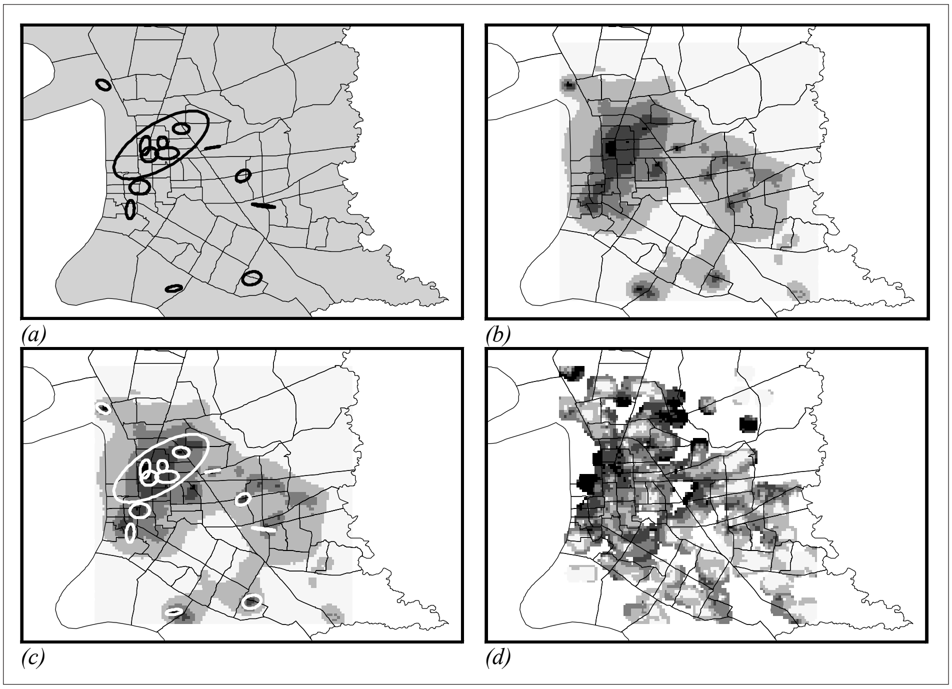
This approach is based on the interpoint distances of the point pattern. As such, the bandwidth reflects the spacing between the points, rather than the size of the study area or the number of points. Since the larger the value for k the smoother the density surface, the user has to set an appropriate value for k, depending on how much smoothing is desired.

The adaptive kernel estimation is based on sampling theory, giving the choice of bandwidth a consistent level of precision over the entire study region. This is achieved by increasing the bandwidth until a fixed number of points (i.e., minimum sample size) is counted. Accordingly, in areas of high density, small bandwidths are used to show detailed local variation, whereas in areas of low density, larger bandwidths smooth the point pattern (Bailey & Gatrell, 1995). Adaptive kernel estimation solves the problem of determining a value for *b*, but it still leaves open the question of how to set an appropriate minimum sample size. In general, the higher the minimum sample size, the larger the bandwidth and the more the density surface will be smoothed. Suggestions of how to set the appropriate minimum sample size are lacking in the literature.

The level of resolution that someone is interested in is yet another criteria that can be used to determine the bandwidth length. If, for example, one would like to draw conclusions at the neighborhood level, a reasonable choice for the bandwidth would be the radius of a circle circumscribing the average size of all neighborhoods in the study area. This is appropriate for all limited distance functions (i.e., triangular, quartic, negative exponential, or uniform). For the normal function, conservatively, the bandwidth is approximately one half that distance, since it stretches throughout the study area and tends to oversmooth distributions. Further suggestions of how to determine a fixed bandwidth can be found in Silverman (1986), Härdle (1991), Farewell (1999), Bowman and Azzalini (1997), and Scott (1992).

Figure 3b shows the results of a single kernel density estimation applied to the same study area and data as in Figure 3a. The kernel density was estimated with an adaptive bandwidth, a sample size of six, and a normal kernel function. Density values were calculated for each grid cell of a 100 by 100 regular grid. Darker grays indicate higher density values. The darkest areas can be described as absolute clusters. In Figure 3c, the results of the NNHC method from 3a are overlaid on top of the single kernel density estimation from 3b in order to find out how much both results spatially coincide. Finally, Figure 3d shows the results of a dual kernel density estimation, with the 569 low-birth-weight births in the numerator and all 5,181 births (the “at-risk population”) in the denominator. Density values were not calculated for the entire grid, but only for 0.25-mile buffers around each of the low-birth-weight birth locations. Again, darker grays indicate higher density values. The darkest areas can be described as relative clusters. A visual comparison indicates that the dual kernel density map looks very different from the single kernel density map, with relative clusters not necessarily found in the same locations as absolute clusters.

Figure 3. (a) Results of nearest neighbor hierarchical cluster (NNHC) analysis; (b) Results of single kernel density method; (c) Results of both NNHC and single kernel density method overlaid in the same map; and (d) Results of dual kernel density method

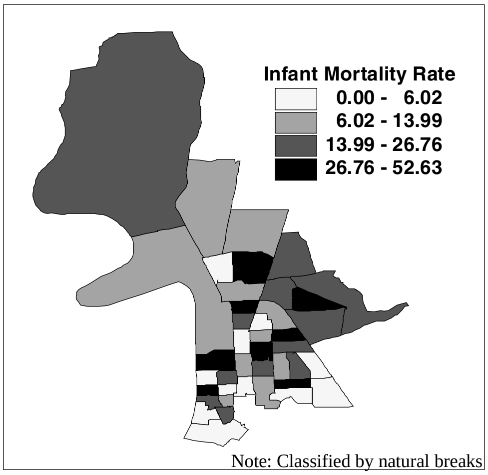


Infant Mortality and Prenatal Risks: The Case of East Baton Rouge

Previous sections of this chapter introduced techniques that can be used to measure clusters of high values (hot spots), clusters of low values (cold spots), and spatial outliers. These methods were discussed for data collected at both the area and the point level. Once a hot or a cold spot has been identified, explanations for the clustering of high or low values need to be found. This can be accomplished through methods of spatial association. For example, after a hot spot of infant mortality has been identified, spatial regression analyses might reveal the underlying factors (infant mortality risk factors) that are responsible for the existence of this hot spot. Once these factors have been measured and recognized, policies can be put into place to lower or even remove these factors completely.

The following analysis is based upon linked birth and death certificate data for the five "at risk" zip code area within the East Baton Rouge Parish and 41 "at

Figure 4. Spatial distribution of the infant mortality rate in an "at-risk" area of East Baton Rouge Parish



risk" census tracts falling within these five zip codes. These areas comprise the Healthy Start program area. The time frame is from 1996 through 1998. Data was first address-matched to the street file of the parish and subsequently aggregated to each of the 41 census tracts. For each census tract, the number of births and infant deaths were queried within the GIS, and the infant mortality rate was calculated as the quotient of the number of infant deaths (numerator) and the number of births (denominator) multiplied by 1,000. Similarly, 21 different prenatal risk factors included in the birth certificate data were also derived through GIS queries at the tract level and expressed as proportions. The next section discusses the spatial variation of infant mortality, followed by a discussion of prenatal risks derived from the birth certificate data. The last part identifies which of the 21 prenatal risk factors can be used as significant predictors of the infant mortality rate (Leitner & Curtis, 2002).

The spatial distribution of infant mortality for 41 census tracts in an "at-risk" area of East Baton Rouge is shown in Figure 4. The overall picture displays a fragmented landscape of infant mortality rates. Census tracts falling in the highest class are scattered throughout the study area, with a majority located in the southern portion. There seems to be a region of higher infant mortality in the east and a higher number of low infant mortality tracts in the south. A large disparity in the rate can also be observed, with seven of the tracts possessing a value of 0.00 and one tract showing a maximum value of 52.63 (Leitner & Curtis, 2002).

The following lists all 21 prenatal risk variables selected for subsequent analysis:

1: Proportion of mothers being less than or equal to 18 years of age.

2: Proportion of mothers being less than or equal to 16 years of age.

3: Proportion of mothers who were equal to or older than 35 years of age.

4: Proportion of babies weighing less than or equal to 1,500 grams (classified as a very low-birth-weight baby)

5: Proportion of babies weighing less than or equal to 2,500 grams (classified as being low-birth-weight).

6: Proportion of mothers making seven or fewer prenatal visits.

7: Proportion of mothers making no prenatal visits.

8: Proportion of mothers making a first prenatal visit in the fourth month or later.

9: Proportion of mothers giving birth in the 32nd week of gestation or before (classified as a very premature birth).

10: Proportion of mothers giving birth in the 36th week of gestation or before (classified as being a premature birth).

11: Proportion of mothers having previously given birth within the preceding 2 years.

12: Proportion of mothers who had previously had a termination.

13: Proportion of mothers who self-reported smoking during pregnancy.

14: Proportion of mothers who self-reported alcohol use during pregnancy.

15: Proportion of women not naming a father on the certificate (seen as a measure of a future lack of male parental involvement).

16: Proportion of African American mothers.

17: Proportion of unmarried mothers

18: Proportion of mothers receiving less than or equal to an 11th grade education.

19: Proportion of mothers who had a previous child die.

20: Proportion of mothers who gave birth for the first time.

21: Proportion of mothers who gained less than or equal to 15 pounds during pregnancy.

Although the study area is relatively small, the results of selected descriptive statistics for the 21 risk factors reveal large differences between some individual variables across the 41 census tracts, indicating a rather diverse region (Table 2). The largest difference (a standard deviation of 27.63) can be found for the proportion of African American mothers (ranged from a low of 4.3% to a high

Table 2. Descriptive statistics of infant mortality risk factors

Risk factor	Min	Max	Mean	Std. dev.	Outlier (Census Tract)
Low birth weight	4.88	22.22	13.38	4.41	None
Very low birth weight	0.00	8.03	3.18	2.06	None
African American mother	4.30	99.61	76.90	27.63	19[1]
Gestation <= 36 weeks	3.25	25.40	14.53	4.83	28[1]
Gestation <= 32 weeks	0.00	10.32	4.80	2.33	None
Age of mother <= 18 years	1.63	26.67	15.26	6.19	28, 19[1]
Age of mother <= 16 years	0.00	11.86	5.48	2.80	13
Age of mother >= 35 years	1.83	31.18	9.07	5.86	19[2], 18,20
No prenatal visit	0.00	6.25	2.74	1.60	None
First prenatal visit >= 4th month	0.00	40.00	26.01	9.89	19[1]
Number of prenatal visits <= 7	2.15	43.75	20.01	8.50	14
Alcohol during pregnancy	0.00	4.21	1.44	1.18	None
Tobacco during pregnancy	0.00	15.63	7.78	3.27	14, 28[1]
Education of mother <= 11th grade	2.44	61.96	29.43	12.56	21
Weight gain during pregnancy <= 15 pounds	5.49	31.60	19.49	5.78	14, 6.02, 20[1], 32.01[1]
First birth	23.31	65.85	40.76	8.33	12, 28
At least one previous termination	15.25	36.26	22.86	4.24	20, 21, 31.02
Unmarried mother	6.45	83.97	63.46	20.56	19[1], 20[1], 28[1]
Previous child died	0.00	6.32	2.41	1.36	1, 31.02
Father not named on birth certificate	1.08	61.02	38.67	15.90	None
Previous birth within the last two years	12.20	44.07	25.82	6.41	21, 13

Note: Outliers without a superscript identify census tracts with prenatal risks that are mild outliers falling 1.5 interquartile ranges above the upper quartile. A superscript[1] is a mild outlier falling 1.5 interquartile ranges below the lower quartile; a superscript[2] an extreme outlier falling 3.0 interquartile ranges above the upper quartile.

of 99.61%), followed by the proportion of unmarried mothers (standard deviation of 20.56 with a range of 77.52%), the proportion of women not naming a father on the certificate (standard deviation of 15.9 and a range of 59.94%) and the proportion of mothers receiving less than or equal to an 11th grade education (standard deviation of 12.56 and a range of 59.52%). Risk factors remaining relatively stable across the study region at a relatively low percentage level include the proportion of mothers who self-reported alcohol use during pregnancy (standard deviation of 1.18), who had a previous child die (standard deviation of 1.36), making no prenatal visits (1.6), and very low-birth-weight babies (2.06). Two census tracts, 14 and 21, possess three mild outliers, each falling 1.5 interquartile ranges above the upper quartile, and only one census tract (19) possesses an extreme outlier (3.0 interquartile ranges above the upper quartile) for the proportion of mothers who were equal to or older than 35 years of age. The interquartile range is a measure of variability and is the difference between the upper quartile (separates the top 25% of the data set from the bottom 75%) and the lower quartile (separates the bottom 25% from the upper 75%). These outliers are nonspatial outliers. If an outlier is also a spatial outlier, it could be an indication of an environmental causation (Leitner & Curtis, 2002).

Regressing Selected Prenatal Risk Factors on the Infant Mortality Rate

One of the most common methods in the social sciences of associating a variable under investigation (such as infant mortality) with several other contributing factors is a multiple linear regression model. Simply put, this model determines how much variation in the variable under investigation (the dependent variable) can be explained by other risks, such as birth weight, number of prenatal visits, and so forth. As 21 prenatal risk factors were initially identified, a variant known as stepwise multiple regression was run. The stepwise selection begins with a forward selection (finds the variable that explains the majority of dependence variance), but after a variable has been added to the model, the resulting equation is examined to see if any coefficient has a sufficiently large p-value that suggests that a backward elimination procedure should be implemented. If a coefficient has such a high probability value, the corresponding variable is removed from the model. This procedure continues until no additions or deletions are indicated, according to specific significance levels. The significance levels chosen for this example were an F probability of 0.05 for adding and 0.1 for removing a variable from the model. The reader not familiar with multiple regression analysis should consult any introductory statistics book.

The results of the final regression model are shown in Table 3. Of the original 21 prenatal risk factors, the following four factors are significant predictors of the infant mortality rate:

1: Proportion of babies weighing less than or equal to 1,500 grams (classified as a very low-birth-weight baby).

2: Proportion of mothers who gave birth for the first time.

Table 3. Results of the multiple regression analysis (stepwise selection)

	Coefficient	R^2 adjusted	t-value	p-value
Constant	41.337		3.415	0.002
Very low birth weight	3.175	0.277	4.169	0.000
First born	-0.854	0.408	-4.114	0.000
Mother's education	-0.546	0.477	-3.442	0.001
Low weight gain	0.813	0.552	2.677	0.011

3: Proportion of mothers receiving less than or equal to an 11th grade education.

4: Proportion of mothers who gained less than or equal to 15 pounds during pregnancy.

These four factors account for 55.2% of the variation in infant mortality. The adjusted R^2 value for each risk factor indicates that the proportion of very low-birth-weight babies explains about the same amount of variation (27.7%) as the other three factors together. The analysis of variance (ANOVA) resulted in F = 13.322 with a probability of 0.000 and 4 and 36 degrees of freedom.

As expected, the coefficients for the two risk factors "very low-birth-weight babies" and "low weight gain during pregnancy" positively influence the infant mortality rate. However, it is somewhat unexpected that mothers with low education levels and first-born babies negatively influence the infant mortality rate. For the latter two significant variables, further analysis is needed before explanations can be offered as to their relationships to infant mortality (Leitner & Curtis, 2002).

Linear regression analysis is based on a set of assumptions. The most important are whether the data is normally distributed, and whether the relationship between dependent and independent variables is linear in nature. Violating any of these assumptions means that the regression model is incorrectly specified and that in turn will result in biased results. In this research, all 21 perinatal risk factors and the infant mortality rate were tested for normality using a one-sample Kolmogorov-Smirnov test. This test revealed that the "proportion of African American mothers" was the only variable for which the normality assumption was rejected at the probability = 0.05.

Spatial regression analysis — that is, regression analysis applied to aggregated spatial data — is based on additional assumptions; for instance, that there exists no spatial autocorrelation in the data. Again, violating these assumptions will lead to biased results. Over the last 10 to 20 years, a suite of diagnostics has been developed to test for such violations and to respecify the original spatial regression model, if one or more of these assumptions have not been met (Anselin, 1988, 1992, 2001). One possibility to account for the effects of aggregated spatial data is the use of local regression models, also referred to as a geographically weighted regression. This new type of spatial regression analysis will be discussed in the next section.

Geographically Weighted Regression

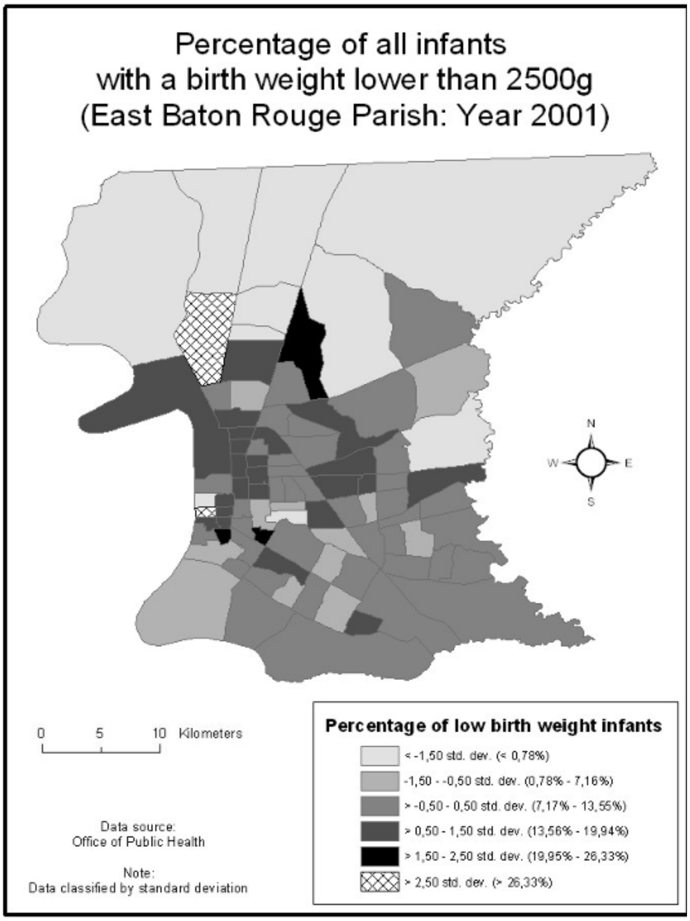
Geographically weighted regression (GWR) is a technique for exploratory spatial data analysis. It is a method for the local analysis of relationships in multivariate data sets (Fotheringham et al., 2002). As such it allows local, as opposed to global, spatial regression models to be estimated. Data near a point i in space have more influence than data located farther away. The result is that the estimated regression coefficients are local and vary with the distance to i (Fotheringham et al., 1999). A GWR requires data to have location information in the form of coordinates. These coordinates are used to determine the distance between the points in space. GWR is based on the ordinary least squares (OLS) regression. The approach of OLS, even when used in the form of GWR, suggests that data are continuous and normally distributed (Gaussian distribution).

In the following analysis, multivariate logistic GWR models are estimated for East Baton Rouge Parish from 1990-2001 on a yearly basis (Arthold, 2004; Arthold, et al., 2004). In this case, the GWR model was specified as a logistic regression or binary logit regression because some of the variables were not continuous. In the logistic regression, the dependent variable is coded in a binary or dichotomous format, whereas the independent variables are coded as binary or as continuous (Brunsdon et al., 2002). The logistic regression is often used to find out what influenced an event to occur or not to occur. For example, what influenced an infant death to occur or why was an infant born with (very) low-birth-weight?

In this example, low-birth-weight deliveries were chosen as the dependent variable (Figure 5). This variable was coded as zero if the infant was a low-birth-weight delivery, and all other births were coded with a one. Low birth weight was chosen as the dependent variable instead of infant mortality for two reasons. First, low birth weight has been repeatedly recognized as one of the main risk factors for infant mortality. Second, the number of infant death events per year in East Baton Rouge Parish is too small for properly estimating the GWR models. For these two reasons low-birth-weight deliveries are used as a surrogate variable for infant mortality. The independent variables included the mother's age, race, educational level, and the number of prenatal care visits. Counts instead of binary values were used for three of the four independent variables. Only the mother's race was coded as a binary variable with a value of one for Caucasians and a value of two for African Americans (Arthold, 2004; Arthold, et al., 2004).

The results of the multivariate logistic GWR models are summarized in Table 5 and the spatial variations in the four estimated local regression coefficients (one coefficient for each independent variable) for 2001 are displayed in Figure 6. The GWR model analyzes the local relationships between the independent and the

Figure 5. Spatial distribution of the percentages of low-birth-weight deliveries in East Baton Rouge Parish in 2001



dependent variables, which means that in contrast to the global OLS regression, the estimated regression coefficients for the GWR are local and vary across the study area. For this reason, each regression coefficient in Table 5 shows a range in values. These coefficients can be interpreted the same way as coefficients in an OLS regression. A positive (local) regression coefficient indicates a positive relationship between the independent and the dependent variable, meaning that an increase (or a decrease) in the independent variable results in an increase (or a decrease) in the dependent variable. Similarly, a negative (local) regression coefficient constitutes a negative relationship that is an increase (decrease) in the independent variable results in a decrease (increase) in the dependent variable.

For example, in 2001 a negative local regression coefficient for the "age" variable means that the older the mother, the lower the risk of a low-birth-weight

Table 5. Variation in the regression coefficients for the independent variables in the multivariate logistic GWR *model*

Year	Independent Variable	Range of the Regression Coefficient
1990	Age	-0.04 to 0.08
	Education	-0.14 to 0.17
	Race	-1.45 to 0.18
	Prenatal Care Visits	-0.01 to 0.19
1991	Age	-0.04 to 0.01
	Education	0.04 to 0.13
	Race	-0.44 to 0.25
	Prenatal Care Visits	0.06 to 0.16
1992	Age	-0.04 to 0.02
	Education	-0.06 to 0.18
	Race	-0.75 to -0.05
	Prenatal Care Visits	0.00 to 0.12
1993	Age	-0.08 to 0.02
	Education	-0.09 to 0.16
	Race	-0.13 to 0.23
	Prenatal Care Visits	-0.02 to 0.25
1994	Age	-0.02 to 0.02
	Education	0.02 to 0.13
	Race	-0.09 to -0.01
	Prenatal Care Visits	0.04 to 0.12
1995	Age	-0.05 to 0.05
	Education	-0.07 to 0.16
	Race	-0.35 to 0.06
	Prenatal Care Visits	0.03 to 0.20
1996	Age	-0.03 to 0.02
	Education	-0.03 to 0.09
	Race	-0.14 to 0.00
	Prenatal Care Visits	0.03 to 0.13
1997	Age	-0.05 to 0.01
	Education	-0.00 to 0.14
	Race	-0.20 to -0.01
	Prenatal Care Visits	0.01 to 0.17
1998	Age	-0.05 to 0.03
	Education	-0.04 to 0.15
	Race	-0.17 to 0.14
	Prenatal Care Visits	-0.07 to 0.18
1999	Age	-0.07 to 0.03
	Education	-0.03 to 0.14
	Race	-0.34 to 0.68
	Prenatal Care Visits	-0.13 to 0.12
2000	Age	-0.05 to 0.03
	Education	-0.13 to 0.21
	Race	-0.39 to 0.14
	Prenatal Care Visits	-0.09 to 0.20
2001	Age	-0.07 to 0.03
	Education	-0.08 to 0.19
	Race	-0.40 to 0.08
	Prenatal Care Visits	-0.08 to 0.19

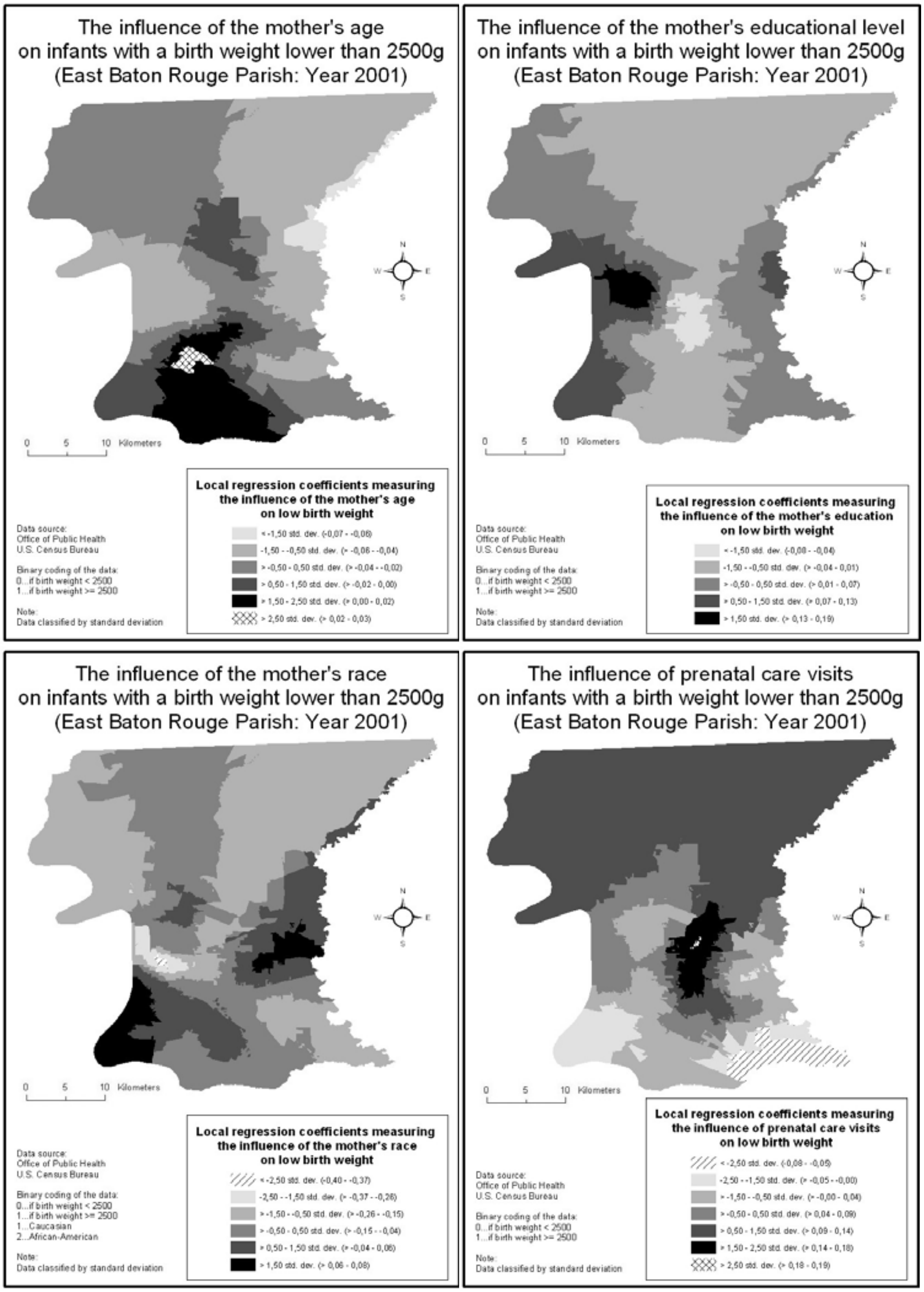
birth, or the younger the mother, the higher the risk of a low-birth-weight birth. Similarly, a positive local regression coefficient for the same variable indicates that with an increased age, the risk of a low-birth-weight birth also increases, or the younger the mother, the lower the risk of a low-birth-weight birth. All other local regression coefficients can be interpreted the same way.

Several observations can be made from these results. First, most local regression coefficients vary from a relatively low negative to a relatively low positive value. This means that the influence that each of the four independent variables has on whether an infant is born with a low birth weight or not is relatively small. The only exception is the race of the mother. In almost every year, this variable exhibits the largest influence on the infant birth weight, with mostly a negative local regression coefficient. However, local regression coefficients in Table 5 need to be interpreted with caution, because the current version of the GWR software (see Chapter III) does not calculate significance values for the logistic regression (Arthold, 2004; Arthold, et al., 2004).

The advantage of the GWR is that estimated local regression coefficients can be mapped and the spatial relationships between the dependent and independent variables explored and interpreted. This provides insight into the local spatial variation in low-birth-weight births and the four risk factors analyzed in this model. The estimated local regression coefficients are displayed in Figure 6. The interpretation of these results is somewhat difficult, because it requires the interpretation of as many maps as coefficients have been estimated. Additionally, the boundaries and spatial extents of the individual classes for each coefficient vary from map to map, and are usually different from any enumeration unit (census tracts, zip codes, etc.) for which other data might have been collected.

Figure 5 displays a census tract north of the downtown area, where the percentage of low-birth-weight deliveries is very high (greater than 26.33%). By comparing this census tract with the same area in the maps depicting the local regression coefficients, the influence each independent variable has on the infant birth weight can be quantified. This comparison reveals that the high percentage of low-birth-weight deliveries results from mothers who are older (the coefficient is slightly negative and ranges between -0.04 to -0.02), have a lower educational level (the coefficient is slightly positive and ranges from 0.01 to 0.07), are predominantly African-American (the coefficient is negative and varies between -0.26 and -0.15), and have a low number of prenatal care visits (positive coefficient that ranges between 0.09 and 0.14). Among all four coefficients, the race of the mother exhibits the largest influence on infant birth weight, followed by the number of prenatal care visits. For a second example, let us look at the census tract in the east of the parish with a very low percentage (less than 0.78%) of low-birth-weight deliveries. Again, by comparing the area for this census tract with the same area in all four maps in Figure 6, the very low

Figure 6. Spatial variation in local regression coefficients in East Baton Rouge Parish in 2001 — Results of a logistic GWR model with low birth weight as the dependent variable



percentage of low-birth-weight deliveries can be explained by younger mothers (the coefficient ranges from -0.06 to -0.04) who have a higher educational level (the coefficient varies between 0.07 and 0.13) and make more prenatal visits (the coefficient ranges from 0.04 to 0.09). The influence of the mother's race is inconclusive, because its coefficient varies from -0.04 to +0.06. A negative relationship would indicate a lower percentage of African American mothers, whereas a positive relationship would indicate a higher proportion of African American mothers. Among the other three independent variables, the educational level influences the infant birth weight the most. Again, the interpretation of these coefficients should be made with caution, because significant values are not calculated by the current version of the software. Therefore it is not known if the coefficient's values are significant or just happened by chance (Arthold, 2004; Arthold et al., 2004).

References

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Boston: Kluwer.

Anselin, L. (1992). *SpaceStat, a software program for the analysis of spatial data*. Santa Barbara: National Center for Geographic Information and Analysis, University of California.

Anselin, L. (1995). Local indicators of spatial association: LISA. *Geographical Analysis, 27*(2), 93-115.

Anselin, L. (2001). Spatial econometrics. In B. Baltagi (Ed.), *Companion to econometrics* (pp. 310-330). Oxford: Basil Blackwell.

Arthold, M. (2004). *Local spatial variation in (very) low birth weight and other risk factors influencing infant mortality: A study from Baton Rouge, Louisiana*. Unpublished master's thesis, School of Geoinformation, Carinthia Tech Institute, Villach, Austria.

Arthold, M., Leitner, M., & Curtis, A. (2004). Local spatial variation of (very) low birth weight deliveries and their risk factors. In J. Strobl, T. Blaschke, & G. Griesebner (Eds.), *Applied geographic information technology XVI* (pp. 29-34). Heidelberg: Herbert Wichmann Verlag.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. Essex: Longman.

Boots, B. N., & Getis, A. (1988). *Point pattern analysis*. London: SAGE Publications.

Bowman, A. W., & Azzalini A. (1997). *Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations.* Oxford, England: Oxford Science Publications, Oxford University Press.

Brunsdon, C., Fotheringham, S. & Charlton, M. (2002). Geographically weighted local statistics applied to binary data. *Lecture Notes in Computer Science, 2478*, 38-50.

Burt, J. E., & Barber, G. M. (1996). *Elementary statistics for geographers*. New York: The Guilford Press.

Can, A. & Megbolugbe, I. (1996). *The geography of underserved mortgage markets*. Paper presented at the American Real Estate and Urban Economics Association meeting.

Chorley, R. J., & Haggett, P. (1968). Trend-surface mapping in geographical research. In B. J. L. Berry, & D. F. Marble, (Eds.), *Spatial analysis*. Englewood Cliffs, NJ: Prentice-Hall.

D'Andrade, R. (1978). U-Statistic hierarchical clustering. *Psychometrika, 4*, 58-67.

Diggle, P. J. (1981). Some graphical methods in the analysis of spatial point patterns. In V. Barnett (Ed.), *Interpreting multivariate data*. New York: John Wiley & Sons.

Environmental Systems Research Institute (1999). ArcView GIS (Version 3.2) [Computer software]. Redlands, CA: Environmental Systems Research Institute.

Everett, B. (1974). *Cluster analysis*. London: Heinemann Educational Books.

Farewell, D. (1999). *Specifying the bandwidth function for the kernel density estimator*. Retrieved April 2, 2005, from http://www.mrc-bsu.cam.ac.uk/bugs/documentation/coda03/node44.html

Fotheringham, A. S., Brunsdon, C. & Charlton, M. E. (2002). *Geographically weighted regression*. Chichester, UK: John Wiley & Sons.

Geary, R. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician, 5*, 115-145.

Getis, A. (1991). Spatial interaction and spatial auto-correlation: A cross-product approach. *Environment and Planning A*, *23*, 1269-1277.

Getis, A., & Ord, J. K. (1996). Local spatial statistics: An overview. In P. Longley & M Batty (Eds.), *Spatial analysis: Modeling in a GIS environment*. Cambridge, England: GeoInformation International.

Gowers, J. C. (1967). A comparison of some methods of cluster analysis. *Biometrics, 23*, 623-628.

Härdle, W. (1991). *Smoothing techniques with implementation in S*. New York: Springer-Verlag.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 2*, 241-254.

Kelsall, J. E., & Diggle, P. J. (1995). Kernel estimation of relative risk. *Bernoulli, 1*, 316.

King, B. F. (1967). Stepwise clustering procedures. *Journal of the American Statistical Association*, *62*, 86-101.

Leitner, M., & Curtis, A. (2002). Spatial variation in prenatal risks associated with infant mortality. In H. Wohlschlägl, M. Leitner, & N. Weixlbaumer (Eds.), *Geographischer Jahresbericht aus Österreich*. Vienna: Department of Geography and Regional Research, University of Vienna.

Longley, P., Goodchild, M., Maguire, D., & Rhind, D. (2001). *Geographic information systems and science*. Chichester, UK: John Wiley & Sons.

Monmonier, M. (1996). *How to lie with maps* (2nd ed.). Chicago: University of Chicago Press.

Moran, P. A. A. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*, 17-23.

Rushton, G., & Armstrong, M. (1997). *Lecture hall, improving public health through geographical information systems: An instructional guide to major concepts and their implementation*. University of Iowa, Department of Geography. Retrieved April 2, 2005, from http://www.uiowa.edu/~geog/health/index.html

Rushton, G., & Lolonis, P. (1996). Exploratory spatial analysis of birth defect rates in an urban population. *Statistics in Medicine, 15*, 717-726.

Scott, D. W. (1992). *Multivariate density estimation: Theory, practice, and visualization*. New York: John Wiley and Sons.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New York: Chapman and Hall.

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin, 38*, 1409-1438.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46(*2), 234-240.

Tobler, W. (1979). Cellular geography. In S. Gale, & G. Ollson (Eds.), *Philosophy in geography*. Dortrecht, The Netherlands: Reidel.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*, 236-244.

Williamson, D., McLafferty, S., Goldsmith, V., McGuire, P., & Mollenkopf, J. (1998, July). Smoothing crime incident data: New methods for determining the bandwidth in kernel estimation. In *Proceedings of the 18th Annual ESRI International User Conference*. Retrieved April 2, 2005 from http://gis.esri.com/library/userconf/proc98/PROCEED.HTM.

Worboys, M. F. (1995). *GIS: A computing perspective*. London: Taylor and Francis.

Chapter VII

Spatial/Temporal Stability in Neighborhoods of Risk: The Mobility of Mothers

Four years have been spent analyzing infant health data for Baton Rouge. As with working with any dataset, one gets a *feel* for the data. One such feeling was that the population being studied was mobile. The first indication came when trying to categorize the different neighborhoods in which Healthy Start program participants lived. The main reason for performing such an analysis was to gain an insight into any program participant entering the Healthy Start program, both in terms of her own previous birth outcomes and those of the neighborhood in which she lived (this will be revisited in the chapter on the Baton Rouge Healthy Start). In order to do this, one requirement was to trace pregnancy histories. As the birth and death certificate data did not come with unique identifiers, an alternative common field between birth certificates was needed. Originally, it was thought that a combination of street address, mother's date of birth, and the date of birth of the previous child would suffice. It soon became obvious that mothers living inside the five zip code region served by the Healthy Start project were mobile. A more in-depth investigation into the degree of this mobility was needed, as there could be implications both in terms of GIS analysis, and the subsequent targeting of outreach into those identified neighborhoods.

Although the benefit GIS can offer the health community is undeniable, it is vitally important that a researcher performing these analyses understand the space within which he or she is working. Although all cities will have their own characteristics, it is possible that a degree of mobility exists within the population under investigation. If this is so, a violation occurs of one of the commonly held assumptions of spatial analysis: that the population is spatially stable. As has been mentioned previously, the "law of small numbers" may produce instability in analysis results due to the increased variance in small data sets. However, to what degree is a further error introduced due to the mobility of the population? For example, if an infant dies, what is the most important address, the residence on the death certificate, the residence on the birth certificate, or the residence during the pregnancy? Three questions need to be answered: To what degree do mothers move, and if they are mobile, how far do they move? And probably most importantly, to what degree is there spatial stability in analysis results?

As an illustration of mother mobility, consider Table 1, which displays results for one of the zip codes within the Healthy Start service region. In this zip code, 657, 619, and 660 births occurred in the years 1996, 1997, and 1998, respectively. Of these births, between 10% and 20% were to mothers who had a previous birth within the 3-year time frame. Of the mothers who had a birth in both 1997 and 1996, only 42% remained at the same address. Of the mothers who had a birth in both 1998 and 1997, 52% remained at the same address, and between 1998 and 1996, 33% remained at the same address. All addresses were taken from the birth certificate. Although errors and variations do occur with birth certificate entry, the suggestion of mobility is strong enough (for this particular risk cohort) to warrant further investigation.

How Far Do the Mothers Move?

For the above 3 years a GIS was used to measure the distances between the two birth residences (if the mother had moved). Of the mothers who gave birth in both 1997 and 1996, 27% stayed within 0.5 miles of their 1996 residence (this percentage also includes those who did not move). Similarly, almost 79% of the mothers who gave birth in both 1998 and 1997 either lived at the same address or had moved no further than 1.5 miles. For the biggest gap between the years, of the mothers who gave birth in both 1998 and 1996, almost 54% either lived at

Table 1. Mobility and distances moved for repeat births between 1996 and 1998

Year	Births	Births in 97	% At Same Address	Births in 96	% At Same Address
1996	657				
1997	619			58	42
1998	660	66	52	117	33

the same address or had moved no further than 1.5 miles. Therefore, from this limited data set, although considerable movement occurs for this at-risk cohort, the majority of the moves are of a short distance, and probably within the same socially defined neighborhood. In order to give the reader some comparison measure, the average distance across a census block is between 0.1 and 0.3 miles.

Mother mobility can affect not only mapped data but also analysis results. This has implications in terms of the interpretation of results, and subsequent interventions. The likelihood of short move mobility is greater with poorer populations due to social "fragmentation," which means there is a lack of cohesiveness in the community (Whitley, Gunnell, Dorling, & Davey-Smith, 1999). In addition, the vast majority of pregnant mothers in the Baton Rouge Healthy Start program area are single, which again can result in a lack of spatial stability.

Several studies have been mentioned throughout this book that have involved the spatial analysis of health data at the intraurban or neighborhood level (Crosse, Alder, Æstbye, & Campbell, 1997; Pearl, Braveman, & Abrams, 2001; Pickett & Pearl, 2001; Roberts, 1997). Also, in Chapter VI several point and aggregated data models were discussed (Anselin, 1998; Duncan, Jones, & Moon, 1996; Fotheringham, Brunsdon, & Charlton, 1998; O'Campo, Xue, Wang, & Caughy, 1997; Openshaw, Charlton, Craft, & Birch, 1988; Reader & Bartolomeo, 2001; Rushton & Lolonis, 1996). These studies and models assume that the underlying population is stable.

Consider some of the environmental risk factors discussed in Chapter IV. Imagine a study that was associating negative birth outcomes to a point source of pollution (Elliott et al., 2001). In such an investigation distance plays a major role, or in geographic terms, the distance decay of interaction, which in this case can be defined as the closer the pregnant woman is to the polluting source, the greater the likelihood of a negative birth outcome. However, what if this residence, as captured on the birth certificate, only represents one temporal snapshot of the mother's life? If the mother had moved during the pregnancy, at what point would the exposure become a harmful influence? The actual birth residence may not be as important as where the majority of the pregnancy, or even where a specific section of the pregnancy, was spent (at what point is she most vulnerable, during the first, second, or third trimester?). Similarly, if a spatial analytical technique is used to identify patterns in infant mortality, what does a

Table 2. Distances moved between births

Distances	97 in 96	%	98 in 96	%	98in97	%	Total	%
0-0.5	27	46.55	52	44.44	43	65.15	122	50.62
0-1	33	56.9	55	47.01	47	71.21	135	56.02
0.1.5	37	63.79	63	53.85	52	78.79	152	63.07

"neighborhood of high infant death" really mean? For a mobile population, should these analyses be performed on the birth residence instead of the death residence? Further analysis of the infant birth and death certificates for 1996-1998 found that of 214 infant mortality deaths in East Baton Rouge Parish, 50 had different addresses on the birth and death certificate. Of these, six had a distance between residences of less than 1 mile, 20 were less than 3 miles, and 13 were greater than 7 miles (six of which moved out of the state). Looking at distance in another way, 10 of the 50 had addresses that were within the same zip code. One explanation for this could be that the mother had moved between the birth and death of the infant. Therefore, the neighborhood identified as having a high infant mortality rate may not reflect where the pregnancy actually occurred, or even where the birth residence was located.

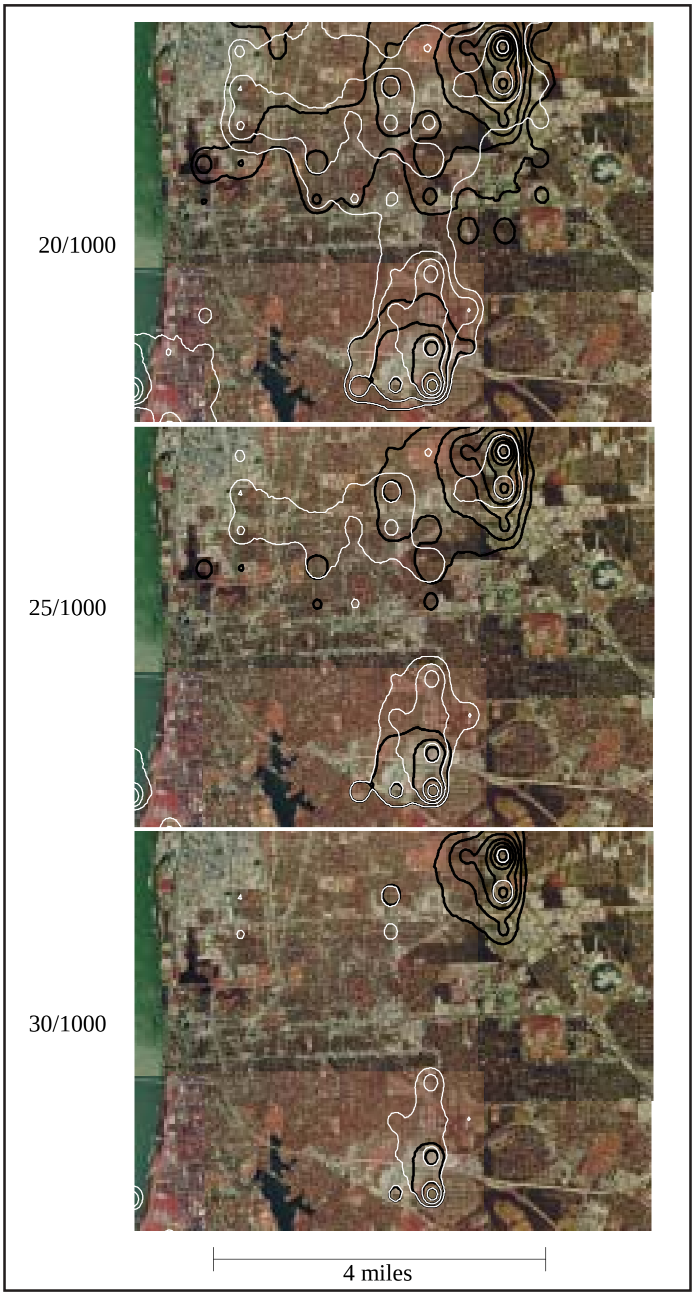
A further element of concern is the accuracy of the residence listed. This does not mean errors in data input, or errors associated with the road file against which the address is matched, but actual unexplainable differences that can occur between birth and death certificates. What is worrying is that in the previous example, for several deaths occurring within the first day(s), a different address was listed. There was no time for the mother to move. In some instances, a different address was listed on the birth and death certificate even when both occurred on the same day. So far no adequate explanation has been found. Could it be that an indigent mother wants to hide the birth from the family, especially if she is single, but then loses this desire for secrecy upon the death?

So can dramatic variations in analysis results occur from using the birth certificate address as compared to the death certificate residence? Consider Figure 1, which displays infant mortality rate isolines superimposed on airphotos for Baton Rouge. The black isolines are generated by spatial filter analysis (as described in Chapter VI) using birth residence of an infant death in 1998 as the numerator.

The lighter contours display rates generated when the death residence was used as the numerator. Of the 81 deaths that occurred in the parish during this year, 21 reported different residences on the birth and death certificate.

From looking at the three rate surfaces, it is obvious that depending on the rate chosen for visualization, considerable differences occur in the areas of the city deemed as being "at risk." Consider the map displaying an infant mortality rate (IMR) of 20/1,000. Although the two isoline surfaces generally overlap, there is a corridor that joins the northern and southern rate areas that is only identified when the death residence is used as the numerator. By switching scales (rates) to 25/1,000 a difference is evident between the two isoline surfaces in the northern part of the map, with the death residence as numerator surface being more marked towards the western side of the map, and the eastern side having a more intense clustering of birth-as-residence in the numerator calculation. By looking at the scale of the map, it is evident that approximately 2 miles of city

Figure 1. Comparisons of different rates of infant mortality



would be considered “at risk,” falling inside a 25/1,000 IMR isoline only if the death residence was used as the numerator. In the last map, a series of isolines suggesting an IMR cluster is found in the northern section of the city, but far more markedly so when the birth residence is used as the numerator.

Although the impact of mobility on health outcomes has been addressed in the literature (Bentham, 1988; Elford, Phillips, Thomson, & Shaper, 1990; Kliewer, 1992; Mancuso & Sterling, 1974; Polissar, 1980; Rogerson & Han, 2002), little if any has been written on the variations between birth and death certificate residence in the spatial analysis literature. I would suggest before any such analysis is performed that investigators identify the amount of variation that exists between residences, as it may be prudent to perform analyses with both numerators.

Temporal Stability and Implications for Outreach

Apart from causing a problem in the underlying assumption of a spatial analysis, an additional problem with a shifting population is "what if resources are targeted to the wrong areas?" This raises a larger question. For irrespective of whether mobility is the causative factor, to what degree are our results temporally and spatially stable? Can we rely on our results to act as predictors? In order to test for this, two investigations have been performed, both variants of analyses used by the Baton Rouge Healthy Start program. The first analysis develops a neighborhood risk index based on birth outcomes surrounding infant death locations. The second approach identifies mothers who had received no prenatal care (NPC) during their pregnancy (1996-1998) in order to develop generalized areas of concern for outreach purposes. Birth data for 1999-2001 is overlaid on these risk areas to validate their being targeted.

Obviously, if outreach workers were to be directed to neighborhoods of risk, it would be important that the data available (in this case 1996-1998) still had relevance for the current birth surface. As has been stated previously, there is often a considerable time lag involved in the application, approval, and receipt of vital statistics data. For the Baton Rouge Healthy Start project, data for 1999-2001 was received midway through 2002. It is therefore necessary to identify which results are temporally stable, irrespective of causation, so that outreach is still a valid option.

Developing a Neighborhood Categorization Scheme Based on Temporal Stability

As has been stressed throughout this book, to effectively bring about change, spatially precise targeting, or profiling (Badr Zahr, 2001), is needed. The purpose of developing a neighborhood categorization scheme is twofold. Firstly, a standardized means for prioritizing intervention can be developed that is understandable and defensible. In this way, for example, a program can justify why resources have been targeted on a particular neighborhood of the city. Program objectives can be set, goals developed, and eventually, results presented as part of a program evaluation. This develops the appreciation that the spatial setting of any program area is complex, "blanket-level" assessments are inappropriate, and that a geographic approach is needed to identify and understand individual neighborhood risks. The second reason is to understand the neighborhood around a single address. For the Baton Rouge Healthy Start, a typical point-based location is the residence of a program participant. In this way, temporally stable risks facing the program participant in her neighborhood can be identified, with tailored responses being developed. Obviously, the data contained on the birth certificate are limited in terms of information that will affect a program participant (for example, the month prenatal visits began will not be relevant), though some information, such as a high proportion of short gestation births or alcohol use during pregnancy, might help inform the caseworker about typical neighborhood stressors. This approach will become more valid with each passing year of the Healthy Start program as more comparison data is collected.

For illustrative purposes, instead of program participant addresses being the focus of these neighborhood investigations, actual mortality locations will be used as the center of the analysis. Table 3 displays the neighborhood categorization scheme designed for Baton Rouge. The first column identifies the temporal stability of infant mortality. The second column displays the temporal stability of risks found in the neighborhood surrounding the death as extracted from the birth certificates. For example, neighborhood category A, in every year, has a temporally stable, high IMR (high meaning the rate exceeds the parish-wide rate), and high risks at the 0.05 level of significance (from the same formulated difference of proportions test described in Chapter V).

In the categorization scheme shown in Table 3, one might think a neighborhood of stable infant mortality should be the first priority in terms of developing a mitigation strategy. Although reducing infant mortality is a noble mission, it is notoriously hard to identify any single causation at the neighborhood level. Having sat on several fetal and infant mortality review board meetings, it is

Table 3. Definitions of risk neighborhoods

A: Stable High IMR, Stable High Risks
B: Unstable IMR, Stable High Risks
C: Unstable IMR, Unstable High Risks
D: Stable High IMR, Unstable High Risks
E: Low / No IMR, Stable High Risks
F: High IMR, No / Little High Risks

evident that many deaths are the result of complex physical and social processes. Of course, we should not lose site of the importance of understanding and learning from each death in terms of effectively directing outreach strategies; it is just that targeting infant mortality may not be the most effective use of resources. Having said that, if the neighborhood surrounding a program participant has a stable high infant mortality rate, and from investigation of international classficiation disease (ICD) codes a common cause of death is identified (such as Sudden Infant Death Syndrome, or SIDS), specific education strategies could be developed during and after the pregnancy. Although causes for SIDS are still largely unknown, it might be that in this neighborhood a "traditional" practice of laying babies on their front will have to be addressed through "Back-to-Sleep" education initiatives.

Neighborhoods in categories B and E are also of importance for Healthy Start caseworkers, because although the IMR is not stable, at least one of the infant risk factors is stable over the 3 years. It is possible these risks could eventually manifest as a higher IMR, though an immediate problem is the relationship these risks have with infant developmental and behavioral problems. Again, when enough previous program participant data is available, the amount of useful analysis will expand dramatically (such as the proportion of previous program participants who experienced post natal depression). Category F might also generate further investigation as no temporal stability in risk, but a stable high IMR might suggest an environmental causation.

Constructing Neighborhoods Around Mortality Locations

The "neighborhood" around each death location (or around each program participant entering Healthy Start) was constructed by drawing a 0.25-mile buffer. This buffer size was chosen after several pretests and can be justified for Baton Rouge, as it is a fragmented city with neighborhoods changing both socially and racially within a few city blocks. In this way, a radius of 0.25 would be small enough to capture a relatively homogeneous population, while at the same time being large enough to minimize the small number effect of having too few births. For other cities, and as was previously mentioned in Chapter V, multiple distances should be tried as different patterns may emerge at different scales.

If two or more 0.25-mile buffers overlap, then the "neighborhood" was enlarged to include all connected deaths inside the buffer area. A neighborhood was only included in the analysis if 30 births were contained within the area of investigation.

The outline shape of each neighborhood was then transferred to the 2 remaining years. Therefore, a death location in year one generated a shape that was placed on the year two and three birth surface. As long as each neighborhood for these years also had 30 births, it was included in the 3-year temporal comparison. The same procedure was repeated for all death locations in year two, with the address being transferred to the year one and three birth surfaces, and for death locations in year three with the address being transferred to the year one and two birth surfaces. In this way, a 3-year comparison could be made based on data originating in each of the years, giving nine total areas under analysis.

A difference of proportions t-test (as described in Chapter V) was used to determine which of the chosen risk variables were significantly different at the 0.05 level, the comparison population being all birth data for that year in the parish of East Baton Rouge. For example, an original death location in year one, generating a neighborhood of 0.25 miles with at least 30 births, would be compared to all birth certificate data for the entire parish for that year. The same location the following year, if generating a neighborhood with at least 30 births, would be compared to all birth certificate data for year two. This would again be repeated for year three. In this way it is possible to see which risks were significant for 1, 2, or all of the 3 years. The 3 years under investigation were 1996, 1997, and 1998.

The one difference between applying this technique to actual Healthy Start program participants (as will be described in Chapter IX) is that the buffer radius is expanded for every previous year until the 30 birth threshold is reached. This is done because we want to know which risks are present in the neighborhood surrounding each program participant.

Temporal Stability in Risks Around Infant Deaths

Table 4 displays the significant risks for the neighborhoods associated with an infant death in 1996. The number of births per neighborhood remained relatively stable over the 3 years, the biggest difference being 31 births (neighborhood 96:9), the smallest being 1 (neighborhood 96:7). Death rates were more volatile because of the impact a single death can have on the relatively small birth population. Neighborhood 96:4 was categorized as A (stable IMR and stable risk), with the IMR exceeding parish levels in all of the years. Two risk categories (the number of prenatal visits and month prenatal care started) were also significant for the 3 years. Neighborhoods 96:2, 96:5, and 96:9 were all categorized as B (unstable IMR and stable risk). For all three of these neighborhoods, 2 high IMR years and 1 year with zero IMR were found. For all three neighborhoods, some combination of the number of prenatal visits and/or month prenatal care started was the stable risk. A possible risk trend was identified for neighborhood 96:2, with short periods between pregnancies being significant for 1997 and 1998. Neighborhoods 96:1, 96:6, and 96:8 were all categorized as C (stable IMR and unstable risk). For neighborhoods 96:1 and 96:6, 1 year also included an IMR spike exceeding 100 per 1,000 in both cases. Neighborhoods 96:3 and 96:7 are both categorized as D (unstable IMR and unstable risk). In both cases only 1 year experienced infant deaths, and significant risks were relatively few, though a possible trend was starting with number of prenatal visits for neighborhood 96:7.

Table 5 displays the significant risks for neighborhoods associated with an infant death in 1997. The greatest range in births was 17 (neighborhood 97:6), the smallest difference being one (neighborhood 97:3). The stability in births for neighborhood 97:3 is also interesting considering the relatively high number of births (80 to 81 per year). Deaths in neighborhood 97:3 were more varied, ranging from one in 1996 and 1998 to five in 1997. Neighborhoods 97:3, 97:6, and 97:7 were categorized as A, though the significant risks changed between the neighborhoods. For neighborhoods 97:3 and 97:6, the month prenatal care started was significant. For both neighborhoods 97:3 and 97:7, the number of prenatal visits was also significant, though in addition for neighborhood 97:7 the number of teenage pregnancies was also significant. Neighborhood 97:4 was categorized as B, with the number of prenatal visits being significant, and a possible trend emerging with the month prenatal care started. Incidentally, these 2 years (1997 and 1998) were the only years with deaths in the neighborhood. Neighborhood 97:9 was categorized as C, with an IMR exceeding the parish level by a magnitude of four during each of the years. Although no stable risks were present, possible trends were emerging with the number of teenage pregnancies and month prenatal care commenced. What is interesting about this neighborhood is that during 1997 no fewer than seven of the risk categories were significant. Neighborhoods 97:1, 97:2, 97:5, 97:8, and 97:10 were categorized as D. Neighborhoods 97:1, 97:2, and 97:5 experienced deaths in only 1 year. Most risks for neighborhood 97:1 were actually significant in the positive, meaning that in 1997 and 1998 pregnant women made significantly more prenatal visits than other women from the parish. A possible trend might be emerging for neighborhood 97:2, with the proportion of women having repeat pregnancies within 2 years being significant.

Neighborhoods 97:8 and 97:10 had deaths in 2 of the 3 years, though neighborhood 97:8 presented a more worrying picture, with deaths occurring in 1997 and 1998, along with significant risks for low birth weight and numbers of prenatal visits for the same 2 years. Table 6 displays the significant risks for neighborhoods associated with an infant death in 1998. The greatest range in births for any neighborhood in 1998 was 18 (neighborhood 98:2) and the smallest was three (neighborhood 98:11). Neighborhoods 98:4 and 98:5 were categorized as A, though the IMR for neighborhood 98:4 is at least a magnitude of three greater

Table 4. 1996 infant mortality locations and significant risks for 1996, 1997, and 1998

1996		Cat	Risk	Birth	Death	IMR	u18	u1500	u2500	#Pren	0pren	Mnpre	u32	u36	Prev B	Term	Tob	Alc
1	1998-1	C	N	52	1	19.23				*	*							
	1997-1			42	5	119	*		*	*					*			
	1996-1			58	1	17.24						*						
2	1998-2	B	Y	52	1	19.23	*			*		*			*	* +ve		
	1997-2			55	0					*		*			*			
	1996-2			49	3	61.22	*	*	*	*		*		*			*	
3	1998-3	D	N	27	0													
	1997-3			36	0					* +ve				* +ve				
	1996-3			39	2	51.28												
4	1998-4	A	Y	64	3	46.88	*	*		*		*						
	1997-4			50	2	40				*		*						
	1996-4			69	1	14.49				*		*		*	*			
5	1998-5	B	Y	45	2	44.44	*			*					*			
	1997-5			51	0		*			*		*						
	1996-5			49	2	40.82				*		*		*			* +ve	
6	1998-6	C	Y	34	1	29.4	*					*			*			
	1997-6			27	3	111.1												
	1996-6			40	3	75	*		*					*				
7	1998-7	D	Y	31	0					*								
	1997-7			30	0					*		*		*				
	1996-7			30	1	33.33	*											
8	1998-8	C	Y	56	3	53.57	*					*						
	1997-8			55	2	36.36			*	*							*	
	1996-8			57	4	70.18							*	*				
9	1998-9	B	Y	67	0	0			*	*		*		*		* +ve		
	1997-9			72	1	13.89						*						
	1996-9			98	5	51.02		*	*	*	*	*		*	*		*	*

IMR =Infant mortality rate per 1000 births

U18 = proportion of births to mothers 18 or under

U1500 = proportion of low birth weight babies

U2500 = proportion of very low birth weight babies

#Pren = proportion of mothers making 7 or less prenatal visits

0pren = proportion of mothers making 0 prenatal visits

Mpren = proportion of mothers making the first prenatal visit in the 4th month or later

U32 = proportion mothers with a premature birth

U36 = proportion of mothers with a very premature birth

PrevB = proportion of mothers who had given birth in the previous two years

Term = proportion of mothers who had previously had a termination

Tob = proportion of mothers who self reported smoking during pregnancy

Alc = proportion of mothers who self reported drinking during pregnancy

** = significant at 0.05 level*

Table 5. 1997 infant mortality locations and significant risks for 1996, 1997, and 1998

1997		Cat	Risk	Birth	Death	IMR	u18	u1500	u2500	#Pren	0pren	Mnpre	u32	u36	Prev B	Term	Tob	Alc
1	1998-1	D	N	33	0			*		* +ve			*					
	1997-1			30	1	33.33	* +ve			* +ve					* +ve			
	1996-1			16	0													
2	1998-2	D	N	33	0		* +ve		*					*	*			
	1997-2			39	2	51.28				*	*				*			
	1996-2			43	0											* +ve	* +ve	
3	1998-3	A	N	80	1	12.5				*	*	*			*			
	1997-3			81	5	61.73			*	*		*			*			
	1996-3			81	1	12.35				*		*				*	* +ve	
4	1998-4	B	Y	42	1	23.81				*		*						
	1997-4			38	2	52.63				*		*					*	
	1996-4			37	0					*	*							
5	1998-5	D	N	39	0									*		* +ve		
	1997-5			34	1	29.41				*								
	1996-5			42	0													
6	1998-6	A	Y	68	3	44.12	*	*		*		*						
	1997-6			54	2	37.04						*						
	1996-6			69	1	14.5				*		*		*	*			
7	1998-7	A	Y	43	1	23.26	*			*								
	1997-7			44	3	68.18	*		*	*		*		*				
	1996-7			55	3	54.55	*	*	*	*	*		*	*				
8	1998-8	D	Y	42	1	23.81			*	*								
	1997-8			53	3	56.6		*	*	*			*	*			*	
	1996-8			51	0													
9	1998-9	C	Y	65	4	61.54	*					*						
	1997-9			68	3	44.18	*	*	*	*		*	*	*				
	1996-9			67	4	59.7												
10	1998-10	D	Y	35	0													
	1997-10			30	2	66.67						*				*		
	1996-10			44	2	45.45				*		*			*		*	
11	1998-11		Y	20	0													
	1997-11			30	1	33.33	*											
	1996-11			23	5	217.4												

than the parish level in every year. Neighborhood 98:4 also has more stable risk factors, including proportion of teenage pregnancies, proportion of low-birth-weight babies, proportion of prenatal visits, and proportion of preterm deliveries. The problems of neighborhood 98:5 are concentrated on prenatal care, with both numbers of visits and month prenatal care began both being significant.

Neighborhoods 98:3 and 98:11 are categorized as B; in both cases only one of the years experienced an infant death. The stable risk factor varied between the two neighborhoods, the proportion of teenage births being significant for neighborhood 98:3, whereas the significant risk in neighborhood 98:11 was the month prenatal care began. Neighborhood 98:11 also had a possible trend developing with teenage pregnancies and number of prenatal visits for 1997 and 1998 being significant. Neighborhoods 98:6 and 98:8 were categorized as C, with risk factors

Table 6. 1998 infant mortality locations and significant risks for 1996, 1997, and 1998

1998		Cat	Risk	Birth	Death	IMR	u18	u1500	u2500	#Pren	0pren	Mnpre	u32	u36	Prev B	Term	Tob	Alc
1	1998-1	D	N	44	1	22.73						*		*				
	1997-1			35	0					*						* +ve		
	1996-1			46	0					*				*				
2	1998-2	D	N	63	3	47.62		*	*	*		*	*					
	1997-2			45	0													
	1996-2			50	1	20				*						*		
3	1998-3	B	Y	39	2	51.28	*			*	*	*				* +ve		
	1997-3			30	0		*									* +ve		
	1996-3			30	0		*					*		* +ve				
4	1998-4	A	Y	57	4	70.18	*		*	*		*		*				
	1997-4			63	4	63.49	*	*	*	*				*			*	
	1996-4			69	3	43.48	*	*	*	*	*		*	*				
5	1998-5	A	Y	96	4	41.67	*			*		*				* +ve		
	1997-5			91	2	21.98				*		*						
	1996-5			95	3	31.58	*			*		*			*		*	
6	1998-6	C	Y	39	2	51.28	*			*	*	*			* +ve			
	1997-6			44	1	22.73												
	1996-6			37	1	27.03				* +ve								
7	1998-7	F	N	36	3	83.3	* +ve			* +ve		* +ve						
	1997-7			24	0													
	1996-7			33	1	30.3	* +ve											
8	1998-8	C	Y	41	2	48.78	*											
	1997-8			40	2	50	*								* +ve	*		
	1996-8			34	1	29.41						* +ve						
9	1998-9	D	Y	37	2	54.05						*		*				
	1997-9			31	0		*					*						
	1996-9			27	0													
10	1998-10	D	Y	38	3	78.95		*				*	*					*
	1997-10			28	2	71.43												
	1996-10			35	0					*		*				* +ve		
11	1998-11	B	Y	39	2	51.28	*			*		*						
	1997-11			42	0		*			*		*						
	1996-11			39	0							*						

only being significant for neighborhood 98:6 during 1998, and with a possible trend starting in neighborhood 98:8 with significant proportions of teenage pregnancies. Neighborhoods 98:1, 98:2, 98:9, and 98:10 were categorized as D, with neighborhoods 98:1 and 98:9 experiencing deaths during one time period, and neighborhoods 98:2 and 98:10 experiencing deaths during 2 of the years. There is no stability in the risk factors, though a possible developing trend was in the month prenatal visits began for neighborhood 98:9. Interestingly enough, the year with the highest IMR in neighborhood 98:2 also had five significant risk factors. Similarly, in neighborhood 98:10 the highest IMR year had four risk factors. No other year for any of these neighborhoods categorized as D had more than two significant risks. Neighborhood 98:7 is categorized as F; that is, having a high IMR but no significant risk factors. Indeed, for this neighborhood there are four significantly positive risks (the neighborhood being apparently better than the rest of the parish).

Of all the neighborhoods identified in Tables 4 to 6, three stand out, one for each year, as having constantly high IMR. Yet for both 96:8 (lowest IMR = 36/1,000 and lowest number of births = 55) and 97:9 (lowest IMR = 44/100 and lowest number of births = 65), none of the birth certificate-extracted “risks” are significant at the 0.05 level for all years. It is with 98:4 (lowest IMR 43 and lowest number of births = 57) that statistically significant risks are found for all years, these risks including low birth weight (and almost very low birth weight), teenage births, number of prenatal visits, and short gestation births. Obviously, if a program participant came from this neighborhood, the caseworker should be concerned because we are probably only seeing the tip of the iceberg.

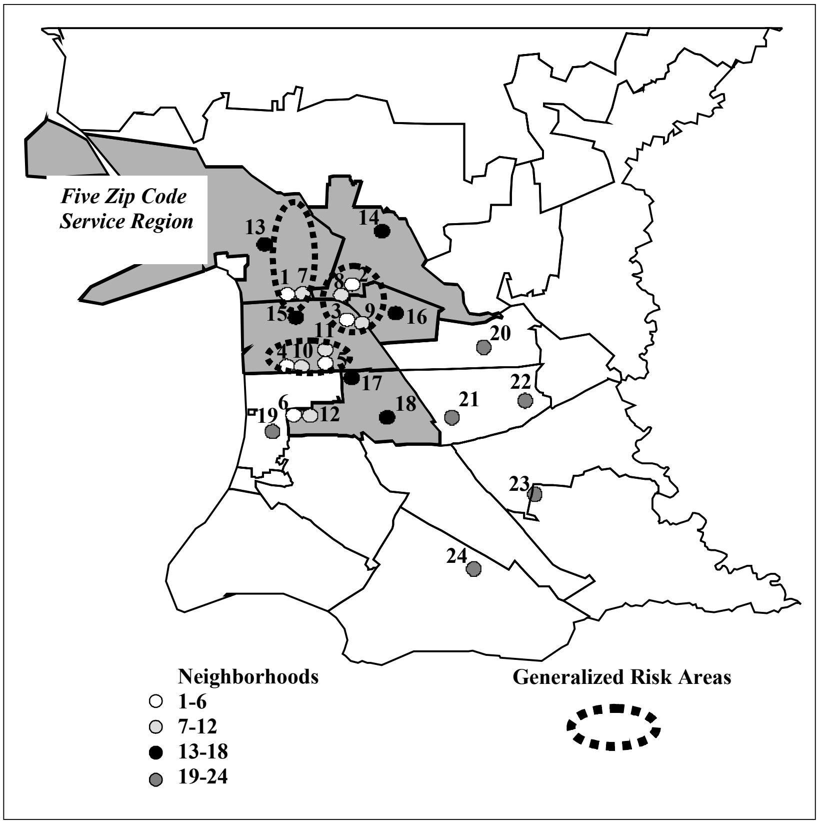
Temporal Stability in a Global Risk Investigation

During the previous analysis, actual neighborhoods were created around a single address. This address could be an Early Head Start Center (such as in Chapter V), a mortality location (as in this chapter), or the residence of a program participant entering Healthy Start. The alternative approach is to identify a complete surface of risk across the map, finding hot spots using techniques and approaches described in the preceding chapters. Again, the purpose of using these techniques is to gain an understanding of the risk surface so that mitigation strategies can be developed. For these to be successful, the risks identified must exhibit a degree of temporal and spatial stability. Confidence is dented if in identifying a neighborhood as having a high proportion of low-birth-weight births, in the year a strategy is developed and implemented no low-birth-weight births occur (presuming the strategy is not the cause of the success).

We have also discussed how temporal stability may be compromised through natural variability in the data, or because the population is relatively mobile. This next section will investigate the temporal stability for a risk surface generated from 1996 to 1998 births, with birth outcomes for 1999 to 2001 being overlaid for validation. The risk in question is arguably one of the greatest challenges for community based pregnancy action groups: women who receive no prenatal care during the pregnancy.

A spatial filter analysis (as described in Chapter VI) was used to identify hot spots of women receiving no prenatal care (NPC) across Baton Rouge. Using birth certificate data (the numerator was women making zero prenatal visits, the denominator was all births), three general areas of risk were identified across the city. These generalized areas were oval shapes designed to cover spatial filter contours at and above a rate of 50/1,000 NPC (See Figure 2). Within these generalized “risk” areas the GIS was used to create a window (a circle of 0.25-

Figure 2. Generalized areas of risk for women receiving no prenatal care and neighborhoods investigated for temporal stability



mile radius), which was moved to capture "neighborhoods." Any neighborhood that had a NPC birth in each of the 3 years was identified. In some ways, this is a similar approach to the categorization analysis which focused on mortality locations, the difference being that a global analysis was used to first concentrate the investigation, and then an NPC residence became a determination of the neighborhood — though not the center point. This approach is also useful as the smoothing that occurs with both the spatial filter calculation, and the subsequent contouring, may reveal hot spots that actually contain no addresses where the risk in question occurs. This secondary window approach can help tighten the focus within the general area of risk. In total, six of these neighborhoods (from this point referred to as 1-6) were identified. A further six neighborhoods (referred to as 7-12) of the same size were selected from within the generalized area to provide comparison to the original six. Neighborhoods 7-12 are immediately proximate to neighborhoods 1-6. Similarly, a third and fourth set of neighborhoods, again of similar size (referred to as 13-18 and 19-24) were selected for comparison. Both these third and fourth sets of neighborhoods originated outside the generalized area, although group three still came from within the five zip code service area of the Baton Rouge Healthy Start. The location of all neighborhoods, the five zip code area of the Baton Rouge Healthy Start, and the three generalized areas of concern can be seen in Figure 2. Although these "neighborhoods" were identified on the basis of NPC births, other traditional birth risks were also extracted from the birth certificates for comparison. For each of the four neighborhood groups (1-6, 7-12, 13-18, 19-24), the difference of proportion t-test was again used to identify statistically significant proportions of risk presence.

The purpose of this exercise was to see if the neighborhood contained within the generalized area of risk, or the more focused window focused on actual NPC addresses, provided better predictors of this risk than the two other comparison neighborhoods. In other words, is there enough temporal and spatial stability to allow for prediction, or are the risks randomly distributed?

Temporal Stability in the Four Neighborhoods

Table 7 displays NPC births for each of the neighborhoods (1-6, 7-12, 13-18,19-24). No specific outreach intervention was aimed at any of these neighborhoods during the 6 years under review. The total births for the four groups of neighborhoods were as follows: During 1996-1998 neighborhoods 1-6 had 377 births, neighborhoods 7-12 had 384 births, neighborhoods 13-18 had 222 births, and neighborhoods 19-24 had 268 births. Similarly, during 1999-2001 neighborhoods 1-6 had 327 births, neighborhoods 7-12 had 413 births, neighborhoods 13-18 had 216 births, and neighborhoods 19-24 had 301 births. When comparing the NPC rate per 1,000 between these two time periods, neighborhoods 1-6 dropped from 82.23 to 15.29, neighborhoods 7-12 rose from 23.44 to 33.9, neighborhoods 13-18 were stable at 27.03 to 27.8 and neighborhoods 19-24 dropped from 14.9 to 9.97.

If we consider the actual number of NPC births per individual neighborhood in the period 1999-2001, no neighborhood in 1-6 had an NPC birth in each of the 3 years. Two neighborhoods had NPC births in 2 of the 3 years, and two neighborhoods had no NPC births during the 3 years. The rate for these neighborhoods dropped from being considerably the highest to third for 1999-2001. Neighborhoods 7-12 displayed a general increase between 1996-1998 and 1999-2001. Only neighborhood 10 had NPC births in 2 of the 3 years between 1996-1998, however three neighborhoods had NPC births in 2 of the 3 years for the period 1999-2001. Neighborhood 7 had an NPC birth in each of the 3 years, the number increasing with each year, which would draw a flag as a potential area needing outreach. The NPC rate for this group of neighborhoods also rose from third to first. Neighborhoods 13-18 remained stable between the two time

Table 7. Number of births receiving no prenatal visits per neighborhood

Neighborhood	1996	1997	1998	1999	2000	2001
1	2	2	2	1	0	0
2	4	1	1	0	0	0
3	2	2	3	0	0	0
4	1	1	1	1	0	1
5	1	1	4	0	0	0
6	1	1	1	1	0	1
7	0	0	1	1	2	3
8	0	0	1	1	0	1
9	0	1	0	1	0	1
10	0	1	2	0	1	1
11	2	0	0	1	0	0
12	1	0	0	0	1	0
13	1	1	0	0	0	0
14	0	0	0	1	0	0
15	0	1	0	0	1	0
16	0	1	0	0	0	0
17	0	1	1	2	2	0
18	0	0	0	0	0	0
19	1	2	0	0	0	2
20	0	0	0	0	0	0
21	0	0	0	0	0	0
22	0	0	0	0	0	0
23	0	0	1	0	1	0
24	0	0	0	0	0	0

periods, both in terms of total NPC births and the rate. One neighborhood, 17, had NPC births in 2 of the 3 years for both time periods. Neighborhoods 19-24 had the lowest NPC rates for both time periods, with the majority of those births falling in neighborhood 19.

Results from the Difference of Proportions t-test

It should again be remembered that the resulting statistical significance level for the difference of proportions test is dependent on the selection of the comparative population, which here was East Baton Rouge Parish. For this analysis, all the births and the associated proportion of risks were combined for each neighborhood to give a group total and reduce the impact of small numbers. After all neighborhoods were combined together in their respective groups, only group 1-6 displayed a statistically significant proportion of NPC births for any of the years (1998, $p = 0.2$). Group 1-6 also had a statistically significant high proportion of births with seven or fewer prenatal visits in 4 of the 6 years (1996, $p = 0.05$;

Table 8. Significant risks for each of the neighborhoods

Neighborhood	No Prenatal	Less 7	After 5	Less 36	Less 32	Less 2500	Less 1500
1996							
1 to 6		***	**				
7 to 12							
13 to 18							
19 to 24							
1997							
1 to 6							
7 to 12		**	*	*			
13 to 18							
19 to 24							
1998							
1 to 6	*	**					
7 to 12		**	***				
13 to 18							
19 to 24							
1999							
1 to 6		*		*		*	
7 to 12					**		
13 to 18					*		
19 to 24							
2000							
1 to 6			*				
7 to 12		*	**				
13 to 18						*	
19 to 24							
2001							
1 to 6		*					
7 to 12		*		*			
13 to 18							
19 to 24							

(*) p = 0.2, (**) p = 0.1, (***) p = 0.05

1998, p = 0.1; 1999, p = 0.2; 2001, p=0.2). For 2 of these years, the same group had a statistically significant proportion of births to women initiating prenatal care in or after the fifth month (1996, p = 0.1; 2000, p = 0.2). The only other statistically significant risk for this group was low-birth-weight births and short gestational births (both in 1999, p = 0.2).

Group 7-12 had statistically significant numbers of women having seven or fewer prenatal visits in 4 out of the 6 years (1997, p = 0.1; 1998, p = 0.1; 2000, p = 0.2; 2001, p = 0.2). Similarly, in 3 of the years a statistically significant proportion of births in this group were to women who began prenatal care in the fifth month or later (1997, p = 0.2; 1998, p = 0.05; 2000, p = 0.1). Other statistically significant risks for the group include short gestational births (1997, p = 0.2), and very short gestational births (1999, p = 0.1). Of the two remaining groups (13-8 and 19-24), only group 13-18 had statistically significant findings, very short gestational births (1999, p = 0.2), and low-birth-weight births (2000, p = 0.2).

Conclusions on Temporal Stability

The results presented in Table 7 suggest that a general smoothing of the risk area (which does occur with spatial filtering and contouring) helps eliminate the fluctuations of selecting study areas that are too small. Again, this fluctuation results from the numbers of births, meaning that it is a problem of spatial scale. However, 19 NPC births were "predicted" in the two neighborhoods within the generalized area of risk, as compared to nine in the two comparison neighborhoods outside. Also, there was more stability in the other risks found in these two neighborhoods. In addition, Group 13-18 had more risks than in Group 19-24, which is reassuring as Group 13-18 falls inside the five zip code Healthy Start area, which was originally identified by spatial filter analysis. Therefore, spatial analysis performed on 1996 to 1998 data was useful in predicting outcomes for 1999 to 2001 births.

Suggestions for further research would include the analysis of risks for each of the individual neighborhoods, as the averaging effect of combining them into a group analysis could lose single neighborhoods that have consistently elevated risk factors. This is also a necessity for outreach, as neighborhoods that have stability in their risks over all years should be prioritized for help. A second direction for investigation is the impact the size of the neighborhood plays on capturing any inter-neighborhood moves. The original size of a 0.25-mile radius neighborhood was chosen because of outreach considerations, which include the need to be able to direct resources into a manageable area. If we consider Table 2, approximately 50% of all repeat births occur within 0.5 miles of both residences specified on the birth certificate. It is probable that a larger neighborhood of this size would display more temporal stability in the risk surface, as evidenced by results from the generalized risk region. It is certainly worthwhile to further investigate the impact that mobility has on temporal stability. This could be achieved by tracking program participants during their pregnancy, as is being done in the Baton Rouge Healthy Start project. The database has been constructed so that every change of address is captured, and distances, through a simple GIS manipulation, are recorded.

References

Anselin, L. (1998). *Spatial econometrics: Methods and models.* Dordrecht: Kluwer Academic.

Badr Zahr, L. K. (2001). Quantitative and qualitative predictors of development for low-birth-weight infants of Latino background. *Applied Nursing Research, 14*(3), 125-135.

Bentham, G. (1988). Migration and morbidity: Implications for geographical studies of disease. *Social Science and Medicine, 26*, 49-54.

Crosse, E., Alder, R., Æstbye, T., & Campbell, M. K. (1997). Small area variation in low birth weight: Looking beyond socioeconomic predictors. *Canadian Journal of Public Health, 88*(1), 57-61.

Duncan, C., Jones, K., & Moon, G. (1996). Health-related behavior in context: A multilevel approach. *Social Science and Medicine, 42*, 817-830.

Elford, J., Phillips, A., Thomson, A. G., & Shaper, A. G. (1990). Migration and geographic variations in blood pressure in Britain. *British Medical Journal, 300*(6720), 291-295.

Elliott, P., Briggs, D., Morris, S., de Hoogh, C., Hurt, C., Jensen, T. K., et al. (2001). Risk of adverse birth outcomes in populations living near landfill sites. *British Medical Journal (clinical research edition), 323*(7309), 363-368.

Fotheringham, A., Brunsdon, C., & Charlton, M. (1998). Geographically weighted regression: A natural evolution and expansion method for spatial data analysis. *Environment and Planning A, 30*, 1905-1927.

Kliewer, J. L. (1992). Influence of migrants on regional variations of stomach and colon cancer mortality in the western United States. *International Journal of Epidemiology, 21*, 442-449.

Mancuso, T. F., & Sterling, T. D. (1974). Relation of place of birth and migration in cancer mortality in the U.S.: A study of Ohio residents, 1959-67. *Journal of Chronic Disease, 27*, 459-474.

O'Campo, P., Xue, X., Wang, M., & Caughy, M. (1997). Neighborhood risk factors for low birth weight in Baltimore: A multilevel analysis. *87,* 1113-1118.

Openshaw, S., Charlton, M., Craft, A., & Birch, J. (1988). Investigation of Leukemia clusters by the use of a geographical analysis machine. *The Lancet, 1*, 272-273.

Pearl, M., Braveman, P., & Abrams, B. (2001). The relationship of neighborhood socioeconomic characteristics to birth weight among 5 ethnic groups in California. *American Journal of Public Health, 91*(11), 1808-1814.

Pickett, K., & Pearl, M. (2001). Multilevel analyses of neighborhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health, 55*, 111-122.

Polissar, L. (1980). The effect of migration on comparison of disease rates in geographic studies in the United States. *American Journal of Epidemiology, 111*(2), 175-182.

Reader, S., & Bartolomeo, B. (2001). *Identifying geographic areas of abnormally high and low birth weight incidence in Florida*. Paper presented at the Papers and Proceedings of the Applied Geography Conferences, Fort Worth, TX, November.

Roberts, E. M. (1997). Neighborhood social environments and the distribution of low birth weight in Chicago. *American Journal of Public Health, 87*, 597-603.

Rogerson, P., & Han, D. (2002). The effects of migration on the detection of geographic differences in disease risk. *Social Science and Medicine, 55*, 1817-1828.

Rushton, G., & Lolonis, P. (1996). Exploratory spatial analysis of birth defect rates in an urban population. *Statistics in Medicine, 15*, 717-726.

Whitley, E., Gunnell, D., Dorling, D., & Davey-Smith, G. (1999). Ecological study of social fragmentation, poverty, and suicide. *British Medical Journal, 319*, 1034-1037.

Chapter VIII

Patient Confidentiality

Different governmental agencies have long stored information in restricted-access databases. The advent of online data entry and analysis, and subsequent distribution of data to the public, has created a need for a more rigid set of visualization rules that preserve individual confidentiality. For example, when crime data are disseminated to the public in the form of crime maps via the Internet, law enforcement agencies have to balance between citizens' rights to know the dangers they face in their neighborhoods, while at the same time preserving the confidentiality rights of the victim. Similarly, in health data it is important to know which "risks" pregnant mothers face in particular neighborhoods, while preserving the actual birth outcomes of women living in those neighborhoods. An ex-graduate student of mine, who currently works for the Office of Public Health (OPH) in Baton Rouge, recently asked for a copy of the original paper on which this chapter is based. Her role in the OPH was as a GIS specialist with a particular emphasis on bioterrorism. The problem that she and the rest of OPH faced was how to publish information on the public's health on a Web site, while at the same time preserving patient confidentiality.

Patient confidentiality is a frequently discussed topic in both the medical and public health fields (Schwartzbaum & Hirschberg, 1991; Carman & Britten, 1995; Woodward, 1996; Bishop & Bishop, 2000; Cohen, 2000). Although articles have discussed how new technologies will impact patient confidentiality, particularly data input and dissemination via the Internet (Regan, 2001; Fowler, et al., 2000) relatively little has been written on how a visual output from a GIS, usually in the form of a map, can violate patient confidentiality. For example, there have been a considerable number of papers within both the geographic and public health literature that have employed GIS to either manipulate and visualize spatial data (Acevedo-Garcia, 2001), or to combine it with an expanding arsenal of point pattern and area association techniques (Fotheringham & Zhan, 1996; Gatrell, et al., 1996; Gatrell & Senior, 1999; Anselin, 1999). To what degree do the mapped outputs of these analyses violate confidentiality?

The display of confidential information in maps is determined by different factors, including map scale, symbol size, type of data, and intended audience. The smaller the scale of the map (which means a large geographic area is covered), the less detail can be shown and the lower the risk that confidentiality will be compromised. The risk of violating individual privacy is also reduced if larger point symbols are used to display the location of confidential information in a map. The type of data is also important. Some data are more private than others. There is less restriction needed when showing cases of flu in a neighborhood, because the general public provides the denominator population, than showing cases of HIV/AIDS, which has a decidedly smaller denominator base.

Another important factor influencing the display of confidential, personal data is the audience viewing these displays. In general, GIS-generated maps including medical data are intended for three different groups of audiences. The first is for publication, usually as articles in a journal, reports or books. For this audience, the primary importance is in preserving the underlying relationships between the different spatial data layers, as the graphics are companion pieces to the text. An example display would consist of three different spatial data layers: a map showing locations of infant deaths in the form of residence locations of mothers, a map showing significance contours of infant mortality (produced as output from an analysis such as the spatial filter), and a choropleth map displaying the proportion of African Americans per census tract. The text to the paper would comment on these relative associations. Unfortunately, access to new technology now allows for these displays to potentially reveal information about actual medical cases. For example, a map produced for a journal could be removed, scanned, and imported as a graphic into a GIS. The map could be georegistered and used to reveal where actual infant death residences are located. It is therefore important to manipulate the displayed residences to prevent this from happening. However, if the data were manipulated, such as the infant death residences being randomized, care would be needed in preserving the spatial relationships so that the deaths would still fall within the correct contours of significance and within census tracts displaying the correct proportion of the at-risk cohort.

The second audience is for a public presentation. This type of presentation is usually for concerned community residents, or possibly local area politicians. It could also be a formal presentation at a conference, the audience being from academia, government, or the private sector. This second group of people will probably not be familiar with the area or region shown in the presentation. However, in both cases the presentation is virtual, in that the image lasts for the length of the associated segment of talk (usually a matter of minutes) and no hard copy is distributed. For this type of presentation less vigilance is needed in masking actual areas of the city. Indeed, local groups will want actual references to show the extent of the problem in their neighborhood. Even individual residences, randomized slightly, can be displayed to show visual clusters, as the chance of patterns leading to recognition of individuals is virtually nil. The audience is shown a clustered pattern of their neighborhood, and then the image is removed. The gravity of the situation is expressed without compromising patient confidentiality. One word of warning with this category though, as most presentations are now made through PowerPoint and a typical session consists of five to six presenters — the usual procedure is to load all presentations onto a central computer. It is imperative that all copies of the presentation are removed — indeed, ideally the presentation should be run from an external drive.

The third audience can be classified as "expert," usually comprised of local health workers who will act on the findings of the analyses. As long as appropriate clearances have been met, no masking of information is needed for this audience. Indeed, for this audience the data has to be presented in an accurate fashion so that contingency plans can be formulated. For example, the exact address of the "at-risk" apartment complex is needed so that testing for lead paint can be performed.

Confidentiality in Maps

Whenever information is displayed in a map, two types of confidentiality issues need to be addressed: (1) statistical or attribute confidentiality, and (2) spatial or locational confidentiality.

Statistical (Attribute) Confidentiality

This type of confidentiality is associated with individual information usually recorded as text, in a table, or in a spreadsheet format. This information can be referred to as statistical or attribute information. It may include generic information (age, date of birth, marital status, religious affiliation, etc.) and more specific information of interest (tobacco or alcohol use, criminal background, class transcripts) to the collecting organization (hospital, police, university registrar, etc.). For this type of information, standards and rules have been developed in order to protect individuals' privacy. For example, the U.S. Department of Health and Human Services issued the Privacy Rule to implement the requirement of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). This rule will be discussed in detail below.

Spatial (Locational) Confidentiality

This type of confidentiality is mostly associated with the visualization of individuals' statistical information in maps. For example, in a detailed map of a city's neighborhood, the residence of an individual can be easily displayed with a dot at the exact location where the individual lives. If the statistical information that is associated with these individuals is "mothers who recently had an infant die," then every mother could easily be identified through the location of her residence on the map. Such a map would clearly violate the privacy of such individuals.

Preserving Confidentiality in Governmental Agencies

Current standards and rules to protect an individual's privacy have been developed for statistical or attribute data. Unfortunately, little exists for the protection of spatial or locational confidentiality. Due to the lack of such rules and to play it safe, organizations would disclose individual-level records at an aggregated level (county or state). This certainly protects the individual's privacy. On the other hand, aggregating individual records to this level will undermine people's right to know about the distribution of health-related or other data in their neighborhood.

In the remainder of this chapter, standards for protecting individual privacy will be addressed. This discussion is focused on preserving statistical or attribute confidentiality. Much of the discussion will be devoted to the Privacy Rule issued by the U.S. Department of Health and Human Services (HHS), followed by recommendations for how to preserve individual privacy from the U.S. Census and the U.S. Department of Justice. After that, new research aimed to develop guidelines for protecting spatial or locational confidentiality is introduced.

U.S. Department of Health and Human Services

The Standards for Privacy of Individually Identifiable Health Information ("Privacy Rule") establishes, for the first time, a set of national standards for the protection of certain health information. The U.S. Department of Health and Human Services issued the Privacy Rule to implement the requirement of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The following is just a brief synopsis about the Privacy Rule. The entire document can be found at the following Web site: www.hhs.gov/ocr/hipaa.

The Privacy Rule standards address the use and disclosure of individuals' health information — called "protected health information" by organizations subject to the Privacy Rule — called "covered entities," as well as standards for individuals' privacy rights to understand and control how their health information is used. A major goal of the Privacy Rule is to assure that individuals' health information is properly protected, while allowing the flow of health information needed to provide and promote high quality health care and to protect the public's health and well-being (www.hhs.gov/ocr/hipaa/privacy.html).

The Privacy Rule protects all "individually identifiable health information" held or transmitted by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral. Individually identifiable health information includes many common identifiers (such as name, address, birth date, and social security number). There are no restrictions on the use or disclosure of deidentified health information. Deidentified health information neither identifies nor provides a reasonable basis to identify an individual. There are two ways to deidentify information: either (1) a formal determination by a qualified statistician, or (2) the removal of specified identifiers of the individual and of the individual's relatives, household members, and employers is required. A long list of identifiers exists, including some locational information, such as street address, city, county, precinct, and zip code. If all zip codes with the same three initial digits contain more than 20,000 people, then these three initial digits can be disclosed; otherwise, the initial three digits are changed to 000 (www.hhs.gov/ocr/hipaa/privacy.html).

The Privacy Rule clearly states how individual and statistical health information should be protected. No guidelines are offered as to how to protect health information visualized in maps generated by a GIS. If the same Privacy Rule established for attribute health information (see above) is also applied to locational data, then health information could only be distributed in map form aggregated to the state or zip code level — as long as all zip codes with the same three initial digits contain more than 20,000 people. This would create highly generalized maps and certainly not comply with people's right to know about the distribution of health-related data in their city or neighborhood. For this reason, better and more appropriate privacy rules for map displays need to be established.

U.S. Census

Title 13 United States Code, Section 9, prohibits the Census Bureau from publishing results in which an individual's or business's data can be identified. The U.S. Census Bureau uses different methods of disclosure limitations to protect the confidentiality of data, including suppression, data swapping, and protection of microdata files (see http://factfinder.census.gov). Suppression is a method of disclosure limitation by not showing (suppressing) the cell values in tables of aggregate data for cases where only a few individuals or businesses are represented. Data swapping is another method of disclosure limitation designed to protect confidentiality in tables of frequency data. Microdata files contain data from censuses of the United States population and household surveys. These files contain individuals' responses that represent only samples of the population and have had all individual identifiers (such as name and address) removed from the records. In addition, to protect confidentiality, the Census Bureau may modify distinguishing characteristics (such as high levels of income) and restrict geographic identifiers (such as the name of a city) so that populations are composed of at least 100,000 people. Among the three methods of disclosure limitation, suppression might be the only one applicable to preserve spatial confidentiality. A detailed discussion of how suppression would be applied to spatial confidentiality will be discussed later in this chapter.

U.S. Department of Justice

Law enforcement agencies throughout the United States have been providing crime maps and data to the public for many years. Every department has its own policy on what data can be released, in what form, and to whom. The Freedom of Information Act, state law regarding public records, and each agency's philosophy shapes policies (Wartell & McEwen, 2001). According to a Web site updated by the Mapping and Analysis for Public Safety (MAPS) program, about 50 local law enforcement agencies in the United States now provide online data and maps (see www.ojp.usdoj.gov/nij/maps/weblinks.html). A look at these Internet maps reveals that indeed no general rules for displaying confidential crime data exist. For example, in some maps, crime data are summarized by police beats, whereas in others, crime data are placed at the midpoint of the street segment, or at the closest street intersection.

Geographically Masking the Location of Confidential Point Data

The information presented in this section summarizes the main research results of an ongoing project that the authors of this book have already published elsewhere (Leitner & Curtis, 2003, 2004). This research discusses different geographic masking strategies of point locations that protect the confidentiality of individuals, and at the same time preserve essential visual characteristics of the true, original spatial distribution of those locations. Five different global and five different local geographic masking methods (for a total of ten) were tested and compared. Armstrong et al. (1999) introduced the term "geographic masking" into the literature, suggesting similar geographic masking methods that are applied in this research. Whereas their research discusses the influence geographically masked data have on the results of geographically based analyses, the research reported here identifies acceptable design solutions for presenting confidential point data on a map. Acceptable design solutions define any geographic masking method that would preserve as many visual spatial characteristics as possible, while reducing the likelihood of individual identification to an acceptable level. A review of the literature suggests that the methods reported here are the first to utilize empirical perceptual studies to assess methods for presenting confidential point data on maps.

The subject matter of this research relates to an area of geography and related disciplines that has been receiving a fair amount of attention lately, especially as it pertains to the mapping of health and crime information (Armstrong et al., 1999; Wartell & McEwen, 2001; Monmonier, 2002; Leitner & Curtis, 2003, 2004). There is a real need to develop appropriate guidelines for mapping the location of individual-level data that is considered to be confidential, because at this time no standard guideline exists even though more and more governmental agencies have begun to disseminate their data on maps via the Internet. According to a Web site updated by the MAPS program (sponsored by the National Institute of Justice – NIJ), about 50 local law enforcement agencies in the United States now provide online data and maps (see www.ojp.usdoj.gov/cmrc/weblinks/welcome.html). Access to this type of data will certainly continue to increase in the future.

Experimental Testing

An experiment was developed to see how accurate test subjects could identify hot spots, such as infant mortality clusters, if those mortality locations were randomly varied across space. The experiment included a series of map pairs. One map of each pair always showed all deaths in their correct location. In the second map, all deaths were geographically masked, either locally or globally. The deaths in each map pair were displayed in yellow on top of either a blue background, census tract boundaries, or a street network. All 82 participants completing the experiment were asked to complete two tasks. The first task was to compare the two point pattern maps and decide if the two were similar or different. The second task involved identifying areas within one of the two point pattern maps (in most instances this was the map for which the deaths were geographically masked, but the participant did not know that) that showed a high mortality concentration.

The results of both tasks satisfy the one important objective of this study; namely, to investigate how much each geographic masking technique changes the original pattern of deaths. This objective is addressed in two ways: first, by comparing the change in the overall point pattern, and second, by comparing the change in the number and location of hot spots. Both comparisons were tested in the experiment described above. A second important objective is to investigate to what extent one might be able to identify an individual death residence after this location has been geographically masked. The interpretation and conclusions of the results from this second objective are difficult, because there are no standard guidelines to preserve spatial confidentiality. The overall premise of this test is to find geographic masking methods that would preserve as many visual spatial characteristics as possible (first objective), while reducing the likelihood of individual identification to an acceptable level (second objective).

Results for Global Geographic Masking

In global geographic masking, individual deaths are spatially displaced by the same exact amount. Figure 1 compares the original death locations (Figure 1a) with the same death locations after being geographically masked by five different global methods (Figure 1b through 1f). For privacy concerns, no other reference points (such as street network, churches, schools, etc.) are shown on the map. Three deaths are labeled to show the effect that each global masking method has on the mortality surface.

In Figure 1b, the original deaths are flipped about the horizontal central axis of the map, and subsequently placed on top of the closest street segment. In Figure 1c, all deaths are flipped about the vertical central axis of the map and then moved on top of the closest street segment. In Figure 1d, deaths are flipped about both the horizontal and the vertical central axes of the map and then placed on top of the closest street segment. In Figure 1e, all deaths are rotated around the map center either by 60 degrees to the right or by 120 degrees to the left (Figure 1f) and subsequently placed on top of the closest street segment.

The global masking methods presented in Figure 1 are just some examples of how death residences could be manipulated. Many more global masking methods exist. One major advantage of all global masking methods is their quick and efficient implementation. On the other hand, global masking methods change the locations of points within the map rather drastically, as can easily be seen when comparing the original mortality distribution in Figure 1a with any of the five globally masked point patterns (Figures 1b through 1f). The visual impression is confirmed by the result of the experiment. All but one globally masked point pattern was perceived to be more different (than similar) to the original point pattern. Only when points were flipped about the vertical central axis of the map did test subjects perceive the masked point pattern more similar to the original pattern. But when the size, shape, location, and number of perceived hot spots from this globally masked point pattern is compared to the original point pattern, differences become apparent. In the masked point pattern, hot spots are clearly flipped about the vertical central axis of the map. A second observation is that no differences exist across the three base maps for both unmasked and masked point patterns (Figure 2).

These results point to a particular danger of using an inappropriate masking method: It can lead to hot spots appearing in neighborhoods where no such risk occurs. As this “revelation” could lead to increased pregnancy stress for women residing in these neighborhoods, the masking may actual contribute to negative birth outcomes.

An additional drawback of global masking methods is that it would be easier (compared to local masking methods) to recognize the type of manipulation and to recreate the original point distribution. If this were done, individual privacy would be compromised and no longer preserved. For these reasons, none of the five global masking methods presented here are appropriate for visualizing the location of confidential point data on a map.

Results for Local Geographic Masking

In local geographic masking, mortality locations or groups of deaths that are close together are spatially displaced by different amounts (Figure 3). Among the many local geographic masking methods that exist, five are presented here. For three of these five methods, a regular grid is first superimposed over the selected study area. The size and shape of each grid cell can vary, but in this example it is a square measuring 500 meters per side. Incident locations falling into the same grid cell are spatially displaced by the same exact amount, but this displacement vector changes randomly between grid cells (Figure 3).

Figure 1. Examples of global geographic masking methods: (a) Original death locations; (b) Flipping about horizontal central axis of the map; (c) Flipping about vertical central axis of the map; (d) Flipping about both central axes of the map; (e) Rotating around the map center by 60 degrees to the right; and (f) Rotating around the map center by 120 degrees to the left (Courtesy of M. Leitner and A. Curtis and Cartographic Perspectives, 2004)

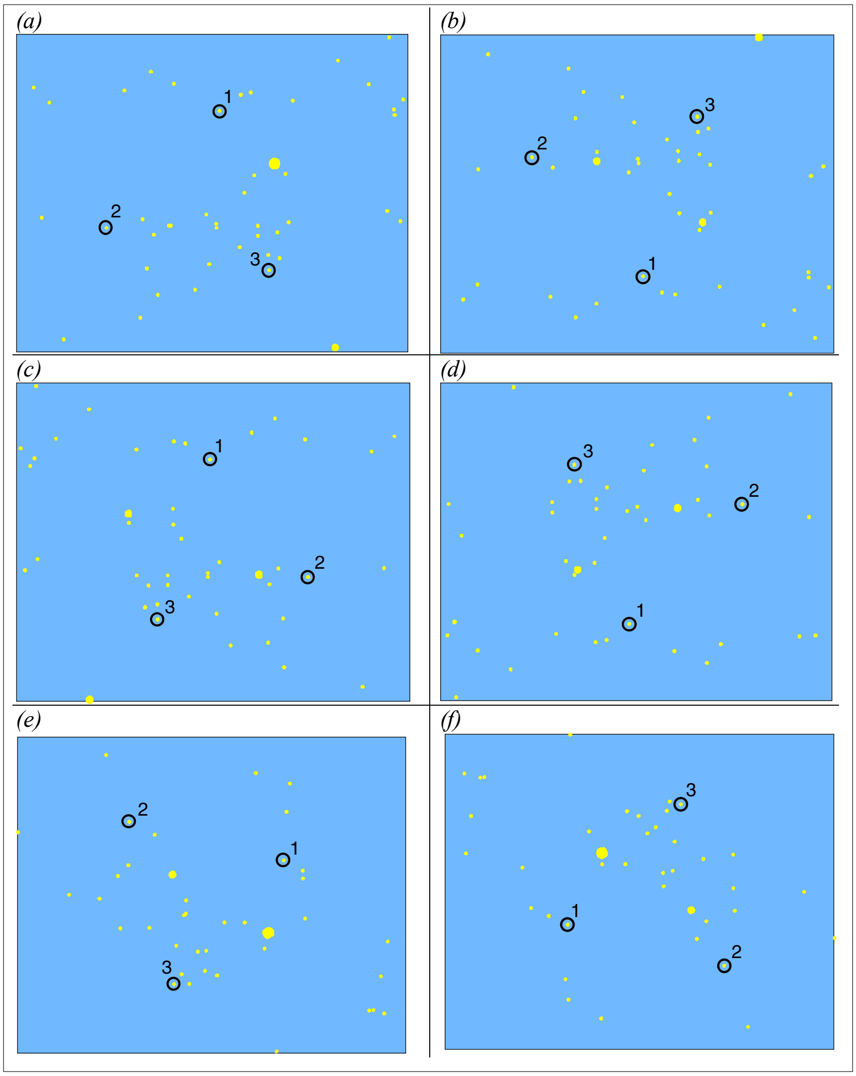


Figure 2. Comparing hot spots between the original and one globally, geographically masked point pattern for three different base maps (the base map information is not included in any of the maps) (Courtesy of M. Leitner and A. Curtis and Cartographic Perspectives, 2004)

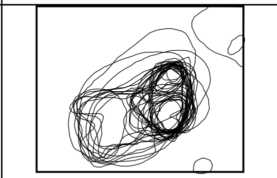
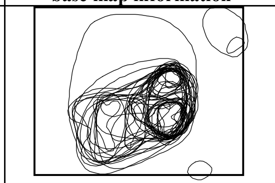
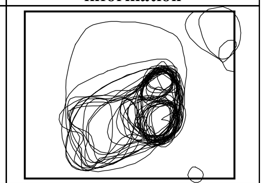
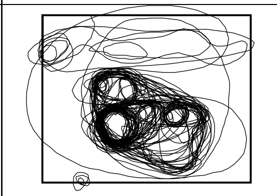
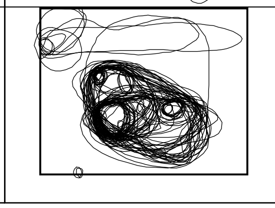
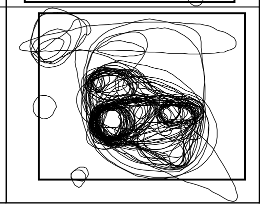
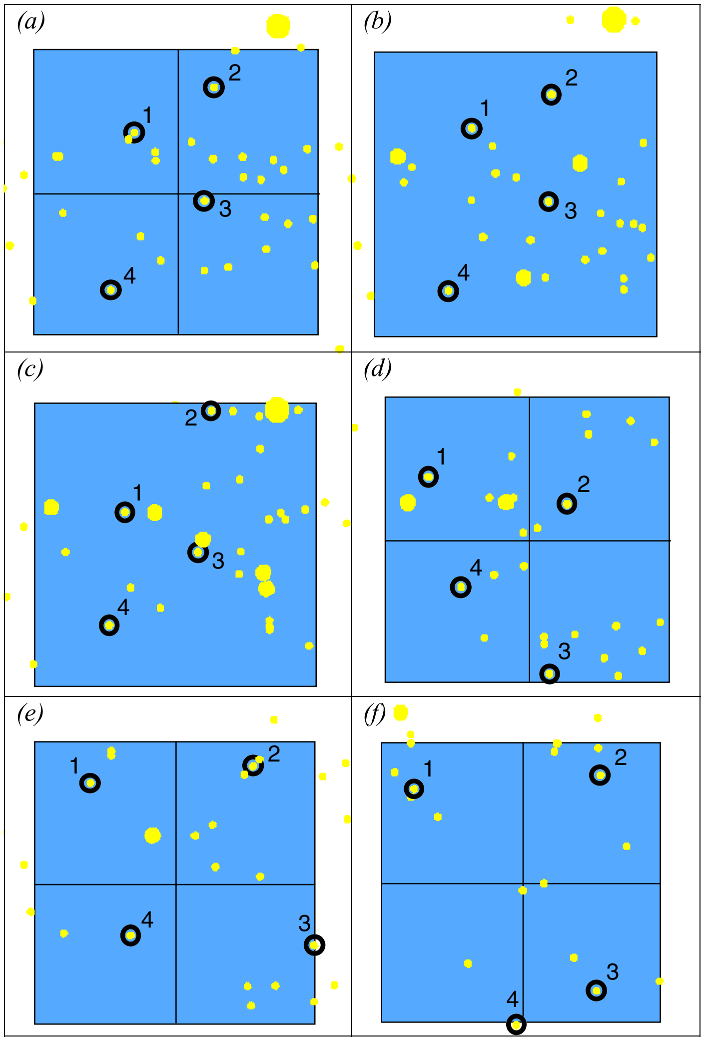
	No base map information	Census tract boundaries as base map information	Street network as base map information
Original, geographically unmasked point pattern			
Geographically masked by flipping point locations about the vertical central axis of the map			

Figure 3 displays the original deaths (Figure 3a) together with the same locations after being geographically masked by five different local masking methods (Figures 3b through 3f). To protect the privacy of the individual, only the blue background is used in all maps. Four sample locations are selected and labeled so their locations can be traced between the original and the five locally masked maps. Note that Figure 3 shows only a subset of the point pattern depicted in Figure 1.

The first local masking method aggregates deaths at the midpoint of their street segments (Figure 3b). A street segment is defined as the portion of a street between two adjacent street intersections. In Figure 3c, deaths are aggregated to their closest street intersection, which is defined as the intersection of three or more streets. The next three local masking methods are based on a regular grid. In the first instance, deaths are flipped either about the vertical, horizontal, or both central axes of each grid cell and then moved on top of the closest street segment. The type of flipping changes randomly between the cells of the regular grid (Figure 3d). For example, the mortality location labeled as #1 is flipped vertically; the other three deaths (#2, #3, and #4) are each flipped horizontally. In Figure 3d, deaths are first rotated by some random degrees around the center of each grid cell and then placed on top of the closest street segment. Mortality location #1 is rotated by 120 degrees to the left; #2 by 60 degrees to the right; #3 by 240 degrees to the right; and #4 by 120 degrees to the right. In Figure 3f, deaths are first translated by some random distance and then moved on top of their closest street segment. The mortality location labeled as #1 is translated 150 meters in x- and 350 meters in y-direction from the lower left corner of its grid

Figure 3. Examples of local geographic masking methods: (a) Original incident locations; (b) spatial aggregation at the midpoint of the street segment; (c) spatial aggregation at the closest street intersection; (d) flipping randomly either about the vertical, horizontal or both central axes of each grid cell; (e) rotating by some random degree around the center of each grid cell; and (f) translating by some random distance (Courtesy of M. Leitner and A. Curtis and Cartographic Perspectives, 2004)

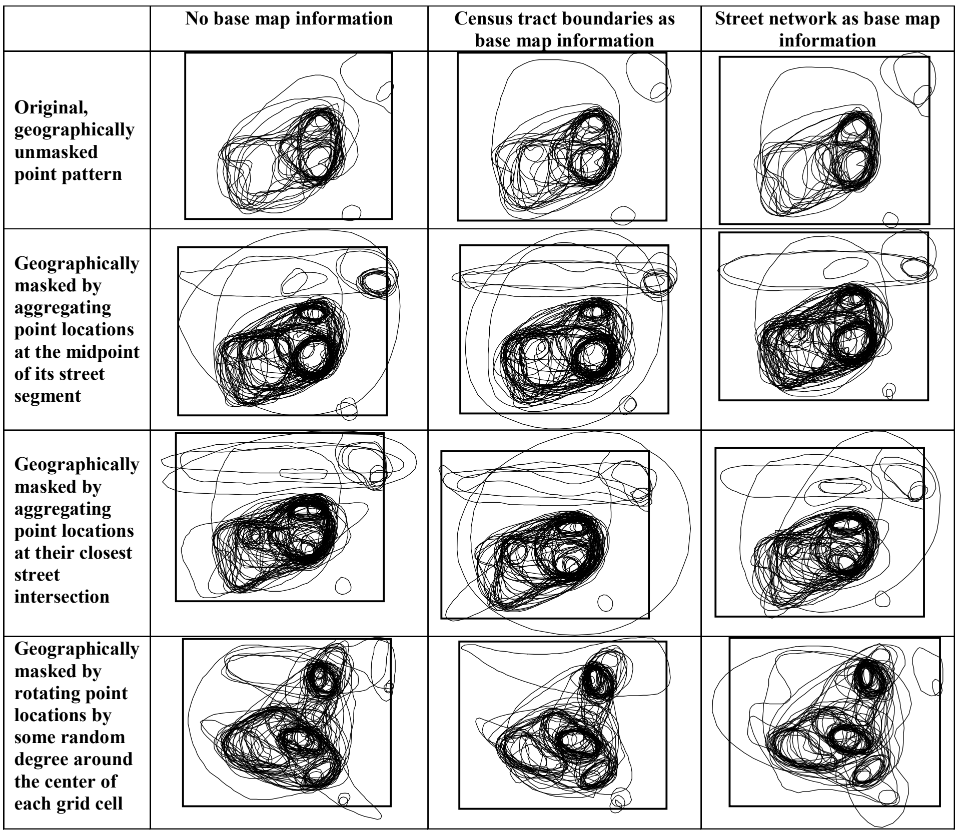


cell; #2 is translated 300 meters in x and 400 meters in y; #3, 450 meters in x and 50 meters in y; and #4, 300 meters in x and 200 meters in y.

In this example, the size of the grid cells, the angles of rotation and the translation distances are chosen arbitrarily when masking point patterns. Clearly, different choices may have yielded different results. For example, smaller rotation angles and shorter translation distances in Figures 3e and 3f would have resulted in shorter point movements. Consequently, these two masked point patterns would have been more similar to those patterns exhibited in Figures 3b and 3c. In addition, smaller cell sizes would have shortened point movements and vice versa.

The results of the experiment revealed that three of the five locally masked point patterns were perceived to be more similar (than different) to the original point pattern. The perceived hot spots for these three point patterns are displayed in Figure 4. Similar to all global masking methods, the results clearly show that the

Figure 4. Comparing hot spots between the original and three different geographically masked point patterns for three different base maps (the base map information is not included in any of the maps) (Courtesy of M. Leitner and A. Curtis and Cartographic Perspectives, 2004)



size, shape, location, and number of hot spots change little across the different types of base maps (no base map, census tract boundaries, and street network). There seems to be evidence that the perception of hot spots is not dependent on the type of base map that is used with the point pattern.

The visual comparison between the unmasked and each of the three locally masked point pattern is more complex. Specifically, there seems to be much agreement in the size, shape, location, and number of hot spots between the original, unmasked point pattern, and the two masked point patterns that aggregate their mortality locations either at the midpoint of their street segment or at their closest street intersection. This is an important result, and it makes these two local masking methods very useful, appropriate tools to visualize the location of confidential data. A further observation is that of a hierarchical arrangement of differently sized hot spots. There are two small (local) hot spots to the right of the center of each map. These two are combined into one medium-sized hot spot. Then there is a larger (regional) hot spot apparent across the lower portion of each map that encloses the two small and the one medium-sized hot spot.

However, differences in the size, shape, location, and number of hot spots are visible when comparing the original point pattern with the third locally masked point pattern. In this case, a distinct, local hot spot appears in the upper half of the map, just right of the center. This particular hot spot is not visible in the original and in any of the other masked point patterns. This again points to a particular danger of using an inappropriate masking method, namely the likelihood of an incident cluster to appear in a neighborhood of a city, where such a cluster does not exist in reality.

Preserving Spatial Confidentiality of Two Locally Masked Point Patterns

The previous analysis found that only two of the five local and none of the five global masking methods are appropriate for visualizing the location of confidential point data on a map. The following final analysis investigates to what extent one might be able to identify an individual residence after this location has been geographically masked by aggregating at either "the midpoint of the street segment" or "at the closest street intersection."

To answer this question, a sample of fourteen different street addresses were randomly selected from the total of 48 addresses used in the test. All addresses are located in a residential neighborhood in Baton Rouge with mostly single-family houses. Churches, schools, office buildings, etc. are interspersed between the residences. Each of the fourteen residences associated with the street address from the sample was visited, and the number of residences on either side of the street segment that included the residence of interest was counted. Accordingly, all residences that were located on either side of all street segments of the street intersection to which the residence of interest was closest to were also counted. The results in Table 1 show that the number of residences along either side of a street segment ranges from a minimum of two to a maximum count of 29 residences. If the number of residences along the street segments with a common intersection are counted, then the minimum number is 15 and the maximum is 58. Both minimum and maximum numbers include the confidential street address.

The question that needs to be addressed is: What is the minimum number of all residences, including the residence of interest, so that the privacy of an individual residing at this address is not compromised? Recall from the beginning of this chapter that current standards and rules to protect individuals' privacy have been developed for statistical or attribute data. Unfortunately, no such standard exists for the protection of spatial or locational confidentiality. The following discussion explores what would happen if the current standards developed for attribute data would also be applied to protect the privacy of locational data.

Table 1. Comparing the number of residential buildings along one street segment or along all street segments with a common intersection (the location of one of the residential buildings is confident) (Courtesy of M. Leitner and A. Curtis and Cartographic Perspectives, 2004)

Confidential location of individual residences, numbered consecutively	Number of residences on either side of the street segment containing the confidential location	Number of residences on either side of all street segments with a common intersection. One of these street segments includes the confidential location
1	22	50
2	29	52
3	4	20
4	12	29
5	9	17
6	2	25
7	28	54
8	6	18
9	2	24
10	9	15
11	25	26
12	27	58
13	24	34
14	29	34
TOTAL	228	456
MEAN	16.3	32.6
MINIMUM	2	15
MAXIMUM	29	58

Applying the HHS Privacy Rule to locational data would mean that health information could only be distributed in map form aggregated to the state or zip code level — as long as all zip codes with the same three initial digits contain more than 20,000 people. This would create highly generalized maps and certainly not comply with people's right to know about the distribution of health-related data in their city or neighborhood. For this reason, better and more appropriate privacy rules for map displays need to be established.

Among the disclosure limitation methods used by the Census Bureau, suppression will be applied to protect the privacy of locational data. In an example to explain suppression, the Census Bureau provides two tables, one with the original data not being suppressed (Table 2a), the other one with the same data being suppressed (Table 2b). Comparing the content of both tables indicates that cell values for up to three individuals or businesses are suppressed, but a cell value of seven is not. Accordingly, the U.S. Census Bureau must use a cutoff value for data suppression that lies somewhere between four and six.

If the same guidelines for data suppression were applied to the information in Table 1, then four of the fourteen residences (#3, #6, #8, and #9 from Table 5) for the masking method "aggregating incident locations at the midpoint of its street segment" would have been suppressed. In other words, such geographically masked locations should not be displayed in a map, because the privacy of individuals living at these residences would not be guaranteed. In cases like this,

Table 2. (a) Value of shipments by county and industry-example with data not suppressed (Adapted from http://factfinder.census.gov)

	Industry				
County	**W**	**X**	**Y**	**Z**	**Total**
Alpha	15	1*	3*	1*	20
Beta	20	10	10	15	55
Gamma	3*	10	10	2*	25
Delta	12	14	7	2*	35
Total	**50**	**35**	**30**	**20**	**135**

*Note: * indicates cells in which data may be identifiable due to the low number in the cell*

Table 2. (b) Value of shipments by county and industry-example with data suppressed (Adapted from http://factfinder.census.gov)

	Industry				
County	**W**	**X**	**Y**	**Z**	**Total**
Alpha	15	D	D	D	20
Beta	20	10	10	15	55
Gamma	D	D	10	D	25
Delta	D	14	D	D	35
Total	**50**	**35**	**30**	**20**	**135**

one solution would be to display each of these four residence locations at their closest intersection. By doing so, the confidentiality of the individuals would be protected.

The U.S. Department of Justice has not put forward any rules or guidelines for how confidential crime data should be displayed on maps. Individual police departments and law enforcement agencies apply different rules to preserve an individual's privacy. Such rules may range from aggregating data by police beat to displace confidential crime locations to the midpoint of the street segment or the closest street intersection. Accordingly, the two local masking methods identified above would be appropriate methods for visualizing the location of confidential crime data on a map.

Manipulating Both Area Boundaries and the Location of Confidential Point Data

The discussion in the previous section focused on methods to geographically mask the location of confidential point data, while leaving the base map information (street network and census tract boundaries) unchanged. This section introduces a method that manipulates base map information (census blocks, census tracts, zip codes, etc.) to preserve the privacy of an individual's residence. The idea behind this approach is to manipulate area boundaries through the creation of Thiessen polygons. A further manipulation of the mapped surface by performing an image modification (such as flipping about an axis) in a drawing package raises the preservation of patient confidentiality to an even higher level. This technique is relatively simple and easy to perform, and has the advantage of maintaining correct associations between different spatial data layers, while preserving patient confidentiality.

An illustration of point locations, such as infant deaths, needs no manipulation if no other reference layer is provided on the map. Indeed, such surfaces have already been presented in this book. For example, the maps shown in Figure 1 could not be scanned and georegistered in a GIS, as there are no spatial identifiers. Such a map would be limited to displaying patterns of deaths, whether they are clustered or not. It would not be possible to combine these clusters with other unmodified spatial data layers, such as socioeconomic information, by census tract, as the boundaries could be recognized, scanned, and georegistered. The following manipulation of spatial data allows for point and area data to be combined, preserving their correct interrelationships (such as most deaths are found in blocks with a high proportion of African Americans), but not containing any recognizable spatial features. East Baton Rouge Parish census tracts will be used to illustrate this approach.

Centroids are created for all census tracts. A centroid is the midpoint of an area (polygon), also known as the center of mass. It can be calculated with any common GIS software. A point file is generated that contains the original census tract ID number. A Thiessen polygon coverage is created from this point file. Again, this is a common procedure in any generic GIS. Thiessen polygons create "regions of influence" (DeMers, 1997) so that all map space is assigned to its closest centroid. In other words, if the map were comprised of a fine resolution grid, each grid cell would be assigned to its closest centroid. The resulting map would contain areas of grid cells assigned to the same centroid. The resultant coverage, when completed without the option of a known boundary such as county or parish line, preserves the spatial relationships of the original census tracts but with different generated boundary shapes. By completing a simple GIS join on the census tract ID number contained in both the census tract and Thiessen polygon coverage, all socioeconomic information contained in the census tracts are placed inside the corresponding Thiessen polygons, allowing for the identical mapping procedures to be performed. Figure 5 displays proportions of African Americans by census tract (left) and for the equivalent Thiessen polygon surface (right) for East Baton Rouge Parish. Onto these maps are overlaid randomly generated surfaces of "infant death" locations, and output significance contours from a spatial filter analysis showing clusters of infant deaths. The infant death locations have been randomly generated so as to not compromise the privacy of the individual(s) living at that location.

Two potential "errors" can occur during the process. The first is that irregular-shaped census tracts may have centroids calculated by common procedures that fall outside their boundary (for example, most centroid routines would place a point outside the shape of a crescent). This can easily be corrected by spatially editing the point back into the shape. Secondly, actual point locations (infant death locations) can fall into a different Thiessen polygon if they were originally

Figure 5. Census tracts (left) and equivalent Thiessen polygons (right) for the East Baton Rouge Parish (see text for explanations of what is displayed in both maps)

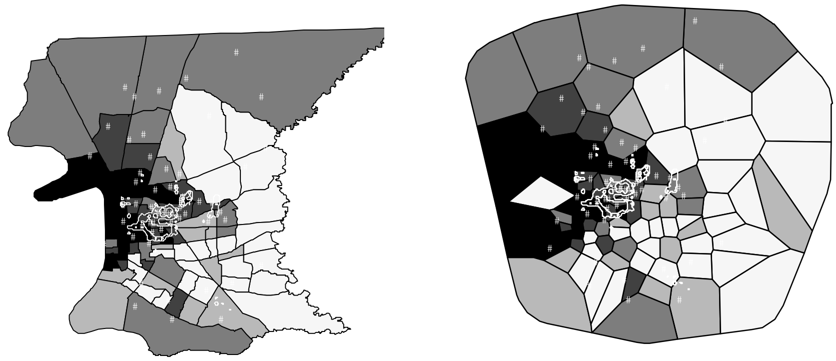
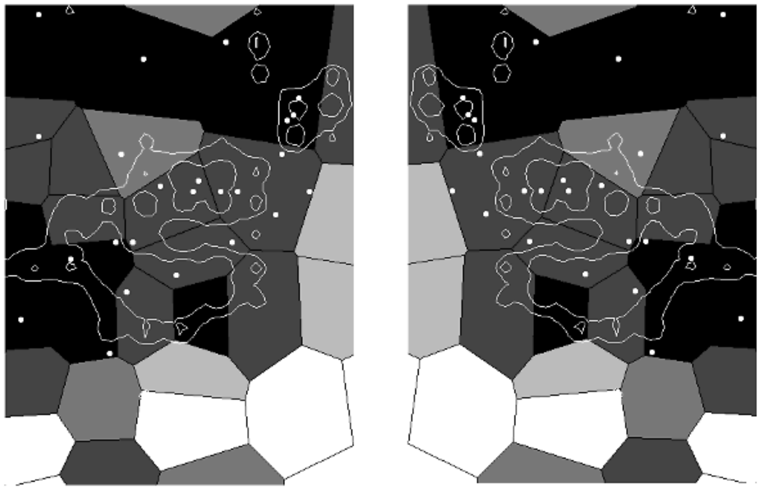


Figure 6. Original Thiessen polygons (left) and Thiessen polygons after being flipped horizontally (right)



located close to a border between two census tracts. Again, a relatively simple spatial edit could move the point into the correct polygon.

Although no physical reference remains with this manipulation, especially if displaying a magnified city neighborhood, it still might be possible (if unlikely) for someone to take the map and replicate the manipulation in order to again potentially identify actual point locations. A final manipulation virtually eliminates this possibility. The final image is exported into a drawing package, such as Adobe Photoshop, where it is rotated, or mirrored, or stretched, in a number of different ways. Figure 6 displays a magnified city neighborhood from the East Baton Rouge Parish with the same spatial data layers as displayed in Figure 5 before (left) and after (right) the image has been exported to Adobe Photoshop and flipped horizontally. The final graphic maintains the correct relationships between the different spatial layers, provides good supporting visual evidence to the text of the article, and maintains patient confidentiality, especially if the writer does not report on exactly how the image has been manipulated.

The reader may think that the manipulations of data presented in this chapter are too technically involved and probably more than a community health unit would need, especially if spatial data is never reported to an outside audience. The reality is that even small health units, especially if receiving federal funding, are now expected to disseminate their findings through articles and presentations. It is therefore important, especially with little in the way of a cartographic guideline to preserve confidentiality, to at least be aware of the problem before a mistake is made.

References

Acevedo-Garcia, D. (2001). Zip code level risk factors for tuberculosis: Neighborhood environment and residential segregation in New Jersey, 1985-1992. *American Journal of Public Health, 91*, 734-741.

Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. In P. A. Longley, M .F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical information systems* (Vol. 2, pp. 251-264). New York: John Wiley & Sons.

Armstrong, M. P., Rushton, G., & Zimmerman, D. L. (1999). Geographically masking health data to preserve confidentiality. *Statistics in Medicine, 18*(5), 497-525.

Bishop, C. W., & Bishop, G. C. (2000). Cascaded surrogates can protect confidentiality in outcome studies [Suppl.]. *Journal of the American Medical Informatics Association, 970*.

Carman, D., & Britten, N. (1995). Confidentiality of medical records — The patients' perspective. *British Journal of General Practice, 45*, 485-488.

Cohen, S. M. (2000). Patient confidentiality. *American Journal of Nursing, 100(*9), 24HH.

DeMers, M. N. (1997). *Fundamentals of geographic information systems*. New York: John Wiley & Sons.

Fotheringham, A. S., & Zhan, F. A. (1996). Comparison of three exploratory methods for cluster detection in spatial point patterns. *Geographical Analysis, 28*, 200-218.

Fowler, D. L., Hogle, N. J., Martini, F., & Roh, M. S. (2000). The use of a personal digital assistant for wireless entry of data into a database via the Internet. *Surgical Endoscopy-ultrasound and Interventional Techniques, 16*, 221-223.

Gatrell, A. C., Bailey, T. C., Diggle, P. J., & Rowlingson, B. S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers, 21*, 256-274.

Gatrell, A. C., & Senior, M. (1999). Health and health care applications. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical information systems* (Vol. 2, pp. 925-938). New York: John Wiley & Sons.

Leitner, M., & Curtis, A., (2003). Cartographic guidelines for the visualization of confidential data using geographic masking. Paper presented at the *21st International Cartographic Conference*, Durban, South Africa (pp. 101-110).

Leitner, M., & Curtis, A. (2004). Cartographic guidelines for geographically masking the location of confidential point data. *Cartographic Perspectives, 49*, 22-39.

Monmonier, M., (2002). *Spying with maps: Surveillance technologies and the future of privacy*. Chicago: University of Chicago Press.

Regan, M. (2001). Utilizing and integrating the Internet in disease management. *Disease Management and Health Outcomes, 9*, 131-139.

Schwartzbaum, J. A., & Hirschberg, J. G. (1991). Exploring the odds ratios using the bootstrap. *Statistics in Medicine, 10*, 1273-1281.

Wartell, J., & McEwen, J. T. (2001). *Privacy in the Information Age: A guide for sharing crime maps and spatial data*. Washington, DC: U.S. Department of Justice, Office of Justice Programs.

Woodward, B. (1996). Disclosure and use of personal health information — Widespread access is likely to erode patient confidentiality. *British Medical Journal, 312*, 653-654.

Chapter IX

Creating the Baton Rouge Healthy Start GIS

I recently gave a presentation to an Introduction to Gender and Minority Studies class at Louisiana State University. This was an interesting experience, as most of my talks tend to lean on the technical aspects (meaning GIS) of how to solve the problem of racial disparity in birth outcomes in the city. This audience was nontechnical and more interested in the injustices facing the poor (and more often than not African Americans) in the city. I would not have been comfortable giving this talk a couple of years previously, but 4 years of being attached to the Baton Rouge Healthy Start, writing continual updates and reports, applying for new grants, and attending all the associated meetings (e.g., evaluation meetings, project area committee meetings, Fetal and Infant Mortality Review (FIMR) meetings) have given me a reasonable insight into the causes of the problems as well as the distribution of dots on a map.

One point I discussed caused a certain degree of consternation, as it usually does, but with an unexpected and satisfying outcome. I outlined the commonly considered risks associated with pregnancy, one of which is receiving the first prenatal visit in the second trimester. I explained that this “risk” was partly due to the way the local medical system treated women in poverty. If a woman realizes she is pregnant, and many clinics will want a woman to wait 2 weeks after her missed menstruation for pregnancy verification, probably 6 weeks have passed. At this point she can apply for presumptive eligibility for Medicaid, and she will receive her temporary card 2 weeks later. Unfortunately, the full Medicaid card will not arrive for 4 weeks after application. Many doctors and providers in Baton Rouge (at time of writing) will not schedule a prenatal visit until this card has arrived. It often takes 4 weeks of waiting time until a free appointment can be found. Therefore, 6 + 4 + 4 = second trimester. Solutions to this problem are being sought, including the electronic automation of Medicaid distribution which, it is hoped, will reduce 4 weeks to a matter of days. *Hopefully this situation will no longer exist by the time this book rolls off the press.*

What made this lecture interesting was when an African American student talked to me afterwards and explained I had just described her situation, and she had as of yet not received a prenatal visit. I gave her the contact information for the Baton Rouge Healthy Start and told her to go and introduce herself. I also asked about her transport situation (which is a huge deterrent for many women). Maybe I am making more of this than I should, but academics, and especially geography academics, are not used to intersecting so dramatically with the real world. It was a nice feeling.

This is an argument I have championed for a long time. At the 2001 meeting of the Association of American Geographers I presented the beginning stages of the Baton Rouge Healthy Start GIS in a session, organized by Dr. Charles Croner from the Centers for Disease Control and Prevention (CDC). This session included several other impressive names from the world of GIS, spatial analysis, and health: Gerrard Rushton, Susan and Grant Thrall, Art Getis, and Peter Rogerson. A major point of my presentation was that the GIS analysis of health data had to have tangible outcomes, the results of which could be applied in the field.

As mentioned in Chapter V, there has been a long-standing debate in academic geography about the sanctity of theoretical research as compared to applied geography, as well as whether GIS is a science or not (Pickles, 1997; Wright, Goodchild, & Proctor, 1997). Again, this book is not the forum to address the issue. We will simply except that geography and GIS have been an integral part of the creation and continuation of the Baton Rouge Healthy Start.

Beginnings

I was in my first year at LSU (early 2000) when a world-renowned epidemiologist asked me to join him at a CityMatCH meeting. CityMatCH is an organization of city and county Maternal Child Health (MCH) programs, the mission of which is to "improve the health and well-being of urban women, children and families by strengthening the public health organizations and leaders in their communities" (CityMatCH, 2004).

At this first meeting the group told me about a presentation they had recently seen on one of the CDC's monthly broadcasts that had them excited. The presenter had been Dr. Rushton and his talk had focused on infant mortality hot spot identification. I informed them that I was familiar with the technique and explained a little about how a GIS worked. The epidemiologist was also familiar with GIS, though his expertise was in the animal realm (Dr. Martin Hugh-Jones was the director of the World Health Organization's Collaborating Center for Remote Sensing and GIS for Public Health before I became director. I will always be indebted to him as we continue to work on several exciting projects together. His favorite saying to the women of CityMatCH and Healthy Start was "I eat my patients").

At this point it was decided that we, as a community, would apply for a Healthy Start grant.

The mission statement of Healthy Start was:

…to promote the development of community-based maternal and child health programs, particularly those addressing the issues of infant mortality, low birthweight and racial disparities in perinatal outcomes. As part of its mission, the NHSA supports the expansion of a wide range of activities and efforts that are rooted in the community and actively involve community residents in their design and implementation. (Baton Rouge Healthy Start Project, 2001)

The Healthy Start program really began in 1991 when Health Resources and Services Administration (HRSA) of the Department of Health and Human Services (DHHS) funded 15 urban and rural locations with an infant mortality rate ranging between 1.5 to 2.5 times above the national average. An additional seven sites were added in 1994, 75 more in 1997, 19 in 1999, three in 2000, and 12 in 2001. At the time of writing there are 96 funded projects in 37 states, the District of Columbia, the U.S. Virgin Islands, and Puerto Rico.

Healthy Start programs are community driven and located in the poorest and most needy neighborhoods in the United States. Since its initiation in 1991, the Healthy Start program, has served hundreds of thousands of families. Over 90% of all Healthy Start families are African American, Hispanic or Native American. Healthy Start specializes in outreach and home visiting as the surest ways to reach the most at-risk women. (National Healthy Start Association Inc., 2004, p. 3)

There are six specific grant types that can be applied for; the Baton Rouge group decided to target the Eliminating Disparities in Perinatal Health program, building on a previous application which had been unsuccessful. The three major grant writers would be Dena Morrison from Family Road, Jamie Roques from the Office of Public Health, and myself representing the Department of Geography and Anthropology at LSU. It was decided that this application would include four GIS components: the result of a spatial filter analysis identifying our program area, some basic cartographic displays of the program area (Figures 1 to 3), the role GIS and spatial analysis would play in determining baseline data for objectives, and the creation of a central database/GIS for program evaluation. The following extracts from the original proposal give an insight into the geography of the program area and the role GIS was described as playing (Baton Rouge Healthy Start Project, 2001). I must say, going through this document again made me realize how much work Dena Morrison put into the writing of the original grant, and I certainly do not want to present the picture that the GIS sections contribute a third of the total document, as they do not.

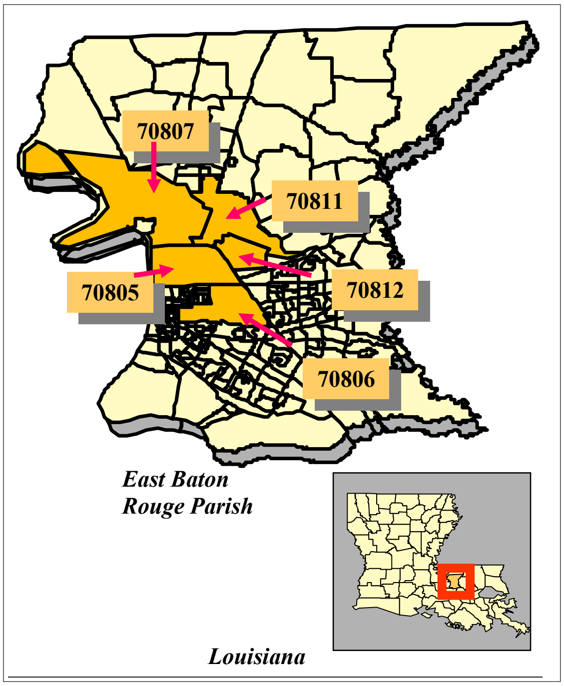
The first part of the narrative was to identify the service region. Although not mentioned in the text, this geographic selection was based on the results of a spatial filter analysis of birth and death certificate data.

...selected the 39-square-mile area in the central part of East Baton Rouge Parish as its project site. The project area embraces five zip codes, 70805, 70806, 70807, 70811, and 70812 [see Figure 1]*. This area has one of the highest concentrations of infant deaths in the Parish. Nearly half (49.4%) of the African American women of reproductive age living in East Baton Rouge Parish reside in the project area. Similarly, during the years 1996-98, births to African American women in the area accounted for 51% of all African American births in the Parish and 52% of the African American infant deaths (based upon a three year average). Consequently, by targeting this population, the Healthy Start Project will be implementing measures that can significantly reduce the current disparities in perinatal health between the Caucasian and African American populations in East Baton Rouge Parish.* (Baton Rouge Healthy Start Project, 2001)

Once the geographic area had been identified, the next stage was to provide demographic information; in other words. the population to be served. The sources for these data varied, though for our current renewal analysis www.geographynetwork.com, www.atlas.lsu.edu, and www.census.gov have been considerably utilized.

All of the following data were extracted by a combination of spatial (by zip codes) and aspatial queries (by risk) in the GIS:

Figure 1. The five zip code study regions



The population in the project area includes 22,803 women of reproductive age, of which 79% are African American. Between 1996-98 there were 5,475 live births, of which 4,634 or 85% were born to African American women. By contrast, 93% of the infant deaths in the project area, 84 of 90 infant deaths, were African American. Eighteen percent (859) of birth mothers were under 19 years old. Fifty-four percent of all babies in the project area were born to mothers who did not complete the eleventh grade.

The African American infant mortality rate of 18 per 1000 live births in the project area is higher than the parish-wide rate for African American infants of 17. This rate is double the 7.5 infant mortality rate (IMR) for Caucasians in the project area and greater than three times the Caucasian IMR of 5 for East Baton Rouge Parish.

Among African Americans, 57 of 84, or 68%, of the infant deaths occurred within 28 days of birth. Among Caucasians, three of the six infant deaths occurred within the first month after birth. The high concentration of African American infant deaths and the racial disparities in the infant and neonatal mortality rates indicate that the proposed Baton Rouge Healthy Start project will target a population of women and children most in need of the proposed services.

The data on premature births, initiation of prenatal care, and causes of infant death also highlight the racial disparity in perinatal health. Over one *quarter (26%) of babies born to African American women in the project area were premature. The 1,191 African American premature births in the project area between 1996-98 compare unfavorably with the 117 (15%) Caucasian premature births in this area. The percentage of African American premature births in the project area is double the 13% of Caucasian premature births for East Baton Rouge Parish.*

The data on low birth weight births in the project area reveals a similar pattern. Fifteen percent or 232 African American babies weighed 2500 grams of less at birth, of whom over a quarter (26%) weigh 1,500 grams or less (based upon the 1996-98 three-year average). There were 61 babies (4%) of African American live births in this very low birth weight category. By contrast, 7.4% of Caucasian babies weighed 2500 grams or less. Only 1% of the Caucasian babies weighed 1500 grams or less.

One-third of the birth mothers (510) in the project area did not enter prenatal care during their first trimester of pregnancy. Thirty-six births (2.3%) were to African American women who had no prenatal care. This figure is considerably higher for African Americans than Caucasians. In the entire Parish of East Baton Rouge, only 31 of 8,664 birth mothers (0.4%) failed to initiate prenatal care during their first trimester of pregnancy. (Baton Rouge Healthy Start Project, 2001)

The following data were extracted from the international classification disease (ICD) code on the death certificate:

The high percentage of women in the project area receiving inadequate prenatal care may account for the high infant mortality and morbidity rates. A review of the causes of infant deaths between 1996 and 1998 reveals that at least 65% of the causes were attributable to perinatal conditions that can be treated to reduce the probability of infant death or avoided through appropriate prenatal care. These causes include short gestation and low birth weight, respiratory distress, maternal complications of pregnancy, birth trauma, intrauterine hypoxia and birth asphyxia, respiratory conditions of fetus/newborn, infections specific to perinatal period, fetal/neonatal hemorrhage, perinatal disorders of the digestive system, fetus or newborn affected by complications of placenta, cord, and membranes, conditions involving the integument and temperature regulation of fetus and newborn, and other/ill-defined conditions originating in the perinatal period. In addition, sudden infant death syndrome (SIDS) accounted for 5% of the deaths, while 20% were due to congenital anomalies. An additional 10% were due to accidents, homicide, and unspecified causes (8%). Fortunately, accidents have not been a significant cause of deaths of children in the *project area. In 1998, the major cause of accidental death for children under one year of age was suffocation. For children one to four years old, the major causes were drowning or submersion and fire or burns.*

Poverty in the project area creates additional needs in the target population. The percentage of African American children under 18 years old living in families with incomes below the Federal Poverty Level in the project area is 44.3%, almost three times greater than the 17.8% for Caucasian children. (Baton Rouge Healthy Start Project, 2001)

The following data are combinations of multiple data sources, some from the health department, some census, some birth and death certificate, anything that would give the reviewer an accurate indication of the problems faced in the program area:

Other socioeconomic data on birth mothers in the project area both help to explain the levels of child poverty and highlight the special needs of the project area residents. The African American teen birth rate, when nineteen-year-old mothers are added, rises to 26%. Approximately 65% of project area birth mothers are unwed. Ensuring that women in the project area receive early and continuous prenatal care is also hampered by the high percentage of birth mothers who did not complete high school (approximately 54%) as well as the limited availability of quality employment opportunities.

Other data from the project area reflect the consequences of these demographics. In 1999, there were 240 valid cases of child abuse and neglect located in the project area, which accounted for 25% of the 970 cases in the Parish. The area also had more than its "share" of reported domestic violence cases. In 2000, 65% of the new cases were located in the project area. Of the 990 cases, 604 or 61% were African American women and children; 346 were Caucasian. (Baton Rouge Healthy Start Project, 2001)

Although the following section comments on a general lack of domestic abuse data, this has been rectified in the renewal proposal due to these data being collected in the Baton Rouge Healthy Start.

Single parenthood and domestic violence contribute to homelessness. In 1999 East Baton Rouge Parish had 364 children under five years old and 550 children ages five to seventeen classified as homeless. Data reported for a single day noted that of the 332 people receiving service, 54 sought assistance due to domestic violence. (LA Interagency Action Council on the

Homeless, Needs Assessment Survey 1999). Specific data for the project area is not available, but anecdotal evidence from service providers in the area indicates that some pregnant women in the area confront problems of homelessness, particularly those who are victims of domestic violence or who have mental health or substance abuse problems. (Baton Rouge Healthy Start Project, 2001)

Crime can be an important neighborhood issue, as was emphasized in Chapter IV — especially in terms of the stress it causes (fear of leaving the house), the infrastructure it denies (limited clinic opening times), and the negative health outcomes it can lead to (e.g., asthma and lead exposure).

Not surprisingly, project area residents also confront problems of crime. Public safety in the area has been generally summarized by the APB Neighborhood Crime Check ratings developed by CAP Index, a leading provider of crime risk assessment data to corporate America. The ratings, which represent a resident's risk of encountering violent crime, assign areas a number from one to ten, with a "1" meaning the risk is one-fifth the national average and ten indicating the risk is 10 times or more the national average. Four of the five zip codes comprising the project area received ratings between "7" and "9." Only zip code 70811 received a "6" deeming it a borderline moderate risk. (Baton Rouge Healthy Start Project, 2001)

HIV/AIDS is a huge problem in Baton Rouge. Unfortunately, no data is available at the residential level to allow for GIS analysis. There are several current program participants who have been diagnosed with the virus, though numbers are too small for any meaningful neighborhood analysis.

One of the unique risk factors to women in the project area is HIV/AIDS. Given the prevalence of the disease in Baton Rouge and the high number of unmarried, sexually active women, both prevention and treatment services are needed. The Baton Rouge Metropolitan Area has the 12th highest AIDS case rate among metropolitan areas with populations greater than 500,000. (1999 Control and Prevention HIV/AIDS Surveillance Report, Centers for Disease Control 1999 edition, Volume 11, No. 2) Baton Rouge has a new AIDS case rate of 32.6 per 100,000 and a total case rate of 53 per 100,000, which is higher than New Orleans or Los Angeles.

In 1999, 135 of the 246, or 55%, of new cases of HIV/AIDS in East Baton Rouge Parish were in the project area. Of these, 87% were African American, of which 37% were female. For 1996, 1997, and 1998, 17 *babies, 16 babies, and 18 babies, respectively, were born to HIV-positive mothers. In 1996, four of the babies were HIV-positive, while one baby in each of the subsequent years was HIV positive. The intensive HIV/AIDS case management and outreach program in East Baton Rouge Parish, which will be associated with this Healthy Start program, will be responsible for early medical intervention for pregnant women, thus reducing the likelihood of babies born HIV positive.*

In 1997, only 45% of African American infants in the project area under the age of two had received all age-appropriate immunizations, compared to 70% of the Caucasian children in the same area. In 1998, the figure for African American children was 43%. The comparable percentage among Caucasian children had also dropped to 66%. Percentage of all children two years and under in the project area who were up to date on their required immunizations in 1998 was 57%. (Baton Rouge Healthy Start Project, 2001)

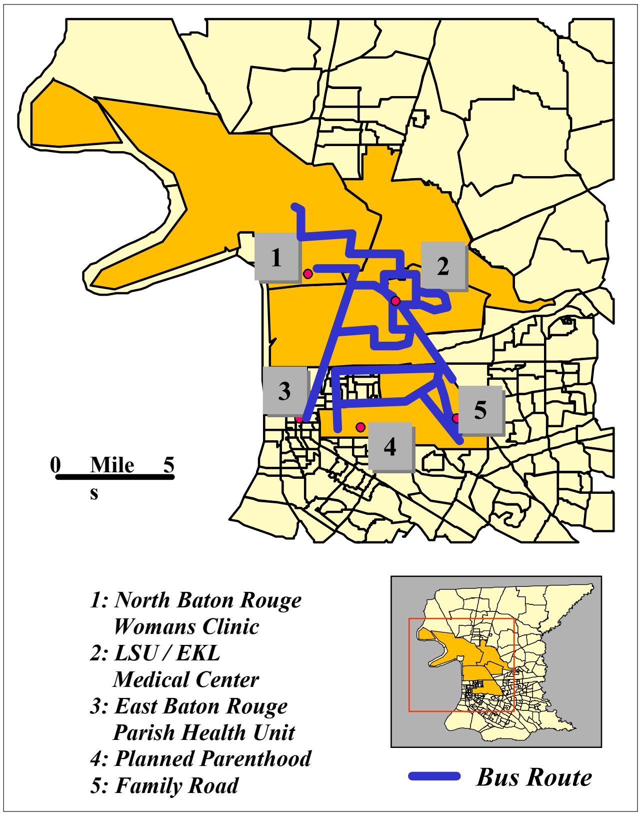
Once the program area and the population residing inside had been defined, the current perinatal health care delivery system was outlined using maps displaying general accessibility.

Women have access to six facilities in East Baton Rouge Parish for walk-in pregnancy testing at no cost or a nominal charge [see Figure 2]. *Of these six, four are within or convenient to the project area. If a test is positive, counselors at these sites provide initial counseling and referral for prenatal care. They also provide each woman with documentation of her pregnancy, which is a requirement for Medicaid eligibility determination.*

To determine Medicaid eligibility, women are referred to the East Baton Rouge Parish Health Unit, which is located within a convenient distance of the project area. A bus stops directly in front of this health unit, which is frequently visited by Medicaid clients. The Health Units capacity for this service is 35 applicants per week or 140 applicants per month. There is a need for increased access to Medicaid eligibility determination services. Title V MCH is presently negotiating with Family Road of Greater Baton Rouge, a multi-service center for women and infants located in the project area, to enable this site to provide this service.

At the time of eligibility determination, women are referred to one of seven facilities within the parish of East Baton Rouge for intake to complete the Medicaid or LACHIP [the Louisiana Children's Health Insurance Program] application. Six of those facilities are located in or in close proximity to the project area [see Figure 3]. *Women deemed ineligible for Medicaid benefits are referred to LSUMC [Louisiana State University Medical Center] /Earl*

Figure 2. Pregnancy testing facilities

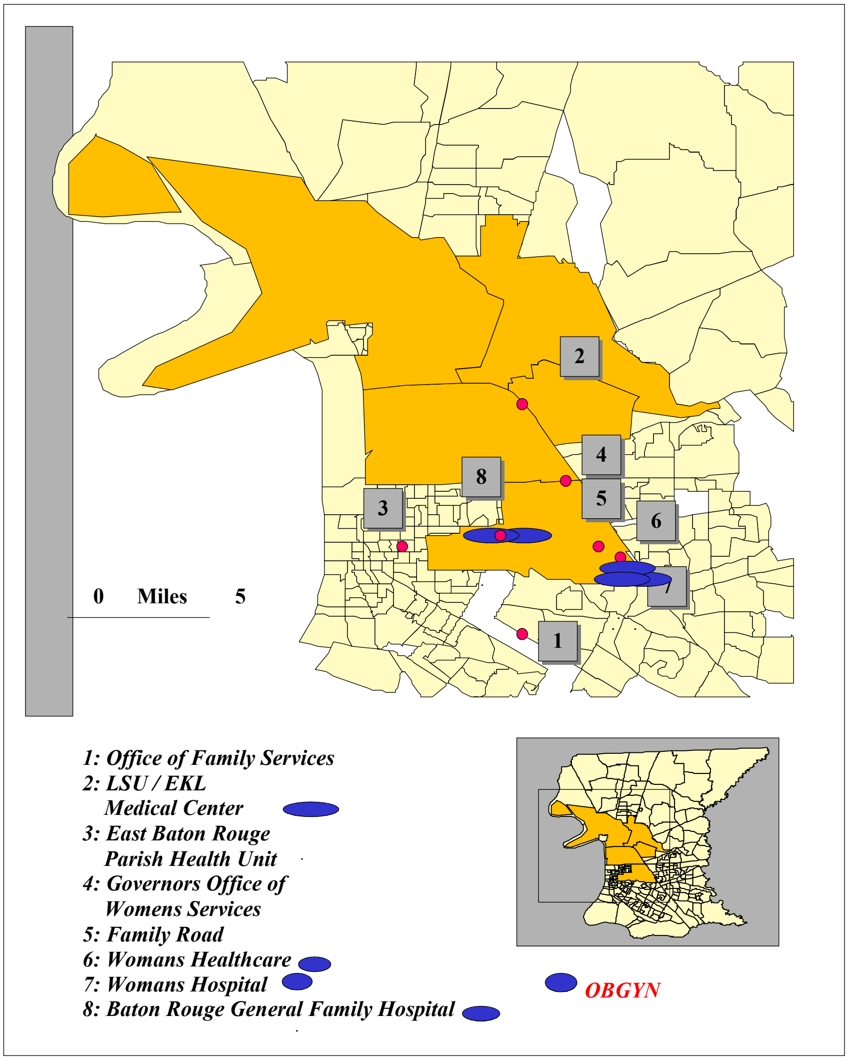


K. Long or to a provider of their choice if they choose to pay out-of-pocket for prenatal care.

Women Infant and Children (WIC) services are provided to project area residents at four sites. The Parish Health Unit discussed above is the largest provider of WIC services. Family Road of Greater Baton Rouge serves as a branch of the Health Unit in providing WIC services to prenatal clients. In January 2001, 391 women received WIC services at these sites. Two other providers in the project area are Child Health America, which currently has 158 prenatal WIC participants, LSUMC/Earl K. Long, which currently serves 234 prenatal WIC clients. (Baton Rouge Healthy Start Project, 2001)

Several sections of the project narrative discussed the use of GIS explicitly. The following statement details the eventual "no prenatal" information system described later in this chapter.

Figure 3. Medicaid of LACHIP Application Centers, OBGYN, and major perinatal health care



Recruitment of outreach teams' members and implementation of the program will be an incremental process. The Baton Rouge Consortium proposes initially to divide the project area into the five areas coextensive with the zip code boundaries. The boundaries will be set to insure that outreach workers are assigned to residential neighborhoods with the highest concentrations of women of child bearing age as determined by the GIS survey, census tract information, and residential information collected from birth certificates during the past five years. Outreach workers will seek to provide services to at least 42% (2,600) women in the target *population during the four years of the Project period.* (Baton Rouge Healthy Start Project, 2001)

It was in the program evaluation section where GIS use, and the spatial analysis of birth and death certificate data were brought to the fore. Although there were many examples, the following excerpts provide a flavor of what was written — in effect spatial analysis was used to identify the problem areas, and then compare results between the years by reapplying the analysis:

Previous Birth certificate data (1996-98) will be analyzed to develop a profile of women (both by individual characteristics and from the neighborhood where they come from) most likely not to seek care in the first trimester for two of the zip codes in the study area. This information will be provided to all Project personnel, outreach workers, Network Neighborhood Support Team members, and members in the "Partners for Healthy Babies". In this way, women who are contacted and a: exhibit identified risk factors and/or b: come from a high-risk neighborhood, will be flagged for special attention.

Birth certificate data will be continually analyzed during the project period. The evaluation team will document the total number of women receiving prenatal care during their first trimester over the duration of the project. The four-year trend will then be compared to the trend as identified in the objectives (continual drop from 67% to 80%).

The evaluation team will compile neighborhood profiles from 1996-98 birth certificate data to identify those areas that traditionally have high teen pregnancy rates.

By 12/31/2001, the evaluation team will map the distribution of children below the age of 2 who have not received their full schedule of age-appropriate immunizations. This map of under-coverage will be used as a guide by the Office of Public Health to target immunization programs in these areas.

By 6/30/2002, the evaluation team will map the distribution of children below the age of 2 who have not received their full schedule of age-appropriate immunizations. This map will be compared with the map from 12/31/2001 in order to judge effect of programs placed in those high-risk areas. If there has not been an increase from 57% to 69%, programs will be reevaluated. This process will again be repeated on 12/31/2002. (Baton Rouge Healthy Start Project, 2001)

It was during the overview of the evaluation component of the project narrative that the plan for the Baton Rouge Healthy Start GIS was (if somewhat briefly) outlined:

General Evaluations:

Evaluation of the above objectives will result from compiling a sophisticated Geographic Information System (GIS) for the area. A Database system connected to this GIS will be built on a secure web site. Any program officer, through a laptop or office computer, and via their own password, will enter data on any client as dealt with. For example, a new client entering a clinic will be entered onto the web site. Any subsequent visit will be added to this initial entry building a complete profile of the mother. Designated workers associated with the program will also be allowed to access the database for immediate information on any client. In this way immediate and informed client help can be offered — as well as more general analyses being performed.

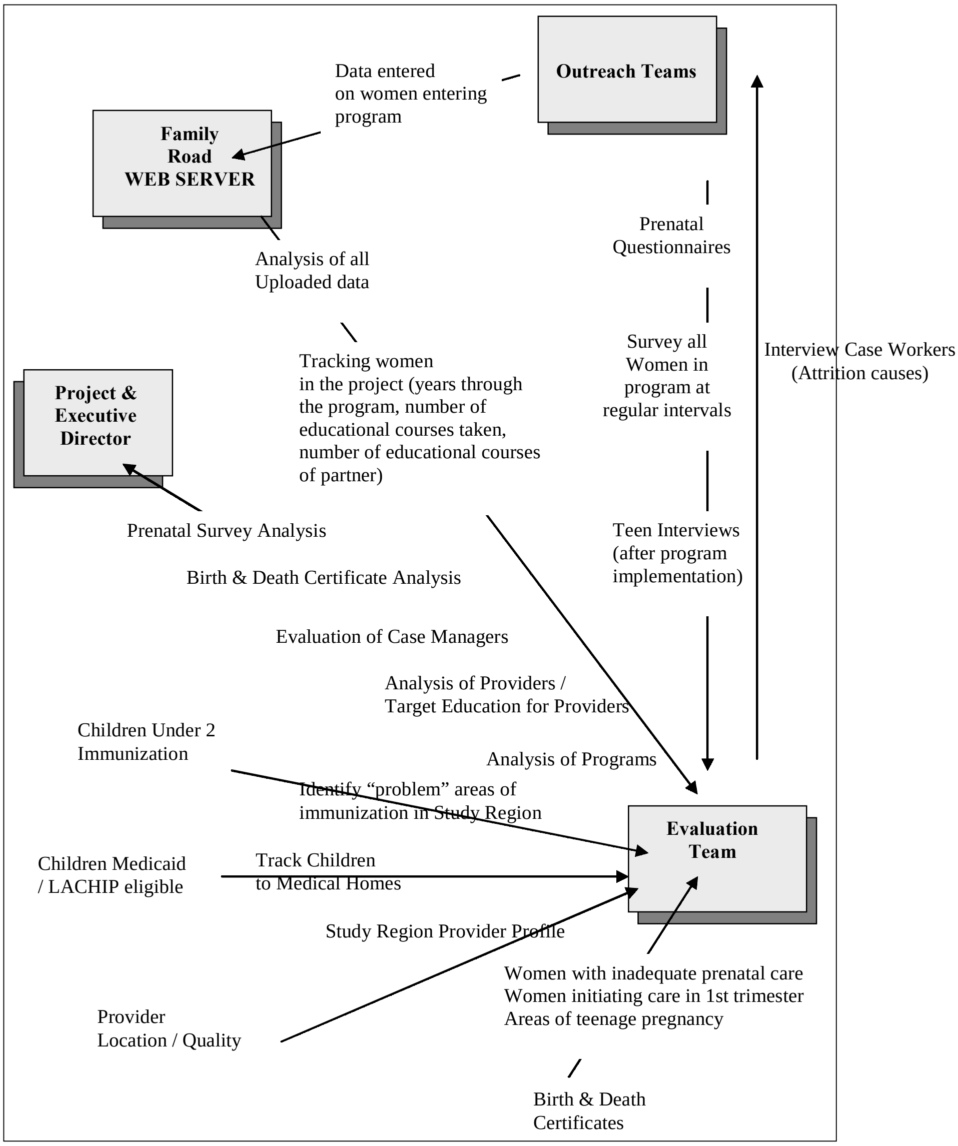
These data will be supplemented with Interview data and Birth and Death certificates, which will be located in a secure section of the database that only the Evaluation team can access. These three sources of data provide all the information necessary to evaluate any of the objectives [see Figure 4].

Other data that will be analyzed by the evaluating team from this database, which should be impacted by the program but has not been directly listed as an objective, include:

- *Facilitating services, including transportation, child care, and/or translation;*
- *Psychosocial services, including substance abuse treatment and counseling, HIV/AIDS screening/counseling and/or treatment, nutrition counseling, male support services, housing assistance referrals, prison/jail initiatives, and job training;*
- *Training, including consortium, consumer, staff and provider training;*
- *Spatial patterns in low birth weight and associated risk factors;*
- *Spatial patterns of mother medical risk factors (such as hypertension);*
- *Infant mortality.* (Baton Rouge Healthy Start Project, 2001)

Fortunately, the proposal was successful.

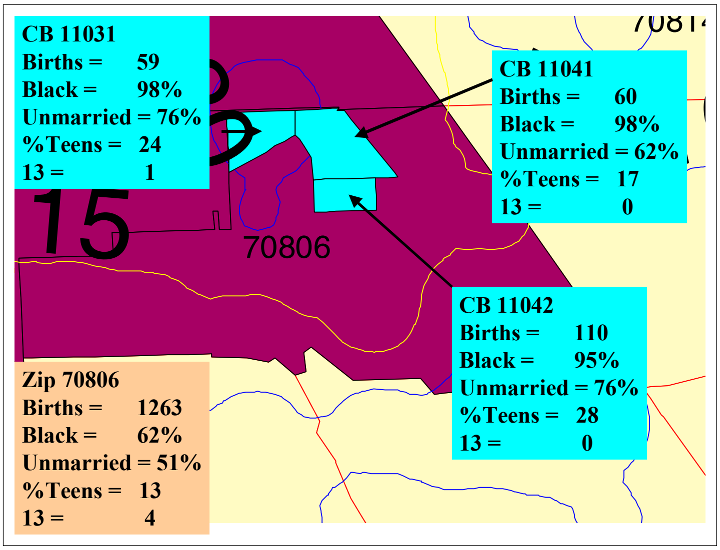
Figure 4. Data collection flow chart



Determining the Program Area

As has been mentioned previously in this book, a spatial filter analysis (using the DMAP program) was used to determine infant mortality hot spots for the years 1996-1998. In order to perform this analysis, several of the points already discussed in this book had to be navigated. For example, the initial application for

Figure 5. Changes in geographic scale



data was incredibly time-consuming. Once the data had been received, the arduous (if done correctly) task of geocoding was performed. The infant mortality surfaces were generated and the highest-risk zip codes were identified. At the time of writing the initial proposal, program areas were supposed to be comprised of zip codes, so our hot spots were translated into this level of aggregation. The renewal grant instructions now allow for more flexibility with portions of zip codes being allowed (these areas being defined by census tract boundaries). This will help when a problem area straddles two zip codes. As an example of how IMR can change within a zip code, consider Figure 5, which shows the variation in a few select birth measures at both the zip code level and for census block groups contained within. Notice how the proportion of African Americans, single mothers, and teenage mothers rises in these subset areas.

Identifying Areas With No Prenatal Care

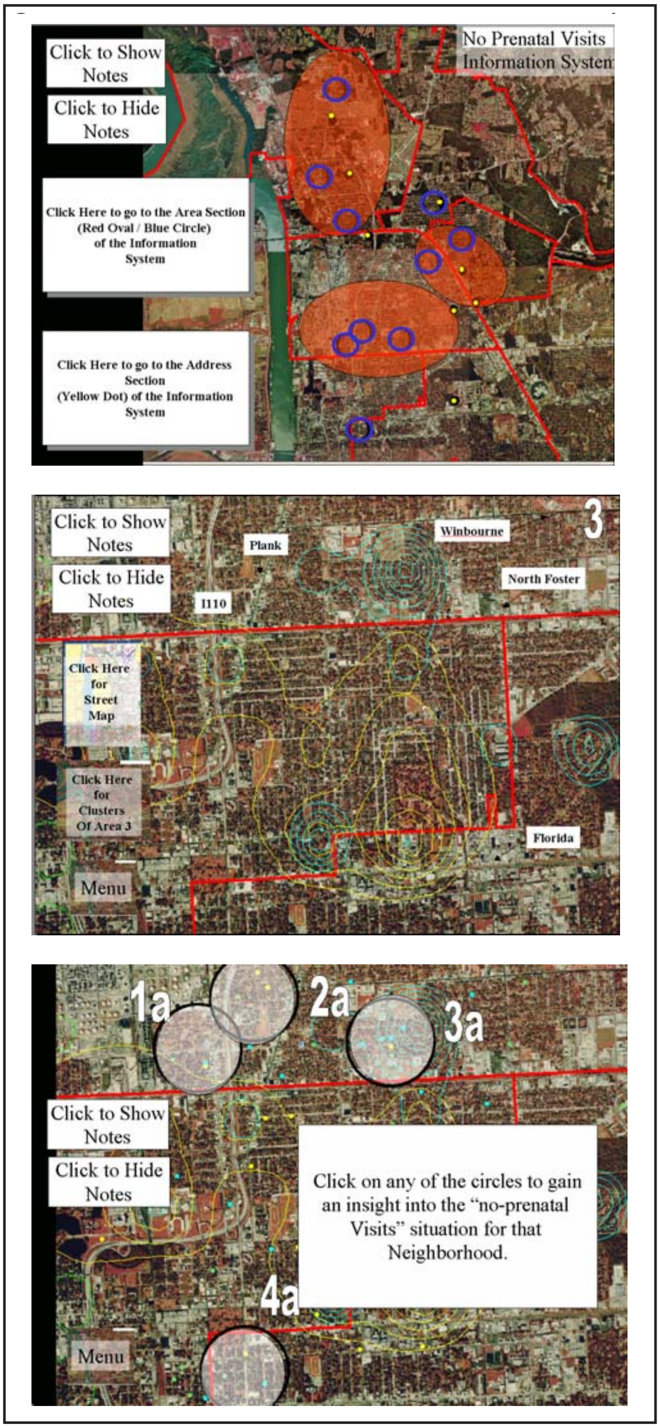
The spatial filter analysis was also applied to other fields captured in the birth certificate, such as birth weight, gestation, and prenatal care. At the 2003 MCH meeting held in New Orleans, a guest speaker commented that one of the greatest problems facing the outreach arm of programs such as the Baton Rouge Healthy Start was being able to identify women who received no prenatal care.

Our approach to this problem was to identify neighborhoods with temporally stable, high proportions of "no prenatal visits" as listed on the birth certificates. By identifying these neighborhoods, outreach workers, or paraprofessionals (local women who receive some enumeration and training from the program), could inform about the risks associated with pregnancy, and generally "keep an eye out" for potentially pregnant women.

The spatial filter was again used in a variant of the analysis described in Chapter VII. The question then was how best to translate the results of the analysis to outreach workers who had no familiarity with GIS and spatial analysis. The solution was to create a crude information system constructed in PowerPoint® that the caseworkers could have on their laptops. This information consisted of several hot buttons that allowed easy navigation between slides. Three example slides are shown in Figure 6. The opening slide of the information system displays an aerial photo of the program area. On this were three generalized ovals displaying areas of risk where, in previous years, high proportions of women receiving no prenatal care resided. As described in Chapter VII, these ovals were actually generalized hot spot contours from the spatial filter analysis. In fact, the spatial filter contours are visible in the second of the images. Within each of these ovals, circles indicated neighborhoods where particular problems were found (based on the actual number of women who previously had not received prenatal care). By clicking on any of these circles, a zoomed-in map was displayed of these "clusters." An alternative approach was to click on the larger oval, which zoomed the user into a map showing the problem neighborhoods for that area. For all three of these scales of display, a notes section allowed the caseworker to read an overview describing the risks found in that particular neighborhood. Each slide also had a navigation button that would take the caseworker back to any of the higher levels of the information system. One final addition on the slides was the option of switching from an aerial photo display to a Baton Rouge street map which also included driving directions. The information system contained one final level of investigation, dots that identified addresses (usually apartment buildings) where women had received no prenatal care in multiple years. These were thought to be good starting points for the display of Healthy Start posters. Again the caseworker could click on the address and get risk information and driving directions.

Although this information system was designed to be stand-alone, it was first presented to a group meeting of caseworkers. This meeting proved to be very insightful, as "problem neighborhoods" were immediately identified on the aerial photographs. This meeting also taught me the colloquial names for these areas. Although the caseworkers already knew about these problem areas, they liked how their own beliefs were corroborated by this new technology. My satisfaction went the other way, that the GIS clusters really did contain problems needing to be addressed.

Figure 6. The "no prenatal visit" information system



Neighborhood Profiling

A further use of the birth and death certificates was to create neighborhood profiles for any program participant entering the Healthy Start program. Similar profiles have been described in Chapters V and VII. The basic premise was to use a difference of proportions test to identify the most significant risks found for multiple years in the 0.25-mile neighborhood surrounding each program participant. Raw numbers could also be used; for example, how many low-birth-weight deliveries, how many infant deaths, etc. had occurred in that neighborhood over the last 5 years. It was hoped this additional information could help better prepare the caseworker. A more effective approach is planned whereby an information system will allow the caseworker to feed in an address on his or her own laptop and immediately see summary information. This system should prove invaluable when enough program participants have passed through the Baton Rouge Healthy Start, allowing for more detail to be used in the creation of the profile.

Creating the Database

In Chapters IV and V, examples were shown of the Baton Rouge Healthy Start database interface built by Farrell Jones, Associate Director of the CADGIS (College of Art and Design and Department of Geography and Anthropology) lab at Louisiana State University. The purpose of this database was to create a system that allowed Healthy Start to keep control of its data, and at the same time perform whatever GIS analysis was needed. Although other database systems are available (often at considerable cost), these are often inflexible in terms of what can be requested. These external database options provide excellent alternatives for the generation of end-of-year evaluation reports, but if the program wants to investigate a particular issue, the lack of flexibility can be frustrating.

From a GIS perspective, spatial analyses can be performed on any of the attributes collected about the program participants. Data could include neighborhood stressors, quality of housing, number of family members residing in the same unit, psychosocial stresses, domestic abuse, problems with diet, and so forth. All of these variables could be analyzed spatially for the program area. If program participants were considered to be samples of the neighborhoods in which they reside, then outreach intervention could also be planned if particular risks were identified.

This flexibility in data manipulation has also been important as MCH reporting requirements change from year to year. Anecdotal evidence suggests that other

Healthy Start programs sometimes face data crises when asked to report on a new variable from a previous year of their program. To extract these data from paper records is time-consuming and usually full of errors. A well-constructed database provides an excellent alternative.

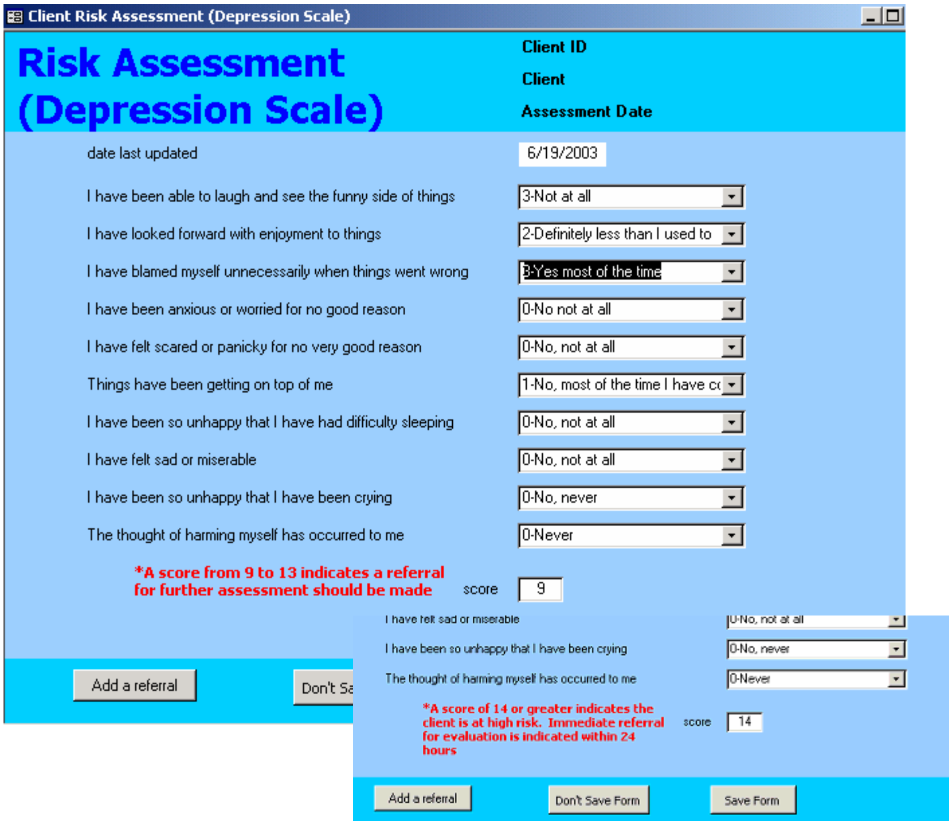
Developing the database was a slow process, and to this day still requires constant tweaking. A recent MCH push has been to address interconceptional care — this will again require a modification of our existing database. Even something simple like the changing of "client" to "program participant" requires data base maintenance. The initial "brainstorming" meetings involved myself, Farrell Jones, who was to develop the interface, the program director, and a selection of nurses and social workers. The purpose of these gatherings was to construct a system that captured all data required to write reports on the objectives we had listed in the initial proposal, along with any additional MCH required data. Other data was collected to satisfy program or research needs; for example, I requested a mechanism to track address changes through the duration of the participants attachment to the program.

The interface itself was designed with caseworkers in mind, the data forms being easy to follow and complete. Date stamps were automated, meaning when a first contact was made, the initial date was recorded. Each program participant had a unique identifier, as did every baby born. Many data pages comprised of "click" boxes, some of which opened a further data collection window. As many validation checks were placed in the system as possible; for example, if the date meant a child was born before the initial contact, or that the mother was over 100 years old, the entry was flagged. Other automation included flagging datasheets when entries were missed, posting reminders when the program participant was expected to make a visit, and prompting the caseworker to make a referral when a threshold was reached (Figure 7). Examples of different data sheets can be found scattered throughout Chapters IV and V. Figure 8 provides three further examples including infant immunization, driving directions to the program participant's residence and a follow-up contact sheet.

Of course, there were also problems, two of the trickiest being program participants that enter the program, leave the program, and join again later. Also, clients can become prenatal, then postpartum, then interconception, and again become prenatal. There were also problems with caseworkers incorrectly filling in information. We still have to go back through cumulative reports and check why some program participants have information missing, or why two dates do not make sense, and though each round sees more validation and quality control checks entered into the system, it is likely this system will never be perfect.

What should be remembered is that the construction of the database is a time-consuming process, and is expensive unless a computer programmer is found who also believes in the mission of the program. The construction of the database is not the end of the story, though. As has been mentioned, new data entries will

Figure 7. A data sheet with referral



need to be programmed, reports will need to be generated, and mistakes and crashes will have to be salvaged.

From a GIS perspective, these databases can easily be imported by geocoding the initial address on the client contact form. All other data sheets can be joined using the unique program participant identifier. In this way, maps can be made based on queries for any single record, or combinations of multiple records. Examples can again be found in Chapter IV.

Data Input

Data input for the Baton Rouge Healthy Start was originally supposed to be via laptop and secure wireless uploads. The current system is still laptop-based, but requires monthly data dumps. As with all technology, things change rapidly. It was soon found that laptops were too intrusive and sometimes created distrust between program participant and caseworker. Personal digital assistants (PDAs)

Figure 8. Example data sheets

provide a possible alternative, but their limited viewing screen may prove too problematic. A possible alternative is moving to a computer tablet, though this could also be viewed as being too intrusive.

On the issue of the Internet, it is hoped that eventually all data will be uploaded to a secure server via a wireless connection. This is obviously the way all data collection is moving for all agencies. Just as with the initial database construction, the development and maintenance costs of these Web portals will not be cheap. This cost impediment has so far hindered its development in Baton Rouge.

Reaching Out

The success of using GIS in the Baton Rouge Healthy Start has not been lost on other community health groups. Other Healthy Starts around the country have requested additional information about how to implement a GIS, usually after seeing a presentation at an MCH meeting. The same interest has been generated at the local level, with other health groups, such as the YWCA, asking for help in identifying patterns of alcohol use during pregnancy, breast cancer screening deficiency, and identifying teenage pregnancy risks around Early Head Start Centers. Most groups, once they have seen what a GIS can do, realize the potential of the technology.

What Next?

So what else can GIS offer? For starters, it can offer an alternative to standard evaluation techniques. The program reporting form required by MCH does not really detail the success or problems of a Healthy Start. One improvement on this reporting tool is to show how program participants fare against other births in their neighborhoods. Figure 9 shows how any program participant birth can be compared to all births within a 0.1-mile radius of the home residence. Although a single comparison such as this might not be too insightful, if all the deliveries were evaluated in this way, caseworkers could be judged as to their general effectiveness. Of course, more than this single evaluation mechanism should be used, as one caseworker may have a caseload in a particularly “difficult” area of town. This evaluation approach will be even more useful when the program has generated several years of data so that current births are compared to previous program births in the same neighborhood. This has the additional advantage of working with a similar population, and being a litmus test to see if the program was improving.

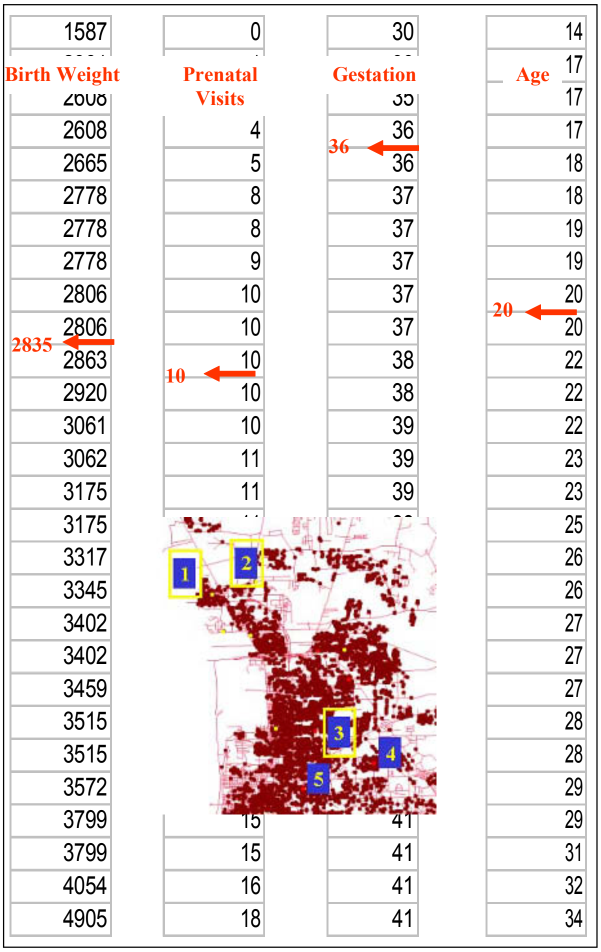
Post Script

In December of 2004 a further 4 year renewal grant was written for the Baton Rouge Healthy Start. Although funding levels cannot increase, a further three zip codes have been added to the program area.

References

Baton Rouge Healthy Start Project. (2001). *Eliminating disparities in perinatal health*. Application for funding under Healthy Start Initiative, submitted to HRSA Grants Application Center. Baton Rouge.

Figure 9. Using GIS as an evaluation



CityMatCH. (2004). *General information*. Retrieved November 14, 2004, from http://www.citymatch.org

National Healthy Start Association Inc. (2004). *The Healthy Start guide to risk factor assessment and how to communicate about perinatal risk to local communities* (Field test copy). Baltimore.

Pickles, J. (1997). Tool or science? GIS, technoscience, and the theoretical turn. *Annals of the Association of American Geographers, 87*(2), 363-372.

Wright, D., Goodchild, M., & Proctor, J. (1997). GIS: Tool or science? Demystifying the persistent ambiguity of GIS as "tool" versus "science." *Annals Association of American Geographers, 87*(2), 346-362.

Chapter X

Bioterrorism, Pregnancy, and Old White Men

By the title you might think this chapter is a departure from the general theme of this book. I beg your indulgence as I plead my case, because pregnant women as a cohort have been tragically forgotten in the hazards and disaster literature. Although mapping of vulnerable populations is a recognized approach, I argue (and frequently have) that we need to reconsider our disaster mitigation, response, and recovery plans, directing attention toward neighborhoods with high proportions of high-risk pregnant mothers. I will argue my case over the course of this chapter and at the same time suggest a potential funding strategy that can tap into Homeland Security funding. This chapter will focus on one particular disaster, a bioterrorist (BT) attack, and suggest how community health units could leverage funding to start their own GIS in the form of a syndromic surveillance system.

Vulnerability in the U.S.

September 11th changed the world forever. Although there had previously been terrorist attacks on the United States, most notably the World Trade Center and

Oklahoma City bombings, the collapse of the twin towers opened the eyes of the U.S .populace to their own vulnerability. It should be remembered, however, that this vulnerability extends beyond such dramatic international attacks to include smaller terrorist statements, such as the Salmonellosis outbreak in Oregon in 1984 (Torok et al., 1997). In this case, a religious group "tested" the effectiveness of a biological agent spread on salad bars to manipulate a local election outcome.

In order to develop a comprehensive disaster response system that could be mobilized for any terrorist attack, a series of Homeland Security presidential directives were created. Presidential directive #8 states, "The term 'all-hazards preparedness' refers to preparedness for domestic attacks, major disasters, and other emergencies, There are three general interventions associated with a disaster: mitigation, response, and recovery. If the house is too close to the river, it may have to be protected against flooding. If the river floods, the residents of the house will need to be cared for. Once the river subsides, the family will need financial help to once again make their house a home" (http://www.whitehouse.gov/news/releases/2003/12/20031217-6.html). The purpose of this directive is to make sure these three stages are coordinated under an Emergency Operation Center (EOC) that can act efficiently to any occurrence.

Bioterrorism and Pregnancy Risk

Even if a community attempts to adhere to presidential directive #8 and develop preparedness strategies to cope with any hazard or disaster, it is still important to recognize that there is a difference between a BT attack and most other disasters. The first difference is the extent of the affected area, both in terms of geography (Flowers, Mothershead, & Blackwell, 2002) and time. For any disaster there will be a geography to the impact — the tornado's path, the flooded valley, the oil spill. A BT attack could literally be at any geographic scale, from the local salad bar to a nation-hopping disease (for a recent example of "what might be," consider the global impact of SARS during 2003).

Simply put, the larger the outbreak, the more a community will have to rely on its own resources. For any disaster, a common time frame of federal intervention is 72 hours. In other words, for the first 6 days local detection, response, and recovery will have to cope with the "situation." This time frame will be lengthened if, as many believe, a BT attack will have multiple foci, and even multiple agents. We will return to the consequence of this for pregnant women in a moment.

The geographic area under "attack" can also exceed the actual area impacted by the release of the agent. Consider the anthrax letters of 2001. Although relatively limited in agent release, at least in terms of how many cities actually received or were infected by the tainted mail, the geographic scope was immense. There was a national run on Cipro, the antibiotic traditionally prescribed to counter the effects of *anthrax bacilli*. Nationally, there were countless hoaxes, ranging from copycat letters to white powder being dumped on an office-mate's desk as a joke. A visiting epidemiology professor to LSU told me that his lab tested many hoax letters after the U.S. attack, and his lab is in Kazakhstan! None of these letters were positive, all were copycat hoaxes, and yet this was still an attack of sorts, an attack that if publicized can generate anxiety and stress. The uncertainty generated by a hoax can lead to an elevation in societal fear of the unknown (Slovic, 1987) and what is referred to as group "dread risk" (Slovic, Fischhoff, & Lichtenstein, 1985).

The second difference between a BT attack and most other disasters is one of time (Inglesby et al., 1999). Again referring back to the anthrax letters, a considerable amount of time was needed to "clean" infected property. Therefore, the number of impacted people is likely to be greater than a "normal" disaster, or at least the impact on them will be longer and therefore more stress-inducing. Although geography and time are tangible differences of a BT attack, it is a third difference, that of the psychological impact of a BT attack, that has ramifications for pregnant women. A great deal of uncertainty will surround a BT attack. It is likely that even official responding agencies will be unsure of what is happening, especially during the initial stages of the attack. Without being overly critical of local communities, it is likely that verification, treatment, and response will not be as well organized as at the federal level. Uncertainty will arise because of several factors. Was there an attack happening? Where was it happening? How contagious was the release? Was it being contained? Even if the "attack" was only a hoax, for example a media message stating an agent had been seeded into the water supply, the same uncertainties and anxieties would occur.

Much of this uncertainty arises from the fact that BT is genuinely something to be frightened about. Many officials have never seen the symptoms associated with a typical BT agent (such as smallpox), let alone a modified weaponized version. This fear easily transcends into the public sector. Many of us have seen films such as *Outbreak*, which irrespective of their scientific accuracy leave their mark. Further anxiety can result if relocation is required (Carus, 2001) or if quarantine is imposed, as this "preventative" measure is often accompanied by considerable social distrusts and fear. Extreme examples could include a break down of social norms, including rioting and looting (O'Toole, 1999), though it is also suggested that with appropriate, accurate, and culturally sensitive information releases, these same neighborhoods can be turned into response resources (Dynes & Tierney, 1994; Freimuth, Linnan, & Potter, 2000; T. A. Glass & Scoch-Spana, 2002).

GIS and Vulnerability Mapping

GIS use is well established in the field of disasters and hazards (for examples see Hodgson & Cutter, 2001; Marcello, 1995; Newsome & Mitrani., 1993; Radke et al., 2000). For example, it has previously been used in the transportation planning of hazardous material (Brainard, Lovette, & Parfitt, 1996; Estes, McGwire, Fletcher, & Foresman, 1987), evacuation planning (Cova & Church, 1997) and social vulnerability analysis in terms of proximity to toxic locations (Chakraborty & Armstrong, 1997; Cutter, Hodgson, & Dow, 2001). It is this last use of GIS in targeting and mapping vulnerable communities (Diaz & Pulwarty, 1997; Walker, 1995) that has particular relevance to indigent pregnant women. It is widely acknowledged that different cohorts will experience disasters in different ways (Weiss et al., 2002). The most obvious difference comes with proximity to the incident (Galea et al., 2002), though communities will have fewer resources to cope with the resulting suffering. It is therefore important to identify who is most vulnerable, whether as cohorts or by geographic areas, so that appropriate mitigation and response strategies can be formulated. Hill and Cutter (2001) break vulnerability into three risk categories: individual (choices), social (surrounding neighborhood), and biophysical. Similarly Morrow (1999) and Jaspers (1999) partition vulnerability into the individual behavioral realm such as the activity space, the physicality of environments including the location and construction of the home as well as who owns it, and the surrounding social and political fabric. These risks combine to present complicated vulnerability landscapes, comprised of social, racial, and economic layers overlaid on the community's physical and service infrastructure.

The reaction of different cohorts can vary in the face of the same risk, therefore the hazard must be placed into the context of those it is likely to impact (Palm, 1990) with both macro (city) and micro (neighborhood/individual) decision-making strategies being involved in any response plan (Palm & Hodgson, 1992). The neighborhood, as described in Chapters IV and V, is a milieu of political, social, racial, and economic conditions, which should be viewed in terms of disaster vulnerability as was previously described for general health vulnerability. The same GIS approaches used to identify the neighborhoods of infant risk can be used to identify neighborhoods of general vulnerability. Unfortunately, the risk and hazard literature has largely ignored pregnant women when considering vulnerability analysis and disaster response. This is a gross oversight, especially as has been previously mentioned, stress can result in a negative birth outcome. Disasters are definitely stress-inducing, and especially a BT attack. All things being equal, women coming from areas of a city that continually experience high rates of negative birth outcomes are arguably the most vulnerable of all cohorts, not only at the time of the incident, but also afterwards as a stress echo follows any woman who is pregnant or becomes pregnant.

At a recent course development meeting held at FEMA, when I brought this point up I was ridiculed for even suggesting it. The argument was that in the event of a disaster, how could services be prioritized, how could individual pregnant women be identified, and should we not serve everyone equally in an unfolding situation? Although I can understand the thinking behind such comments, it is very much a white male attitude displaying an ignorance of the needs of pregnant women, the role the neighborhood plays in their pregnancy, and the additional stresses they will face in the recovery stage of a disaster. We will return to these points presently.

Identifying the Vulnerable

It is not only neighborhood level social risks that make the poor more vulnerable during a disaster; in terms of infrastructure the poor often live in overly crowded central city areas (which could have implications in the event of a contagious agent release), and have the furthest to travel to safety outside of the city. In addition, one of the traditional reasons for women not making prenatal visits, that of lack of transport, also plays a role here. A pregnant woman, especially with little family support, is reliant on either public transport or the efficiency of local emergency evacuation planners to escape the areas of threat (Cova & Church, 1997). An example of this can be found in New Orleans, where many inner-city residents live below sea level and are extremely vulnerable to a hurricane strike. Many of these families have no private transport.

As an aside, this forced evacuation has to be seen to be believed. In the summer of 2004 Hurricane Ivan threatened New Orleans. Three days beforehand, all three lanes of the interstate were bumper to bumper in Baton Rouge, over 70 miles away, as people fled the city. Watching this mass flight was in equal measure impressive and disturbing.

A general lack of understanding, which might result from a lower educational attainment, can also play its part in terms of comprehending what is happening during an incident, especially given the general uncertainty of a BT attack. There have been analogous situations in which a correctly targeted information campaign limited the spread of meningitis (Bray, 1996). This lack of education can also have post incident implications affecting the victim's ability to successfully navigate government compensation programs, thus prolonging the impact of the incident (Bolin, 1986; Bolin & Bolton, 1986; Mileti, 1999). In order to diminish these problems of understanding, a mechanism is needed whereby the culture and social structure of the neighborhood under threat is incorporated into response and recovery strategies (Freimuth et al., 2000; T. A. Glass & Scoch-Spana, 2002; Klitzman & Freudenberg, 2003).

It is, though, not enough to simply map indigent neighborhoods and mark them as elevated risk areas (Chambers, 1989). As was stressed in Chapter IV, the multiple intersecting risks present a complex situation (Dasgupta, 1995). Even after the event, a condition that disproportionately affects the poor is increased likelihood to suffer psychological distress in the face of a disaster. As an example, consider Rehner, Kolbo, Trump, Smith, and Reid's (2000) study of Methyl Parathion (MP), a pesticide used to control the boll weevil in rice fields exposure in southern Mississippi. Although not approved for indoor use, 1,800 buildings, mainly homes but also community interaction foci such as day care centers and churches, were treated with MP over a 10-year period by unscrupulous pesticide sprayers. The immediate consequences of the exposure included the psychological stress associated with being forced to leave their homes. However, after the incident many victims also suffered from post-event depression. In the study, the authors found that the groups suffering the most from post-event depression were poor, African Americans, and women. Depression was associated with the length of exposure, and not the degree of exposure. A possible explanation for this relationship was the compounding uncertainty of whether the home was making them ill. A poverty trap was in full effect. Those families not reaching a level of exposure "deserving" of financial assistance to relocate were left with no choice but to pay for clean-up, or live with the exposure. Many families could not afford to move; indeed, the original exposure was caused because it was a cheap treatment option. Of those diagnosed as suffering from depression, only 16% went on to receive treatment.

Two points should be gleaned from this example: First, the poor often suffer elevated and prolonged psychological impacts after a disaster due to a lack of financial power to change their lot. Second, they are often unlikely to (or unable to) seek appropriate medical help for similar reasons. Both these points have implications when considering indigent pregnant women.

Women within poor communities are often hardest hit by hazards or disasters (for examples see Cutter, 1995; Enarson, 1998; Enarson & Morrow, 1997; Fothergill, 1996; Noel, 1998; Rivers, 1982; Stehlik, Lawrence, & Gray, 2000). These impacts range from increased exposure to hazards in the domicile (examples being found in Cutter, 1995)), to the less well-defined changes that occur in the family and neighborhood social structure after an incident. The home is a complex environment in its own right (Massey, 1994), and a disaster can precipitate a relationship into collapse. Relocation and/or the loss of items associated with the memory of the relationship can lead to arguments, separation, abuse, and depression. The way the family fits into the neighborhood support structure can also be affected. The poor are often reliant on extended families, or neighborhood support mechanisms. If that community is seriously disrupted, or even lost due to relocation, this support structure (such as free child care) could be permanently lost. It has been shown that for a woman, maintaining this support structure can be more important than preserving the actual home itself (Fordham, 1998). Obviously, these relationships could have implications in heeding evacuation recommendations, with some groups being less likely to move in the face of a threat. All of these factors are magnified when the woman is also the head of her household, as she is both provider and family caregiver (Morrow, 1999). A further compounding gender difference is that women are more likely to suffer elevated stress levels than men in relation to potential risks facing the home, or immediate environment, such as pollution or infrastructure deficiencies. Although obviously a generalization, these anxieties are likely to be magnified if the woman already has young children in the home, or is pregnant. All of these additional stressors associated with being female in the face of a "threat" can have implications for birth outcomes.

Studies have also shown that minorities often suffer disproportionately in the aftermath of a disaster (Bolin, 1986; Bolin & Bolton, 1986), such as elevated levels of psychological distress (Phillips, 1993). Although some would argue that the risks faced by a minority population are the same as those experienced by a poor population (in other words, minority equals poor), this is rejected for the same reasons as discussed in Chapter IV. For example, an often-voiced criticism is that responders in a disaster are frequently white males. During a time of stress, it is not surprising that a responder distributing orders in an unfamiliar minority community will be unfamiliar with the dynamics of that neighborhood, a situation which might lead to a potential conflict situation. This is particularly important when dealing with a BT threat, as the uncertainty of the attack itself needs to be explained effectively, meaning a culturally sensitive delivery so as to assuage as much anxiety as possible (Phillips, 1993). A common thread through the BT communication literature is that information must be disseminated by culturally sensitive, and ideally, locally accepted voices (D. C. Glass et al., 1973; T. A. Glass & Scoch-Spana, 2002).

So How Do We Bring Healthy Start into This?

Any impacted population is going to face stress during a disaster, though as suggested previously, the general uncertainty surrounding a BT attack is likely to magnify these stresses and anxieties even further in poor, female, minority populations. It is therefore somewhat surprising that although response recommendations exist for nursing mothers in developing world situations, little advice exists for mothers and expectant mothers in the United States. Consider the following example: A *Family Readiness Guide for Bioterrorism Preparedness and Emergency Response*, funded by the Louisiana Department of Health and Hospitals, the Louisiana Office of Public Health, the Centers for Disease Control and Prevention, the Louisiana Office of Emergency Preparedness, and the U.S. Department of Homeland Security sits on the counter at the Family Road center (housing the Healthy Start program) in Baton Rouge. This center is a nonprofit organization dedicated to pregnancy-related issues, with its service population being largely drawn from indigent neighborhoods in the city. The guide, in the form of a newspaper, is a good attempt at educating the population about hazards, and specifically in the event of a BT emergency, how to recognize symptoms. Suggestions for family emergency measures are presented and a checklist covers everything from supplies for babies to the needs of the elderly. Although many of the checklist items would also be relevant to a pregnant woman, and the general message of "knowledge is power" is a good one, there is no obvious mention in the document about precautions to be taken if pregnant. It is likely that a pregnant woman would prefer to read about precautions to be taken if pregnant, and not be forced to infer the correct course of action from such general guidelines. Although previous BT papers have encouraged a targeting of education and information release strategies which are culturally sensitive, and if possible delivered by a respected individual with ties to the neighborhood, this same logic should apply to actual cohorts at risk, such as pregnant women.

Are Pregnant Women Really Vulnerable?

The vulnerability of being pregnant is on one level immediately obvious. Even the most independent, self-reliant woman, irrespective of home environment or financial status, becomes vulnerable during pregnancy. Risks from the environment, such as secondhand smoke, environmental toxins, and access to medical care all cause concern for a pregnant woman. In terms of reacting to a disaster or BT attack, all of these issues could become important if the woman has to be relocated. Depending on the stage of her pregnancy, the physical process of relocation could become traumatic and stress-creating. If she originates from an inner-city area, using the Baton Rouge Healthy Start as a typical population, she may have no spouse, no transport, may tire quickly, and be worried about accessing her needed medications. All of these are stress-inducing situations.

Individual signs of distress can manifest as depression, feeling overwhelmed, and feeling abandoned (DeMarco, 2004). More worryingly, it has also been found that stress associated with the September 11th attacks led to an increase in tobacco and alcohol use (Hall, Norwood, Ursano, & Fullerton, 2003), both of which have been linked to negative birth outcomes.

This connection between negative birth outcomes and living in stress has been gaining more momentum. Recent research has begun to focus on these stressors and the impacts they have; for example, a recent research proposal (December, 2003) announcement from the Department of Health and Human Services focusing on "Reducing Preterm and Low Birth Weight in Minority Families" states:

Increasingly, social stress has become the focus of much research and, at the same time, it has become apparent that it is important to understand the context, especially the neighborhood context, in which women at risk for poor pregnancy outcomes live their lives.

People tend to worry disproportionately about some risks, and worry leads to stress. In the event of a BT attack, fears are likely to continue well after the incident for pregnant women, as uncertainty remains in terms of has the building or area been cleaned effectively, or has any potential exposure led to a problem with the developing fetus?

As was previously stated, the poor and especially minority females are likely to suffer these stresses to an even greater degree, a cohort already most at risk from negative birth outcomes in the first place. The question to be asked, but as far as I can tell has not as of yet, is to what degree will the echo of a disaster, such as flooded areas after a hurricane, lead to increases in low-birth-weight deliveries and infant mortality?

It would seem prudent to take any precautions possible to limit the anxieties these populations will face. It is therefore imperative that education programs are developed for pregnant women which are disaster-specific, though response-general. For example, culturally sensitive information specific to BT, or for that matter any toxic matter release, will help lessen anxiety (Slovic, 1987; Slovic et al., 1985). Questions can be answered such as such as, "How vulnerable am I? What effect could the release have on my unborn child?" Or more topically, "What effect might vaccines have on my unborn child?" (GAO Report 03-578, 2003) Although all pregnant women as a cohort should be educated in this way, an initial prioritization should occur aimed at the most vulnerable women. The GIS approaches presented throughout this book provide an excellent first step in identifying where the most vulnerable live in a city. Neighborhoods showing temporally stable negative infant outcomes would be the first place to start. Even during peacetime, these neighborhoods have serious problems and it is likely that any life disruption such as would likely follow a disaster, and especially a BT attack, will worsen this situation. Simply stated, if an attack impacted all of Baton Rouge, *ceteris paribus*, high-risk neighborhoods would be the most vulnerable to negative infant health outcomes, both during and after the incident.

But how can we reach pregnant women during an unfolding disaster? In two ways: by helping those women who are already enrolled in pregnancy programs, and by using their experiences to develop broader neighborhood initiatives. In other words, the information contained in the Healthy Start GIS could be used as a (admittedly biased) sample of the neighborhood social structure. For example, the proportion of women who are heads of their household, the proportion having young children in the home, the proportion having no family support, and no transportation, the proportion reliant on neighborhood daycare, and the proportion with medication needs — all of these "risks" have disaster scenario implications.

In one query of the Healthy Start GIS, the following risk combination was queried out: Women who already have a young child in the house (in addition to the current pregnancy), are single, heads of their household, and have no transport. Of the 21 program participants identified, only four have full-time employment, nine are unemployed, and one is registered as disabled. Of the 21, 18 are single, nine live in households with no annual income, eight acknowledged they suffer from stress, seven scored at the risk level on the Edinburgh Depression scale, one had suicidal thoughts, and one suffered from domestic violence. If we again see the Healthy Start population as a sample for their neighborhood, these findings could be used to target the response and recovery efforts so as to minimize their stress. If a hurricane were approaching, we could estimate the proportion needing transportation, if relocation was required, we could estimate the likely number needing daycare, and afterwards, which neighborhoods would most likely benefit from posttraumatic stress counseling (based on the psychological testing of the Healthy Start Program participant).

It makes sense that the Baton Rouge Healthy Start GIS should be included in an all-hazards preparedness plan for the city. And yet it is not. A further push is needed to introduce these systems as part of a response plan. That answer might just be syndromic surveillance, a means to identify if a BT attack was happening. Syndromic surveillance systems attempt to illicit patterns from nontraditional surveillance structures, searching for patterns in symptoms (such as fever) rather than reported clinical diagnosis. Examples of these systems range from spikes in over-the-counter drug sales (Goldenberg, Shmueli, Caruana, & Fienberg, 2002) to symptomatic extraction from insurance record entries. Other surveillance systems combine both traditional (hospital records) and nontraditional data (such as absenteeism from work) (Lazarus, Kelinman, Dashevsky, DeMaria, & Platt, 2001; Lewis et al., 2002; Lober et al., 2002; Miller & Mikol, 1999). The benefit of creating an early detection system is obvious, as it would limit the number of infected people and make containment easier through quarantine or vaccination. In addition, an early detection could also help identify the origin of release and the collection of forensic evidence.

Although the Baton Rouge Healthy Start GIS was designed to improve health outcomes in high-risk areas of the city, it could also be modified for syndromic surveillance. The benefit of introducing the GIS into a BT preparedness plan would be the creation of a daily monitored syndromic surveillance network, which could be automated by a decision support system trained to recognize significant elevated levels of sickness. The peacetime public health benefit of the system would be the introduction of a technology and technique that could be used to improve birth outcomes in one of the most disadvantaged cohorts in the United States, courtesy of Homeland Security funding. “Training” information, such as influenza outbreaks, would also offer insight into traditional healthcare networks (where does someone with the flu go?) and behavioral patterns (at what level of sickness does an individual seek help?). However, the primary benefit is that the vulnerability of being pregnant is elevated in disaster response planning. “Risk neighborhoods” would be identified and catered for in the event of a disaster or attack by responders who are familiar and knowledgeable of the local community. This in turn can reduce tensions due to the caseworkers being familiar and culturally sensitive.

The Healthy Start GIS also fits in with the general recommendations for improving BT preparedness, which include the generation of better neighborhood-level detail, especially demographic vulnerability data (Pavlin et al., 2003). Similar comments were made after a BT exercise in Denver, specifically the need for real-time data analysis expressed as spatial outcomes (Inglesby, Grossman, & O’Toole, 2002). The Baton Rouge Healthy Start GIS certainly contains this type of quality neighborhood detail.

The current database includes approximately 250 program participants, but this number is eventually expected to rise to 400 — which is almost 10% of the Baton Rouge pregnant population. Although this GIS is one of the first of its kind, the program would easily be transferable to other pregnancy-related centers (for example, there are currently over 100 Healthy Start programs in the United States). Indeed, one of the objectives of the federal initiative Healthy People 2010 is that 90% of all health reporting agencies, such as Healthy Start, will be creating electronic patient data in a GIS environment by the end of the decade.

Caseworkers could inform their program participants to immediately call in any signs of illness, such as fever. In this way, a sample of program participants spread across the city could be used to detect patterns of emerging sickness. For some cities, the indigent at-risk population may be centralized, in which case other similar systems would have to be included in order to cover the entire city. In Baton Rouge, which is a fragmented city of social classes, although certain cohort neighborhoods (white, middle class) will be missed, the entire city will be adequately covered in terms of areal extent. Instructing the program participants to report any other sickness in their home could also enlarge the sample size. Symptoms would be entered into the GIS, where the location of the program participant is already stored. It would be simple to map these call-ins to identify developing spatial clusters. Indeed, a spatial decision support system could be constructed whereby if thresholds are met, an automatic warning occurs. These thresholds could be based on a percentage of all program participants in the GIS reporting symptoms, or if a spatial cluster develops in any section of the city. For example, a warning threshold could be if proportion X of all program participants residing within a 0.25-mile radius reports the onset of fever.

Criticisms of Syndromic Surveillance

Several problems have been identified with syndromic surveillance, possibly the most important of which is validation (Reingold, 2003). How could a peak in over-the-counter drug purchases be judged? Against what base line would this anomaly become significant? And was the anomaly the result of an attack or a naturally occurring outlier? How much observation time would be needed before the above questions are answered satisfactorily, resulting in a response team mobilization? Reingold (2003) comments that the only sure way of answering these questions would be to compare events to a previous BT attack, which is not an option. Other validation strategies could include the application of the surveillance system to identify or predict non-BT generated disease clusters, such as influenza seasons (Centers for Disease Control and Prevention, 2001) or febrile respiratory illness (Susser, Herman, & Aaron, 2002). Pavlin et al. (2003) suggest an alternative strategy of cross-validating multiple surveillance systems, with anomalies in one being linked to anomalies in another. A further approach, again following their suggestion, is to use a population that was already well-understood by responders who were familiar with both victims and their neighborhoods, so that anomalies could be quickly interpreted. This, in effect, describes the Healthy Start GIS of Baton Rouge, and its (near) real-time data collection strategy.

A useful syndromic surveillance would also have to meet the following points: (1) the system needs to be flexible and allow for early detection; (2) the system needs to be interpreted by a team comprised of both medical and community expertise; (3) validation is required, or at least comparative base-line data is needed; (4) the technology should exist for multiple syndromic surveillance data sets to be immediately analyzed and cross-referenced; and (5) the cost of a BT initiative, such as syndromic surveillance, is often at the expense of other public health programs.

Does the Healthy Start GIS meet these criteria? (1) The system is flexible in that the database could quickly be modified to incorporate new data. The training, data collection and analysis would not have to change if more or different data were needed. (2) The data for the Baton Rouge Healthy Start is collected by either nurses or social workers. Healthy Start caseworkers serve limited numbers of program participants, therefore they "know" their program participants relatively well. This familiarity is extended to the neighborhoods in which their program participants live. (3) Validation would occur in two ways. First, by creating base-line reference data (as the GIS is an ongoing data collection process) against which future space and time clusters can be compared. Once the reporting of symptoms is established as a norm, base-line call-ins could be analyzed both spatially and seasonally. It would certainly be possible to analyze naturally occurring epidemic seasons, such as influenza (therefore complying with CDC recommendations for system validation). These data could be used to help "guide" the decision support system: What is the normal number of call-ins? What does the call-in pattern resemble during an epidemic season? Are these call-ins randomly distributed about the city or are they clustered? The second method of validation would be by direct contact. If a cluster of symptoms occurred, it would be relatively easy for the caseworker to either home-visit, or arrange for the program participant to be brought into the clinic for testing. Obviously the severity of the symptoms would dictate the level of response. (4) Validation can also be achieved if multiple surveillance systems are analyzed for similarities in clusters (Pavlin et al., 2003). Although the Healthy Start GIS would not do this per se, as long as national metadata standards (as defined by the Federal Geographic Data Committee) were met during the creation of the system, output data would be immediately transferable to an EOC Enterprise GIS (a system allowing multiple user access in a non-GIS specialist environment). Therefore, the Healthy Start GIS would at least be able to play its part in the simultaneous analysis of a syndromic surveillance data network.

(5) The cost of the Baton Rouge Healthy Start is relatively inexpensive. The teething troubles experienced in the creation of the initial GIS would not be passed onto similar programs. Indeed, in keeping with Reingold's (2003) suggestion of an academic and public health collaboration being the logical way forward for syndromic surveillance, the task of maintaining and analyzing the data could be performed by a PhD student under a spatial scientist's supervision (as is currently the case for the Baton Rouge Healthy Start). Departments of geography at local universities would be more than happy to collaborate with

these programs. Academics would benefit by having access to excellent data for their research and teaching. Doctoral students would have a source of funding and a potential research topic. The creation of a spatial decision support system linked to a secure Web site where incoming data from caseworkers were uploaded in real time would limit the need for constant monitoring. The researchers could spend most of their time investigating the data for other, normal patterns of risk that could prove of benefit to the program, such as identifying spatial clusters of infant mortality.

Although creating an entire pregnancy-related system from the beginning could be cost-prohibitive, most cities have similar pregnancy programs already in place, many of which are aimed at improving birth outcomes in indigent populations. As was previously mentioned, there are over 100 Healthy Start programs already in the United States. The (relatively inexpensive) cost would be in the addition of a GIS to these programs. The longevity of the GIS would also be related to the program itself. If Homeland Security were to fund the academic liaison on a permanent basis, the program would have every incentive to continue with the GIS. Again as previously stated, Healthy People 2010 will push these health care programs toward this type of data collection in the near future, irrespective of any Homeland Security involvement.

One final comment should be made on the ethics of data collection. Obviously, we must conform to HIPPA requirements in the collection of data. Patient confidentiality must be maintained during the surveillance process. During normal conditions (for example background influenza) there is no need for any member of the preparedness team to know information about program participants — a summary report detailing illness by sections of the city would be sufficient. However, if a threshold were exceeded whereby a sickness cluster generated concern of a possible attack, the caseworker would work with the appropriate medical component of the response team in terms of interviewing the program participant and administering appropriate medical care. The active role of the caseworker not only helps control the creation and distribution of these data, but also helps nullify several of the criticisms often leveled at disaster response, that there is an unfamiliarity with the neighborhoods under threat, with responders also being of a different gender (Enarson & Morrow, 1997) and race (Phillips, 1993). In this way, potential conflicts between responders and victims could be avoided. The caseworker should also be able to relay information to the program participant so as to reduce the anxiety that naturally occurs during a disaster, especially during a BT attack. Indeed, it was noted after the September 11th attack that the recovery of the impacted population would have been quicker had more community voices been incorporated into the emergency response (Klitzman & Freudenberg, 2003).

I admit this is "pretty out there," and criticisms have been made about the ethics of using a poor African American population as "canaries." Hopefully, readers of this book will know by now that my intentions to serve this population are honorable. But this is still missing the two major points of this chapter. First, by directing attention to the vulnerability of being pregnant in high-risk neighborhoods during any disaster, mitigation, response, and recovery strategies will hopefully become more sensitive. Second, the chances of any attack happening are remote, and while funding is available to develop syndromic surveillance, why not use it to create some good at the same time?

References

Bolin, R. C. (1986). Disaster impact and recovery: A comparison of black and white victims. *International Journal of Mass Emergencies and Disasters, 4*(1), 35-51.

Bolin, R. C., & Bolton, P. (1986). *Race, religion, and ethnicity in disaster recovery*. Boulder: University of Colorado.

Brainard, J. S., Lovette, A., & Parfitt, J. P. (1996). Assessing hazardous waste transport risks using a GIS. *International Journal of Geographical Information Systems, 10*, 831-849.

Bray, J. (1996). Public information. *Nurs Stand, 11*(12), 7.

Carus, W. S. (2001). *The illicit use of biological agents since 1990.* Working paper: Bioterrorism and biocrimes. Center for Counterproliferation Research. Washington, DC: National Defense University.

Centers for Disease Control and Prevention. (2001). *Updated guidelines for evaluating public health surveillance systems: Recommendations from the Guidelines Working Group* (MMWR Morb Mortal Weekly Reports. 2001: 50). Centers for Disease Control and Prevention.

Chakraborty, J., & Armstrong, M. P. (1997). Exploring the use of buffer analysis for the identification of impacted areas in environmental equity assessment. *Cartography and Geographic Information Systems, 24*, 145-157.

Chambers, R. (1989). Vulnerability, coping and policy. *IDS Bulletin, 20*(2), 1-7.

Cova, T. J., & Church, R. L. (1997). Modeling community evacuation vulnerability using GIS. *International Journal of Geographical Information Science, 11*, 763-784.

Cutter, S. (1995). The forgotten casualties: Women, children and environmental change. *Human and Policy Dimensions, 5*(3), 181-194.

Cutter, S. L., Hodgson, M. E., & Dow, K. (2001). Subsidized inequities: The spatial pattern of environmental risks and federally assisted housing. *Urban Geography, 22*(1), 29-53.

Dasgupta, P. (1995). Population, poverty, and the local environment. *Scientific American, 272*, 40-45.

DeMarco, P. (2004). *Mental health issues and weapons of mass destruction*. Presentation at the symposium Protecting the Homeland – Focus on Healthcare Baton Rouge, June 3.

Diaz, H., & Pulwarty, R. (Eds.). (1997). *Hurricanes: Climate and socioeconomic impacts*. New York: Springer.

Dynes, R. R., & Tierney, K. J. (Eds.). (1994). *Disasters, collective behavior and social reorganization*. Newark, DE: University of Delaware Press.

Enarson, E. (1998). Through women's eyes: A gendered research agenda for disaster social science. *Disasters, 22,*(2), 157-173.

Enarson, E., & Morrow, B. (1997). A gendered perspective: The voices of women. In W. G. Peacock, B. Morrow, & H. Gladwin (Eds.), *Hurricane Andrew: Ethnicity, gender, and the sociology of disasters*. London: Routledge.

Estes, J. E., McGwire, K. C., Fletcher, G. A., & Foresman, T. W. (1987). Coordinating hazardous waste management activities using geographical information systems. *International Journal of Geographical Information Systems, 1*, 359-377.

Flowers, L. K., Mothershead, J. L., & Blackwell, T. H. (2002). Bioterrorism preparedness II: The community and emergency medical services. *Emergency Medicine Clinics of North America, 20*, 457.

Fordham, M. H. (1998). Making women visible in disasters: Problematising the private domain. *Disasters, 22*(2), 126-143.

Fothergill, A. (1996). Gender, risk, and disaster. *International Journal of Mass Emergencies and Disasters, 14*(1), 33-55.

Freimuth V., Linnan, H. W., & Potter, P. (2000). Communicating the threat of emerging infections to the public. *Emerging Infectious Diseases, 6*, 337-347.

Galea, S., Ahern, J., Resnick, H., Kilpatrick, D., Bucuvalas, M., Gold, J., et al. (2002). Psychological sequelae of the September 11 terrorist attacks in New York City. *New England Journal of Medicine, 346*(13), 982-987.

GAO Report 03-578. (2003). *Smallpox vaccination: Implementation of national program faces challenges*.

Glass, D. C., Singer, J. E., Leonard, H. S., Krantz, D., Cohen, S., & Cummings, H. (1973). Perceived control of aversive stimulation and the reduction of stress responses. *Journal of Personality, 41*(4), 577-595.

Glass, T. A., & Scoch-Spana, M. (2002). Bioterrorism and the people: How to vaccinate a city against panic. *Confronting Biological Weapons, 34*, 217-223.

Goldenberg, A., Shmueli, G., Caruana, R., & Fienberg, S. (2002). Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. In the *Proceedings of the National Academy of Sciences, USA, 99*(8), 5237-40.

Hall, M., Norwood, A., Ursano, R., & Fullerton, C. (2003). The psychological impacts of bioterrorism. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science, 1*(2), 139-146.

Hill, A., & Cutter, S. (2001). Methods for determining disaster proneness. In S. Cutter (Ed.), *American hazardscapes — The regionalization of hazards and disasters* (pp. 13-36). Washington, DC: Joseph Henry Press.

Hodgson, M., & Cutter, S. L. (2001). Mapping and the spatial analysis of hazardscapes. In S. L. Cutter (Ed.), *American hazardscapes – The regionalization of hazards and disasters* (pp. 37-60). Washington, DC: Joseph Henry Press.

Inglesby, T., Grossman, R., & O'Toole, T. (2002). A plague on your city: Observations from TOPOFF. *Biodefense Quarterly*, *2*(2), 1-10.

Inglesby, T. V., Henderson, D. A., Bartlett, J. G., Ascher, M. S., Eitzen, E., Friedlander, A. M., Hauer, J., McDade, J., Osterholm, M. T., O'Toole, T., Parker, G., Perl, T. M., Russell, P. K., & Tonat, K. (1999). Anthrax as a biological weapon: Medical and public health management. *Journal of the American Medical association, 281*(18), 1735-1763.

Jaspers, S. (1999). Targeting the vulnerable: A review of the necessity and feasibility of targeting vulnerable households. *Disasters, 4*, 359-373.

Klitzman, S., & Freudenberg, N. (2003). Implications of the World Trade Center attack for the public health and health care infrastructures. *American Journal of Public Health, 93*(3), 400-406.

Lazarus, R., Kelinman, K., Dashevsky, I., DeMaria, A., & Platt, R. (2001). Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): The example of lower respiratory infection. *BMC Public Health, 1*, 9.

Lewis, M. D., Pavlin, J. A., Mansfield, J. L., O'Brien, S., Boomsma, L. G., Elbert, Y., & Kelley, P. W. (2002). Disease outbreak detection system using syndromic data in the greater Washington DC area. *American Journal of Preventative Medicine, 23*, 180-186.

Lober, W. B., Karas, B. T., Wagner, M. M., Overhage, J. M., Davidson, A. J., Fraser, H., Trigg, L. J., Mandl, K. D., Espino, J. U., & Tsui, F. C. (2002). Roundtable on bioterrorism detection: Information system-based surveillance. *The Journal of the American Medical Association, 9*, 105-115.

Marcello, B. (1995). FEMA's new GIS reforms emergency response efforts. *GIS World, 8*(12), 70-73.

Massey, D. (1994). *Space, place and gender*. Cambridge, UK: Polity Press.

Mileti, D. (1999). *Disasters by design: A reassessment of natural hazards in the United States*. Washington, DC: Joseph Henry Press.

Miller, J., & Mikol, Y. (1999). Surveillance for diarrheal disease in New York City. *Journal of Urban Health, 76*, 388-390.

Morrow, B. (1999). Identifying and mapping community vulnerability. *Disasters, 23*(1), 1-18.

Newsome, D. E., & Mitrani., J. E. (1993). Geographic information system applications in emergency management. *Journal of Contingencies and Crisis Management, 1*, 199-202.

Noel, G. E. (1998). The role of women in health-related aspects of emergency management. In E. Enarson, & B. H. Morrow (Eds.), *The gendered terrain of disasters: Through women's eyes*. Westport, CT: Greenwood Publishing.

O'Toole, T. (1999). Smallpox: An attack scenario. *Emerging Infectious Diseases, 5*, 540-546.

Palm, R. (1990). *Natural hazards: An integrative framework for research and planning*. Baltimore: John Hopkins University Press.

Palm, R. I., & Hodgson, M. E. (1992). Earthquake insurance: Mandated disclosure and homeowner response in California. *Annals of the Association of American Geographers, 82*, 207-222.

Pavlin, J. A., Mostashari, F., Kortepeter, M. G., Hynes, N. A., Chotani, R. A., Mikol, Y. B., Ryan, M. A., Neville, J. S., Gantz, D. T., Writer, J. V., Florance, J. E., Culpepper, R. C., Henretig, F. M., & Kelley, P. W. (2003). Innovative surveillance methods for rapid detection of disease outbreaks and bioterrorism: Results of an interagency workshop on health indicator surveillance. *American Journal of Public Health, 93*(8), 1230-1235.

Phillips, B. D. (1993). Cultural diversity in disasters: Sheltering, housing, and long-term recovery. *International Journal of Mass Emergencies and Disasters, 11*(1), 99-110.

Radke, J., Cova, T., Sheridan, M. F., Troy, A., Mu, L., & Johnson, R. (2000). Application challenges for geographic information science: Implications for research, education, and policy for emergency preparedness and response. *URISA Journal, 12*(2), 15-30.

Rehner, T., Kolbo, J., Trump, R., Smith, C., & Reid, D. (2000). Depression among victims of south Mississippi's Methyl Parathion disaster. *National Association of Social Workers, 25*(1), 33-40.

Reingold, A. (2003). If syndromic surveillance is the answer, what is the question? *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science, 1*(2), 77-81.

Rivers, J. (1982). Women and children last: An essay on sex discrimination in disasters. *Disasters, 6*(4), 256-267.

Slovic, P. (1987). Perception of risk. *Science, 236*, 280-285.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1985). Characterizing perceived risk. In R. Kates, C. Hohenemser, & R. Kasperson (Eds.), *Perilous progress: Managing the hazards of technology*. Boulder, CO: Westview.

Stehlik, D., Lawrence, G., & Gray, I. (2000). Gender and drought: Experiences of Australian women in the drought of the 1990s. *International Journal of Disaster Studies, 24*(1), 38-53.

Susser, E. S., Herman, D. B., & Aaron, B. (2002). Combating the terror of the terrorism. *Scientific American, 287*, 72-77.

Torok, T. J., Tauxe, R. V., Wise, R. P., Livengood, J. R., Sokolow, R., Mauvais, S., Birkness, K. A., Skeels, M. R., Horan, J. M., & Foster, L. R. (1997). A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. *Journal of the American Medical Association, 278*, 389-395.

Walker, B. (Ed.). (1995). *Women and emergencies*. Oxford: Oxfam.

Weiss, L., Fabri, A., McCoy, K., Coffrin, P., Netherland, J., & Finkelstein, R. (2002). A vulnerable population in a time of crisis: Drug users and the attacks on the World Trade Center. *Journal of Urban Health, 79*, 392-403.

Chapter XI

Rural Health Issues and Their Investigation in a GIS Environment

By Guest Author Jackie Mills
Louisiana State University, USA

Introduction

The previous chapters have provided a background in factors related to infant and maternal health and the various ways these individual and neighborhood characteristics can be studied in a GIS environment. Though rich in information for approaching investigations in urban areas, this book has yet to delve into these issues from a rural perspective, wherein some of the salient variables differ and certainly these data require additional understanding for use with GIS. Therefore, this chapter introduces some of the topics that could be included in health investigations in rural areas, along with some of the caveats for working with and interpreting the GIS results.

The Complexity of Rurality

Several years ago, a colleague questioned my interest in the rural South. He said that he found it boring, uneventful, and catatonic. Others overhearing the conversation agreed, the consensus being that the most interesting processes in the human and environmental realms occur in urban places. I respectfully disagree. Although cities and their surrounding areas are dynamic places, rural America has its own ironies, anomalies, and global interconnections; some are hidden in what appears to be a quiet landscape, and others are blatant for those who care to look.

Once it is understood that rural places can display diverse and sometimes complicated patterns, the next issue to address is the question of "what is rural?" This question has both concrete definitions provided by federal agencies, and discursive debate among academics, not surprisingly without an agreed upon definition (Economic Research Service, 2005; United States Department of Agriculture, 2005; Delatorre, Fickenscher, & Luft, 1991; Conrad, 1991). Before the advent of trains and automobiles, what was urban and what was rural was more clearly visible. Since these advances in transportation, however, what was once farmland is being divided into parcels for bedroom communities — havens for employees in nearby urban areas and even developing into cities in their own right (Morrill, Cromartie, & Hart, 1999; Garreau, 1991; Hart, 1991). Urban areas are sprawling and the urban-rural fringe is a blur.

Most of the land area in this country is considered non-metropolitan, or rural, depending on which definition you ascribe to. Though most of the U.S. population lives in cities, approximately 20%, or 59.2 million, people reside in rural areas (Census 2000 Population Statistics). The particularly rural places in which the millions of these people live, work, and raise families have distinct geographies from their urban counterparts and therefore present alternative health issues and patterns which must be understood in order to conduct a reliable GIS investigation of related matters. In order to aid understanding, this chapter provides an overview of rural places and health, specifically targets the health concerns of rural infant mortality, rural ghettoes, and agricultural structure and chemicals, and discusses ways to deal with rural data. Some of these topics have been covered extensively and therefore I see no reason to repeat, but other aspects such as agristructure and rural ghettoes have received little coverage as they pertain to medical issues. For examples, I will refer to my geographic area of study, the Mississippi Delta.

Rural Places and Health

In 2003, Sir George Alleyne, former director of the Pan American Health Organization (PAHO) of the World Health Organization (WHO), observed that, "If one country in the Americas has a high rate of disease, all countries are at risk. But within each country there are vast differentials, between urban and rural, between the inner cities and more affluent areas" (Health and Medicine Week, 2003, p. 396). He goes on to identify research as the tool to help provide rural areas with quality healthcare and improved health outcomes. "Research is relevant to fixing problems of a region, especially when it's rural and you need to get care without dislocating people." (Health and Medicine Week, 2003, p. 397) Understanding that differences in geography have a relationship with differences in health has not gone unnoticed (Arcury, Gesler, & Preisser, 2005; Manious, King, & Garr, 2004; Asthana & Halliday, 2004; Hartley, 2004; Gamm & Hutchinson, 2004; Larson, Hart, & Rosenblatt, 1997; Senior, Williams, & Higgs, 2000). In the United States, rural health is proclaimed to be a national concern; hence, it must be different from urban health and as such, should be approached differently. "In July, [2001] Secretary Thompson announced the creation of a HHS Rural Task Force that will conduct a Department-wide examination of how HHS programs can be strengthened to better serve rural communities" (Centers for Disease Control (CDC), 2001). Also in 2001, the CDC presented research focused on health in the United States as it relates to urban and rural patterns (CDC, 2001). In general, findings from the study that are relevant to rural health are the following: Smoking is more likely in rural teens and adults, absence of health insurance is likely in rural residents, and high mortality rates exist for urban youth and children. Though it is worth noting that for some variables, poor health outcomes exist equally in the most rural and in the most urban places, with suburbia being the healthiest.

It makes sense that some rural health issues are different from those in urban areas. The one glaring aspect is accessibility. Regarding infant health, in many cases rural locations are far removed from the general services of a pregnancy program, such as Healthy Start, to obstetric and gynecological care, to the specialization of a neonatal surgeon or a high-risk nursery should they be required. Granted, in both rural and urban locales the issue of financial accessibility is a common thread, but geographic accessibility may be more of an issue as public transportation is all but nonexistent in rural places, especially in the rural South.

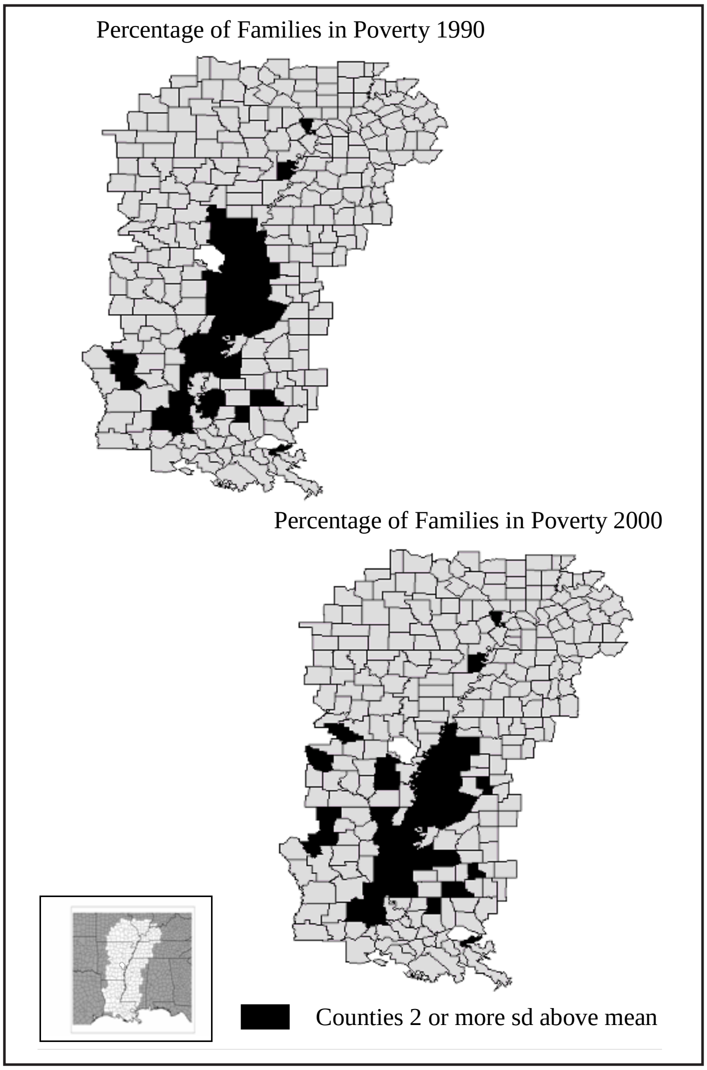
Additionally, in places where agriculture is a dominant land use, exposure to chemicals may also be a relevant issue. In fact, I have listened to a personal account of this case in the Mississippi Delta. A young woman once stated that one of the reasons her mother made sure her daughters left for college and then did not return home to stay, was her observation that many women in the area had difficulty with pregnancy in addition to problems with cancer in early adulthood. The mother believed this situation to be connected to the chemicals applied on the expansive cotton fields in the area. She is not alone in her perception of risk. Since Rachel Carson's 1962 seminal work, *Silent Spring*, the topic of chemicals in the environment and their health ramifications have attracted the attention of researchers. The focus in agricultural areas has been on identifying health risks of pesticide, herbicide, fertilizer, and other chemical applications to those who face an occupational hazard from these activities, but also to the larger rural community (Carson, 1962; Alavanja, Sandler, & McMaster, 1996; Sever, 1995; Efird et al., 2003; Lee, Burnett, Lalich, Cameron, & Sestito, 2002; Uri, Atwood, & Sinabria, 1999; Allen et al., 1997; Zejda, McDuffie, & Dosman, 1993; Mushak & Piver, 1992: Jorgenson, 2004: Sandifer, 1974).

As this book is predominantly about infant mortality and related issues, I feel it is important to note specific rural concerns with the topic. Often, but by no means in every instance, rural areas are also agricultural. Even without agriculture, rural areas face limited access to medical and social services. This limitation becomes further heightened if poverty is combined with limited access (i.e., not owning a car). The Economic Research Service of the United States Department of Agriculture (USDA) provides maps of nonmetropolitan, "persistent poverty" counties. A dominant spatial pattern of these counties is their clustering in the Mississippi Delta and throughout the Cotton Belt of the southern United States and Appalachia. Even using census data on poverty for 2000, this same trend appears (Figure 1). Where plantation agriculture has been a part of the past and the present, poverty is a problem. This general pattern provides a cursory geographic background to distressed rural areas where poverty and limited accessibility to transportation heighten medical and socioeconomic ills.

An Overview of Some Rural Health Issues

Since this book focuses on infant health, it is appropriate that this section begins with an overview of rural infant mortality. I will then introduce two topics not often linked with rural health: rural ghettoes and "agristructure," the structure of agriculture, from commercial to small farms. Rural ghettoes have been identified by the geographer, Charles Aiken, a noted scholar for work on plantations in the South. In general, he has found that distressed neighborhoods, similar to inner-city urban ghettoes, exist in the rural South, specifically in the African American

Figure 1. Rural poverty in the Mississippi Delta



community (Aiken, 1990, 1998). Though his work in this area has not received attention from the medical community, hopefully this chapter will highlight his findings and relate them to understanding rural health environments. In addition, rural sociologist Linda Lobao has written extensively on the relationship between "agristructure" and rural community well-being (Lobao, 1990; Lobao & Schulman, 1991; Lobao, Schulman, & Swanson, 1993; Lobao & Saenz, 2002). Again, this is a literature that should be included in understanding rural places from a health perspective.

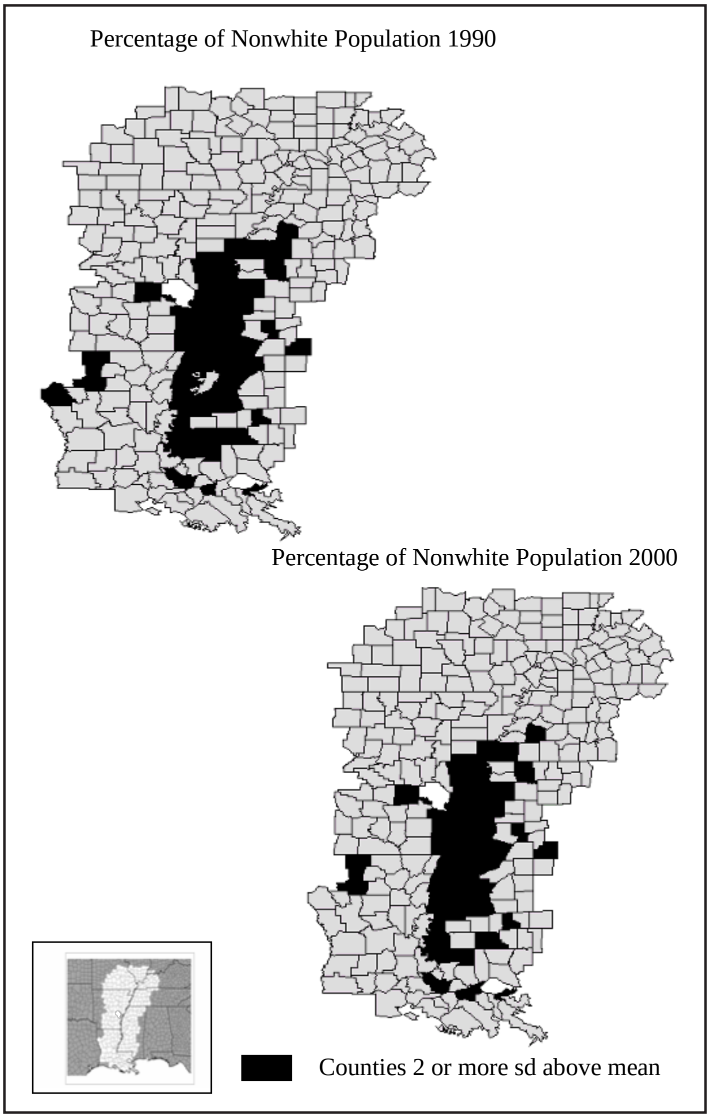
Rural Infant Mortality. Chabot, Garfinkel, and Pratt (1975) find that the degree of urbanization is influential in infant mortality, especially in cases of nonwhite infants in the most rural places. In general, they find that for infants older than one day, the incidence of death increases as urbanization decreases and that racial inequality prevails in the rural areas (as is known to occur in urban areas as well). As the nonwhite infant grows older, the incidence of mortality in rural areas grows along with it. Their study clearly shows the disadvantage to infant health in rural America, particularly in nonwhite populations. These findings lend substantiation to what any resident, observer, or scholar of rural areas with large Native American, Hispanic, or African American populations can tell you — that in many cases the quality of life for these groups in rural areas is bleak. Figure 2 presents the pattern of nonwhite residents in the Delta. Notice the similarity of this pattern with that of poverty in Figure 1. Since this pattern has been established, it begs the question "why." Here again we get into variables of accessibility to any number of services (health care, education, employment opportunities, etc.) that are affected by rural location and sometimes race relations. However, the same factors that impact urban infant mortality are also seen here; they are without regard for urban/rural geographic distinction (teen pregnancy, single-mother head of household, smoking, substance abuse, HIV, crime — yes, rural crime).

In addition, Eisner, Pratt, Hexter, Chabot, and Sayal (1978) find that although infant mortality rates for the nation have improved overall, when comparing these deaths along a continuum from urban to rural, the rural areas still show higher rates than urban areas. Part of the explanation for the geographic inequality is access to high technology for newborns in urban hospitals.

I mention these two brief examples to bring forth a couple of points: First, spatial inequality does exist in infant health from urban to rural places with some of the causes being the same and some being different. Second, both these studies are close to 30 years old, meaning these concepts are not new to us. What is new, however, is that we can investigate these spatial concepts with GIS in order to target areas for funding and more efficiently direct preventative care or treatment to the most needy areas. One of these types of areas in need is the rural ghetto. For those unfamiliar with the topic, spend a day driving through the rural South. The boarded storefronts, graffiti-covered buildings, and people loitering along the mostly abandoned roads describe the rural ghetto, a southern ghost town.

Rural Ghettos. Whatever the definition or connotation the word ghetto brings to mind, the term is associated with urban metropolises. These areas, populated by a racial minority, are characterized by poor quality in safety, health, housing, education, and they are reasonably assumed to be areas where infant mortality

Figure 2. Nonwhite population in the Mississippi Delta



is a problem. However, rural America also possesses similar areas. Aiken (1990, 1998) identifies the existence of rural areas of the South, namely in locations where plantations were or are a part of the economy. These are places where a concentration of African American residents live with the same characteristics as the urban ghettos, complete with federally funded housing projects.

Here is where rural location plays a part in the recognition (or lack of recognition) of this problem. "Small nuclei of poor blacks spread over a large agricultural region hardly have the geographical impact, visually or statistically, of black ghettos in metropolises" (Aiken, 1990, p. 242). Aiken observes that these places are "concealed by spatial distribution." However, given the knowledge of their existence and of the problems that are likely plaguing these areas, those interested in health research in rural areas should be aware of the rural ghetto. Because these places are small, spatial analysis in remote areas should occur at a fine scale, which is contrary to what might be undertaken given the dispersed population over large areas of land.

Agristructure and Chemicals. Another concern to rural areas and rural health is the type of agriculture occurring in the community and the chemicals utilized in its operation. Of course, pollutant-related health problems are occurrences in urban areas, but in rural areas, especially those wherein agriculture is a part of the economy, the types of pollutants differ. The results of exposure are a relevant health concern in the rural areas.

We already know that exposure to certain agricultural chemicals has adverse effects on human health. For example, the National Institutes of Health (NIH) study the effects of these chemicals on reproductive health, certain cancer incidences, pulmonary, and dermatological health (NIH, 1995). In addition, anencephaly and spina bifida occurrence is examined in relation to environmental factors such as agricultural chemicals, among other types of exposures in a search for causality (Sever, 1995). Shaw, Nelson, and Olshan (2002) found paternal exposure to chemicals as a risk factor in infant neural tube defects. They specifically cite farm workers as a risk group. However, because many chemicals are distributed aerially, exposure to areas beyond the fields is plausible.

Related to agricultural chemical use is the type of farming operation (for example, commercial cotton growers need planes to apply chemicals and control of spread may be limited, whereas organic farmers would not be applying chemicals). Farming, though, is not important to rural health only through chemical usage. The type of farm structure that comprises a rural area also has an impact on the local society. Lobao (1990) investigates the relationship between farm structure (including such attributes as average farm size, number of land owners, number of laborers, mechanical and capital investment, etc.) and infant mortality using the general premise of Goldschmidt (1978) that with larger, more corporate types of farms in an area, the socioeconomic well-being of the community is generally low. With the low quality of life, various factors emerge in the populations that are related to incidence of infant mortality and low birth weight. These factors may be direct or more indirect; for example, through corporate agriculture creating economic disparity between those who own and/or manage the land, and those who are no longer employed by it due to mechanization and chemical applications, or general technological progress. A situation of economic inequality develops with “haves” and “have-nots,” those who are a part of and who benefit from this system and those who are relegated to the edges of the fields. With this creation of social inequality come disparities in education and income (Lobao, 1990). More directly, studies (such as Goldschmidt, 1978; Fujimoto, 1977; Swanson, 1980) have identified fewer numbers of community services in areas where large-scale farming is a dominating economic force. With fewer services comes even more restricted access to medical attention than what would exist due only to the distance from an urban center. For her complete argument, see Lobao (1990, pp. 145-157), though reading the whole book is well worth the time.

The need for research into rural health issues has been addressed and various universities and government-sponsored groups do undertake this topic. However, as the spatial aspect of health has gained preeminence (i.e., Healthy People 2010), and as its necessity of study in rural areas has already been addressed, working with spatial data in rural areas presents some different challenges than the same work on health in urban areas. Therefore, it is worth discussing some of the common elements of rural spatial data that any researcher should be aware of and should be cognizant of how they should be addressed.

Hopefully, this section introduced some sources and ideas not usually associated with rural infant mortality in typical medical investigations. They are worth being cognizant of in any spatial investigation into infant health in rural places. After all, the more you know about not only the health outcome under investigation, but the place(s) where it is being studied, the more confidence you have in your findings. However, when employing GIS in these research endeavors, it is essential to understand the specific cautions to rural spatial data, in addition to knowing the medical facts and the relevant geography.

Rural Geography and Dealing With Rural Data

The question of “what is rural?” has yet to be answered in a consistent fashion — the answer will depend on who you talk to. Furthermore, and specifically for spatial research, the delineation of this type of area is not consistent within the U.S. government — again, it depends on which agency you talk to. We have a census designation for rural vs. urban places from the Economic Research Service of the Department of Agriculture. However, its definition is only one of several.

The U.S. Office of Management and Budget (OMB) has traditionally defined metropolitan areas of the country, but since this definition only includes cities of at least 50,000 people and the associated suburban communities, much of the land area of the country is left out. Counties that hold this population are defined as the Metropolitan Area (MA) with the exception of New England, where smaller political units are used instead. Spatially, this definition presents a problem: All of the remaining counties are not really rural. The answer to how you define "rural" places depends on what you are investigating. I suggest taking a look at some of the different classification schemes because they are not created equal. From this point, you can decide what makes sense for your area and particular research question.

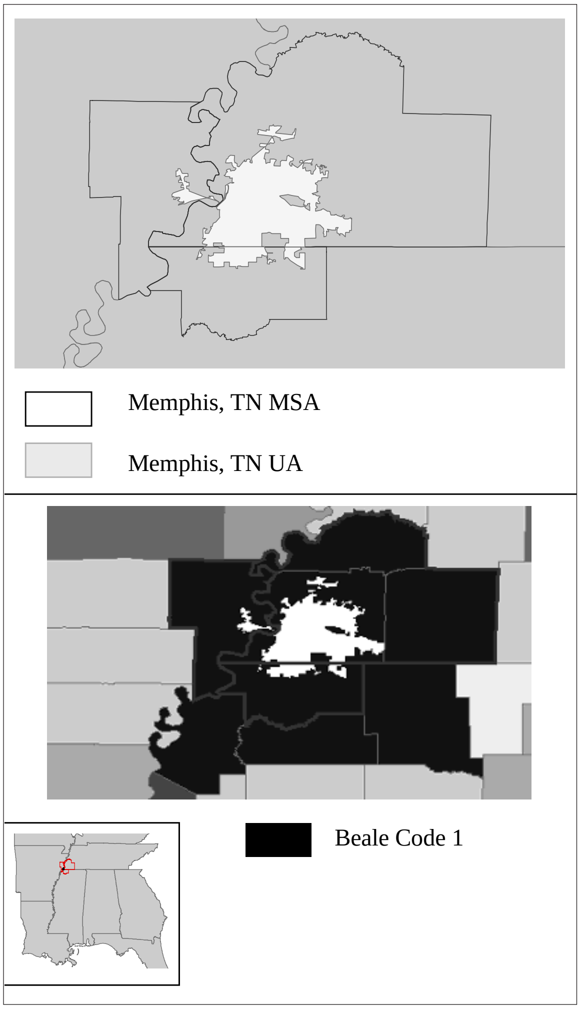
The Economic Research Service (ERS) of the U.S. Department of Agriculture delineates "rural" in more specific terms by providing rural-urban commuting classifications, rural-urban continuum codes to show degrees of urban or "ruralness," and urban influence codes (ERS, 2004).

These classification schemes use the county as the scale of analysis, but what other designations exist at a finer scale? The U.S. Census provides an alternative, albeit one that only takes into account urban areas, and leaves the rest to be rural. To address the overgeneralization of aggregation by counties, ERS has adopted a census method to identify "rurality" and "urbanness" using census tracts. The census considers rural to be any area that is outside an Urban Area (UA) or an Urban Cluster (UC), therefore being defined by what it is not rather than by what it is.

Take, for example, the Metropolitan Statistical Area (MSA) for Memphis, Tennessee. It includes the counties of Shelby, Tipton, and Fayette, Tennessee, DeSoto, Mississippi, and Crittenden, Arkansas. Though much growth has occurred in these counties, driving through all of the counties except Shelby (where Memphis is located) gives the distinct impression that they are not completely urban places. In fact, standing in downtown Memphis and looking over to the Arkansas side of the Mississippi River provides a stark change in landscape—fields, fields, and more fields. This scenery is also quite visible in Tipton, Fayette, and Crittenden counties, yet the counties are considered metropolitan. As shown in Figure 3, use of UA or rural/urban continuum data may provide a more realistic map of the area's geography.

Now that I have presented some background to the ways rural areas can be defined, the more important aspect of how rural data differ from urban data and the relevant aspects of these differences in a GIS can be tackled. Though there are many topics that could be included in this section, I have chosen to focus on representation of aggregated data in rural areas (using the example of Rapides Parish, Louisiana), and rural issues with Topologically Integrated Geographic Encoding and Referencing (TIGER) (this road file has been introduced in a prior chapter). I chose these aspects because in the years I have been utilizing GIS to study rural areas, these just keep popping up. Depending on your research

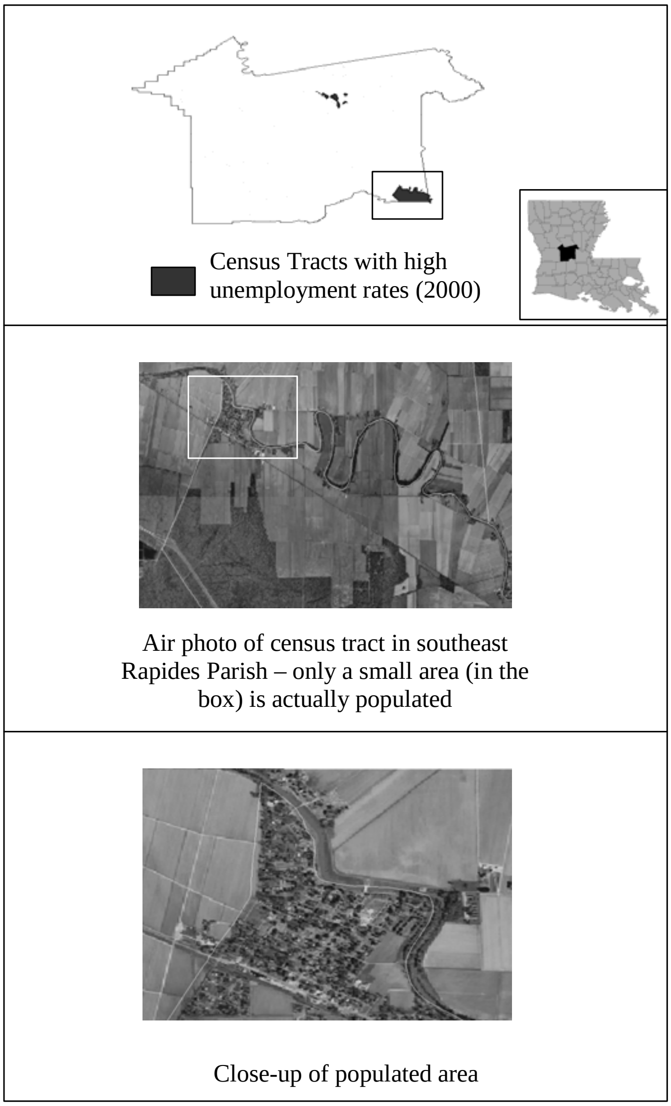
Figure 3. Example of different urban/rural classification schemes



topic, data for rural areas may be comparable to data for urban areas (i.e., availability, reliability). However, rural data also have some characteristics particular to this type of place of which any researcher should be aware. This chapter is not meant to provide an all-inclusive list of any possible problem that might be encountered while working with rural data, but more as a reference for a couple of common challenges I have repeatedly encountered in the area of GIS and some ideas of how to deal with them.

Representation of Aggregated Data in Rural Areas. Though population of a political unit is a primary characteristic used in defining a rural area, it is the population density that poses potential problems (such as Modifiable Area Unit Problem) for the results of spatial analysis. The Modifiable Area Unit Problem (MAUP) is simply that with changing scale or level of aggregation, statistical output will change with it. The MAUP is certainly prevalent in many geographic studies, but rural areas present a special challenge due to low population densities and vast areas without populations (such as fields). A representative example of this situation is Rapides Parish, Louisiana (though many rural counties and parishes could provide similar results). In this case, the variable under investigation is unemployment (Figure 4). In the first map, the unemployment percentage

Figure 4. Example of scale impact on data display



is mapped by census tracts and the areas identified in dark color are +1.5 standard deviations above the mean for unemployment. The tract in the southeast corner shows up as an area for further investigation since it is outside the cluster further to the north. This is a case in which Digital Ortho Quarter Quads (DOQQs), basically an air photo, are a useful source and provide a view we cannot get anywhere else short of flying over the parish. We see in the second image that actually, the vast majority of land in the tract is farmland with a population cluster in the northwest corner. Not only are the resulting unemployment figures not applicable to the entire tract, but in the populated area we may have a case of small numbers making for a high rate (maybe only a few people are unemployed, but the total number of people in the area is small, therefore making a high rate). The final figure zooms to the block group level. The large area of unemployment that is displayed on the first map would not have done so if the scale had been finer. This is a common way you will run into MAUP in rural areas. If you cannot ground truth your data in the field, DOQQs can be a useful alternative (provided they're at a usable resolution for your work and that they are available for a time that is concurrent with the time of collection of your data). This example graphically displays the importance of utilizing as fine a scale of analysis as possible in rural areas. As mentioned in the section on rural ghettos, due to dispersed populations, a fine scale is needed to effectively identify at-risk areas in rural environments.

Rural Issues With TIGER. Incomplete address ranges in the attribute file of TIGER roads is not unusual, even in an urban area. However, I have noticed more missing address range fields in the rural areas wherein I conduct most of my research. Each road file should have columns that represent the address range for the left side of the road segment and for the right side of the road segment. In a TIGER file they are titled FADDR (from address right), TADDR (to address right) and FRADDL (from address left) to TADDL (to address left). Without these numbers in place, the geocoding operation has no way of matching the address you want to see on the map to its proper location. It is not uncommon to face this situation when dealing with sparsely populated areas. This is when you get to be creative, but be careful of your sources. I have four main approaches to finding the proper location for rural addresses, all of which have proved successful in identifying correct address locations: (1) mapquest.com — type in the problematic address(es), zoom into the map that is generated, find the same location on the view of the GIS, heads-up digitize that point into to the shapefile (or layer). (2) DOQQs are useful if you want to double check a suspicious address (you don't think it belongs where the geocoding operation placed it and this does sometimes occur). Add the appropriate DOQQ to your GIS view, add your existing geocoded point file to the view (make sure they are in the same coordinate system — usually Universe Transverse Mercator

(UTM)), and then zoom into the questionable area to see if a building, residence, etc. is truly at this location. One caveat exists, though. You must be sure the date of the DOQQ is proximate to the date of the data you are working with. (3) make a phone call or visit to a knowledgeable source; this technique speaks for itself. (4) hit the road yourself to ground truth the data, if it is relatively close and feasible to do. At this point you can either mark it on a map you take with you, (I like to use USGS quad maps) or take a GPS to attain its coordinates and then upload this information into your GIS.

Another set of problems that arise from rural data in a GIS environment is an abundance of P.O. boxes and the actual offset of street addresses — how far the GIS places a point from the actual location based on road file. P.O. boxes do not represent a residence and therefore are of no value for geocoding. Unfortunately, they are commonly used in rural areas. In these cases it pays to be resourceful — ground truth and heads-up digitize the actual residence if possible. Also, it is not uncommon in rural areas to have the mailbox at the roadside, which may be quite a distance from the actual dwelling. In these cases, and depending on how representative your information must be, DOQQs are a way to verify actual location of a residence.

Conclusion

Infant mortality is problematic in both urban and rural environments, and though some of the causal factors are indiscriminate of geography, others are related to location. People, especially poor populations, face particular risks in rural places owing to inadequate access to transportation, limited local medical services, rural ghettos, agristructure, and agricultural chemical use.

A secondary concern is defining rural areas for research purposes. To date, the existing government definitions show the dominance of the concept of urban, with rural being what is left over. The coarse scale of these classification schemes does little for local GIS analyses and therefore is of limited use. Unfortunately, this perspective of focusing on urban places, while leaving rural places to be defined by what they are not rather then by what they are, has carried over into policy issues and public attention. Rural areas have been described as catatonic, boring, disconnected, and slow — just to name a few adjectives. However, any scholar of rural America knows that they are none of these things, but are a part of the processes that impact any place — urban or rural or anywhere in between.

References

Aiken, C. S. (1990). A new type of black ghetto in the plantation south. *Annals of the Association of American Geographers, 80*, 223-246.

Aiken, C. S. (1998). *The cotton plantation south since the Civil War*. Baltimore: Johns Hopkins University Press.

Alavanja, M. C. R., Sandler, D. P., &McMaster, S. B. (1996). The agricultural health study. *Environmental Health Perspectives, 104*, 362-369.

Allen, R. H., Gottlieb, M., Clute, E., Pongsiri, M. J., Sherman, J., &Obrams, G. I. (1997). Breast cancer and pesticides in Hawaii: The need for further study. *Environmental Health Perspectives, 105*, 679-683.

Arcury, T. A., Gesler, W. M., &Preisser, J. S. (2005). The effects of geography and spatial behavior on health care utilization among the residents of a rural region. *Health Services Research, 40*, 135-155.

Asthana, S., & Halliday, J. (2004). What can rural agencies do to address the additional costs of rural services? A typology of rural service innovation. *Health & Social Care In The Community, 12*, 457-465.

Carson, R. (1962). *Silent spring*. Boston: Houghton Mifflin.

CDC (Centers for Disease Control). (2001). *HHS issues report on community health in rural, urban areas*. Retrieved March 25, 2005, from http://www.cdc.gov/ nchs/pressroom/01news/hus01.htm

Census 2000 Population Statistics. United States Department of transportation, Federal Highway Administration. Retrieved August 15, 2004, from http://www.fhwa.dot.gov/planning/census/cps2k.htm

Chabot, M. J., Garfinkel, J., & Pratt, M. W. (1975). Urbanization and differentials in white and nonwhite infant mortality. *Pediatrics, 56*, 777-781.

Christenson, J., Maurer, R. C., & Strang, N. L. (1994). *Rural data, people, and policy: Information systems for the 21st century*. Boulder, CO: Westview Press.

Conrad, C. (1991). Health care in rural America: The congressional perspective. *Public Administration Review, 51,* 461-465.

Delatorre, A., Fickenscher, K., Luft, H. (1991). The zip (postal) code difference — Methods to improve the identification of rural subgroups. *Agricultural Economics, 5*, 253-262.

Economic Research Service (ERS). (2004). *Measuring rurality: Rural-urban commuting area codes*. U.S. Department of Agriculture. Retrieved August 11, 2004, from http://www.ers.usda.gov/Briefing/Rurality/RuralUrbanCommutingAreas/

Efird J. T., Holly, E. A., Preston-Martin, S., Mueller, B. A., Lubin, F., Filippini, G., et al. (2003). Farm-related exposures and childhood brain tumours in seven countries: Results from the SEARCH International Brain Tumour Study. *Paediatric and Perinatal Epidemiology, 17*, 201-211.

Eisner, V., Pratt, M. W., Hexter, A., Chabot, M. J., & Sayal, N. (1978). Improvement in infant and perinatal mortality in the United States, 1965-1973: I. Priorities for intervention. *American Journal of Public Health, 68*, 359-364.

ERS. (2005). *Measuring rurality*. U.S. Department of Agriculture. Retrieved March 24, 2005, from http://www.ers.usda.gov/Briefing/Rurality/

Fujimoto, I. (1977). The communities of the San Joaquin Valley: The relation between scale of farming, water use, and quality of life. In U.S. Congress, House of Representatives, Obstacles to Strengthening the Family Farm System. Hearing before the Subcommittee on Family Farms, Rural Development, and Special Studies of the Committee on Agriculture, 95th Congress, first session (pp. 480-500). Washington, DC: U.S. Government Printing Office.

Gamm, L., & Hutchison, L. (2004). Rural Healthy People 2010 — Evolving interactive practice. *American Journal of Public Health, 94*, 1711-1712.

Garreau, J. (1991). *Edge city: Life on the new frontier*. New York: Doubleday.

Goldschmidt, W. (1978). *As you sow: Three studies in the social consequences of agribusiness*. Montclair, NJ: Allanheld, Osum, and Company.

Hart, J. F. (1991). The perimetropolitan bow wave. *Geographical Review, 18*, 35-51.

Hartley, D. (2004). Rural health disparities, population health, and rural culture. *American Journal of Public Health, 94*, 1675-1678.

Health and Medicine Week. (2003, November 11). Former world health leader says rural U.S. needs better care [Healthcare Access section], 396-397.

Jorgenson, A. K. 2004. Global inequality, water pollution, and *infant mortality. Social Science Journal, 41*, 279-288.

Larson, E. H., Hart, L. G., & Rosenblatt, R. A. (1997). Is non-metropolitan residence a risk factor for poor birth outcome in the U.S? *Social Science and Medicine, 45*, 171-188.

Lee, E., Burnett, C. A., Lalich, N., Cameron, L. L., & Sestito, J. P. (2002). Proportionate mortality of crop and livestock farmers in the United States, 1984-1993. *American Journal of Industrial Medicine, 42*, 410-420.

Lobao, L. M. (1990). *Locality and inequality: Farm and industry structure and socioeconomic conditions*. Albany, NY: State University of New York Press.

Lobao, L., & Saenz, R. (2002). Spatial inequality and diversity as an emerging research area. *Rural Sociology, 67*, 497-511.

Lobao, L. M., & Schulman, M. D. (1991). Farming pattern, rural restructuring, and poverty: A comparative regional analysis. *Rural Sociology, 56*, 565-602.

Lobao, L. M., Schulman, M. D., & Swanson, L. E. (1993). Still going: Recent debates on the Goldschmidt hypothesis. *Rural Sociology, 58*, 277-288.

Mainous, A.G., King, D. E., & Garr, D. R. (2004). Race rural residence, and control of diabetes and hypertension. *Annals of Family Medicine, 2*, 563-568.

Morrill, R., Cromartie, J., & Hart, L. G. (1999). Metropolitan, urban, and rural commuting areas: Toward a better depiction of the U.S. settlement system. *Urban Geography, 20*, 727-748.

Mushak, E. W., &Piver, W. T. (1992). Agricultural chemical utilization and human health. *Environmental Health Perspectives, 97*, 269-274.

National Institutes of Health (NIH). (1995). Protecting the people who feed us. NIEHS News. *Environmental Health Perspectives, 103*(5), May. Retrieved August 15, 2004, from http://ehp.niehs.nih.gov/ docs/1995/103-5/NIEHSnews.html

Salant, P., & Walker, A. J. (1995). *Guide to rural data* (rev. ed.). Publication of the Rural Economic Policy Program of the Aspen Institute. Washington, DC: Island Press.

Sandifer, S., Hope. (1974). Industrial and agricultural chemicals. *Pediatrics, 53*, 843-844.

Senior, M., Williams, H., & Higgs, G. (2000). Urban-rural mortality differentials: Controlling for material deprivation. *Social Science and Medicine, 51*, 289-305.

Sever, L. E. (1995). Looking for causes for neural-tube defects: Where does the environment fit in? *Environmental Health Perspectives, 103*, 165-171.

Shaw, G. M., Nelson, V., & Olshan, A. F. (2002). Paternal occupational group and risk of offspring with neural tube defects. *Paediatric and Perinatal Epidemiology, 16*, 328-333.

Swanson, L. (1980). *A study in socioeconomic development: Changing farm structure and rural community decline in the context of technological transformation of American agriculture*. Unpublished doctoral dissertation, University of Nebraska, Lincoln.

United States Department of Agriculture (USDA). (2005). National Agricultural Library, Rural Information Center. *What is rural? Defining rural,*

rural values, and rural character. Retrieved March 24, 2005, from http://www.nal.usda.gov/ric/faqs/ruralfaq.htm

Uri, N. D., Atwood, J. D., & Sanabria, J. (1999). The environmental benefits and costs of conservation tillage. *Environmental Geology, 38*, 111-125.

Zejda, J. E., McDuffie, H. H., & Dosman, J. A. (1993). Epidemiology of health and safety risks in agriculture and related industries: Practical applications for rural physicians. *Western Journal of Medicine, 158*, 56-63.

About the Authors

Andrew Curtis, PhD, is an assistant professor in the Department of Geography and Anthropology at Louisiana State University. He is also director of the World Health Organization's Collaborating Center for Remote Sensing and GIS for Public Health. This center currently houses several local, national, and international projects involving GIS and the understanding of disease systems. Topics range from anthrax distributions in Kazakhstan to the cultural surface of Chagas' disease in Mexico. His personal research foci include real-time GIS analysis of emerging infections, the spatial analysis of historic epidemics, and GIS response to bioterrorism. In addition, he has worked for 4 years on a project to reduce African American infant mortality in Baton Rouge, Louisiana. He is the evaluator (and original co-grant writer) of the Baton Rouge Healthy Start project.

Michael Leitner is an assistant professor in the Department of Geography and Anthropology, Louisiana State University (LSU) in Baton Rouge. He received his undergraduate (1987) and graduate (1990) degrees in geography and cartography at the University of Vienna, Austria, followed by his PhD in GIS and computer cartography in the Department of Geography at the State University of New York at Buffalo in 1997. His research interests fall into the area of geographic information science and technology (GIS&T) applied to criminology, medical geography, and cartographic visualization. He has taught primarily GIS&T courses, including computer cartography, map design, GIS, GPS, aerial photo interpretation, and spatial analysis. Dr. Leitner is the current chair of the Cartography Specialty Group of the Association of American Geographers and a member of the editorial boards of *Cartographic Perspectives* and *Cartography and Geographic Information Science*.

***

Jacqueline W. Mills, PhD is an instructor in the Department of Geography and Anthropology at Louisiana State University. She has recently completed her PhD and will be attending Tulane Law School in fall 2005. Her research interests are rural and agricultural geography, GIS, policy analysis, social inequality, and social and environmental justice. She has served in various GIS and research positions for the state governments of Tennessee and Louisiana. For the past 6 years, she has conducted research on several projects focused on the Yazoo-Mississippi Delta region and in rural Georgia.

Index

Symbols

1:100,000 and 1:250,000 Scanned Topographic Maps 33
1:24,000 Scanned Topographic Maps (DRGs) 33
4 P's Plus 147

A

A Civil Action 17, 103
absolute clusters 180
absolute hot spot 180
abuse 9, 273
access to medical care 275
accessibility 289
activity space 271
adaptive bandwidth 187
adaptive kernel estimation 186
address matching 24
aerial photo 33
African Americans 5, 40, 85, 245, 273
age-appropriate immunizations 253
aggregating into spatial units 37
aggregation effects 153
agricultural structure 288
agristructure 288, 294
air photos 54
air pollution 82
Air Quality Index in 2003 107
airborne exposures 104
alcohol 82
alcohol use 276
alcohol-related birth defects (ARBD) 86
alcohol-related neurodevelopmental disorders (ARND) 86
all-hazards preparedness 269
American College of Obstetricians and Gynecologists 44
American Heart Association 104
American Hollow 96
analysis of variance (ANOVA) 64, 193
analyze 21, 52
anencephaly 294
anthrax 269

anxiety 86, 274
APB Neighborhood Crime Check ratings 252
applied geography 246
area analysis 38
area association techniques 225
area samples 150
areal symbolization scheme 57
areas 21
aspatial queries 42, 248
Assessment System for Population Exposure Nationwide (ASPEN) 105
asthma 9, 69, 80, 104
asthma levels 97
at-risk population 180, 187
atherosclerosis 80
attribute confidentiality 226
attribute data 238
audience 54

B

bacterial vaginosis (BV) 85
bandwidth 184
basic summary statistics 63
Baton Rouge 272
Baton Rouge Healthy Start 83, 149, 182, 203, 209, 245, 278
Baton Rouge Healthy Start GIS 28, 117
Baton Rouge Healthy Start Region 53
behavioral characteristics 69
behavioral risks 9, 80, 82
behavioral traits 5
bias 149, 277
binary contiguity 175
binary contiguity matrix 175
binary logit regression 194
biological pathway 80
biological risks 9
biophysical risk 271
bioterrorism 224
bioterrorist (BT) 268
birth certificate 25, 53
birth certificate data 203, 248
birth residence 25
births to teenagers 56
black lung 103
boll weevil 273
Boolean expression 41
Breadline Britain poverty index 43
break point 67
breaks chosen 53
Broad Street Pump 55
buffers 37
busy road 11

C

CADGIS lab 35, 262
cadmium 106
CAL 68
Calculating Deprivation Indexes 42
cancer alley 17
cancer morbidity rate 110
cancer mortality 107
CAP Index 252
carbon monoxide (CO) 106
Carcinogenic Potency Database (CPDB) 107
Carstairs Deprivation Index 43
Cartesian coordinate 151
cartography 52
caseworker 54, 83, 216, 279
CDC 45
census block 25, 182
census block group 182
Census Bureau 229
census tract 25, 182, 225
census tract boundaries 258
census tracts 57, 153, 179, 296
centroid (center of gravity) 30, 240
centroid method 183
Chechnya 152
chemical applications 294
chemicals 288, 294
Chernobyl 103
child care 257
child development problems 102
childhood leukemia 108
children hit by automobiles 102
cholera 15, 154
choropleth 182
choropleth map 37, 52, 57, 225
chronic pollution exposure 104

chronic stresses 97
city infrastructure 104
CityMatCH 45, 246
classification 39
classification scheme 182
"click" boxes 263
client 263
clumping techniques 181
cluster 26, 153, 179, 226, 231, 260
cluster analysis methods 180
cluster techniques 184
clustered pattern 226
clustering 174
Coastwide 2001 Scanned Images 33
cockroach allergens 115
cockroach excretia 102
cohort or social risks 82
cold spots 174, 178
colloquial names 55
colors 53
Committee to Study the Prevention of Low Birth 84
common dot map 60
community advisory boards (CAB) 95
community alert system 68
community facilities 103
community health 21
community voices 281
complete spatial random distribution 183
complete spatial randomness (CSR) 175
complex physical processes 210
complex social processes 210
confidentiality 27
confirmatory analyses 66
confounders 71
confounding by indication 86
congenital malformations 2, 103
contaminated groundwater 103
contextual modeling 38
contextual or multilevel models 71
contiguity 175
contiguity-based matrices 175
contours 38
coordinates 154
coping mechanism 9, 80
crime 10, 105, 252
crime hot spots 80
crime rates 66
CrimeStat® 64
cross-tabs 64
crowded central city areas 272
crowding 98
cultural risks 9
culturally sensitive delivery 274
Cumulative Exposure Project 105

D

dasymetric map 58
data 21
data analysis 153
data classification 53
data forms 263
data reduction 41
data suppression 239
database 256
date stamps 263
death certificate 25, 53
death certificate data 203, 248
degree of poverty 42
demographic information 248
demographic vulnerability data 278
denominator 7, 153
density 154
density analysis 67, 157
density techniques 181
Department of Geography and Anthropology at LSU 110, 248
departments of geography 281
depression 273
descriptive statistics 63
diet 82
difference of proportions t-test 211
difference of proportions test 107, 154, 160
different patterns 153
Digital Orthophoto Quarter Quadrangles (DOQQs) 33, 151, 299
diphenyl-trichloroethane (DDT) 106
disaster 268
disclosure limitation 229
disease ecology 9
disease ecology approach 81

disparity 5
dispersion models 112
disproportional stratified sample 152
distance decay 159
distance-based matrices 175
distance-based spatial weights matrices 175
DMAP 183, 257
DMAP program 111
domestic risks 83
domestic violence 89, 251
dot map 15, 38
“dread risk” 270
drug offences 80
dual kernel density estimates 185
dual kernel density method 181

E

Early Head Start Centers 216, 266
East Baton Rouge Parish 58, 206, 241, 248
East Baton Rouge Parish Office of Homeland Security 68
ecological fallacy 39, 58
ecological studies 38, 71, 82, 99
economic conditions 271
economic disparity 97
Economic Research Service (ERS) 296
economic risks 9
economic traits 5
Edinburgh Depression 117
education 66
eliminating disparities in perinatal health 56
Eliminating Disparities in Perinatal Health program 247
eliminating racial disparities 96
Emergency Operation Center (EOC) 269
Emergency Planning and Community Right-to-Know Act 107
enterprise GIS 68
enumeration units 32
environmental factors 294
environmental hazards 85
environmental injustice 85
environmental justice 106
environmental pollution 103
environmental risks 9, 80, 82
environmental tobacco smoke 80
environmental toxins 275
environmental traits 5
EOC Enterprise GIS 280
EPA’s toxic release inventory 105
equal steps 59
Erin Brockovich 17, 103
ethics 282
European Commission (EC) 33
evacuation planning 271
evaluation 41, 255
evaluation meetings 245
exploratory analyses 66, 146
exploratory spatial analysis 55
exploratory spatial analysis maps 14
exploratory spatial data analysis 194
exposure to toxins 103
extreme outlier 191

F

facial anomalies 86
Family Planning, Access, Care, Treatment (Family PACT) 93
Family Readiness Guide for Bioterrorism Preparedness and Emergency Response 275
Family Road 33, 247
Family Road of Greater Baton Rouge 253
farm structure 294
farm workers 294
Farrell Jones 262
farthest neighbor 183
fatherhood program 41
federal intervention 269
Federal Poverty Level 251
feeling abandoned 276
feeling overwhelmed 276
FEMA 272
female as head of household 44
fertilizer 290
Fetal Alcohol Syndrome (FAS) 86
fetal and infant mortality review (FIMR) 27, 113, 245
fetal and infant mortality review board meetings 210

fetal-infant mortality 45
fewer retail choices 101
first-order clusters 181, 183
fixed bandwidth 187
flagging datasheets 263
Four and 4P's Plus 92
fragmented city of social classes 279
Freedom of Information Act 229

G

Geary's C 175
generational "stress" cycles 95
geocode 53
geocoding 18, 24, 258
GeoDa™ 65
geographic aggregation 53
Geographic Analysis Machine (GAM) 66, 109
geographic inequality 292
geographic information science (GISc) 17
geographic information systems (GIS) 18
geographic masking strategies 230
geographic scales 2
geographic techniques 1
geographically masking 230
geographically weighted regression (GWR) 65, 148, 193
geography 1
georegistered 154
Germ Theory 14
GIS 248
GIS analysis 203, 246
GIS environment 287
GIS/spatial analysis interface 63
global geographic masking 230
global masking methods 232
global Moran's I 178
global OLS regression 195
global positioning system (GPS) 18, 24
global spatial autocorrelation 175
global spatial regression model 194
globally masked point patterns 232
Goldschmidt 294
graduated colors 52
ground truth 299
group averages 183
GWR model 194
GWR software 197

H

hazardous waste site 104
hazards 268
heads-up digitizing 23, 154
Health Insurance Portability and Accountability Act of 1996 (HIPAA) 55, 227
Health Resources and Services Administration (HRSA) 247
Healthy People 2010 18, 28, 278
Healthy Start 7, 189, 289
Healthy Start GIS 277
herbicide 290
hereditary 96
HHS Privacy Rule 238
hierarchical techniques 181
high-resolution air photography 58
"hillbilly" communities in Appalachian Eastern Kentucky 95
HIPPA requirements 281
HIV/AIDS 17, 252
HIV/AIDS screening/counseling 257
hoaxes 270
holistic neighborhood investigations 72
home 273
Homeland Security 28, 268
homelessness 251
hot spots 9, 23, 53, 103, 148, 174, 178, 231, 257
Hugh-Jones, Martin 247
hurricane 272
hypothesis 41
hypothesis testing 146

I

ICD 250
inadequate intrauterine growth 84
inadequate prenatal care 147, 250
index of local deprivation 43
index of social deprivation 33
indigent and/or minority populations 68, 98
indigent pregnant women 271

individual confidentiality 224
individual privacy 225
individual risk 271
individual's "place" 69
industrial emissions 104
industrial plant 11
industrial polluters 12
infant death 82
infant health 45
infant mortality 23, 52, 147, 205, 257, 276
infant mortality clusters 66
infant mortality rate 1, 249
infant mortality surfaces 258
infantile colic 88
influenza 278
information system 54, 260
infrastructure deficiencies 100
inner city 5
inner-city neighborhoods 101
"insults" 2
International Cartographic Association (ICA) 57
Internet mapping engine 30
interpolation 174, 182
interpolation technique 184
inverse distance weights 175
isarithmic (isoline) Map 60
ISCST3 (Industrial Source Complex Model) 113
isolines 206

J

Jamie Roques 247
John Snow 15, 54, 147, 154

K

K-means partitioning technique 181
kernel 158
kernel density 66
kernel density analysis 103, 149
kernel density estimates 185
kernel density interpolation 174
kernel function 184
Kessner Index 44
Kotelchuck Index 44

L

LACHIP 253
lack of community spaces 102
lack of transport 272
Lalee's Kin: The Legacy of Cotton 96
land fills 82, 103
laptop 264
latitude 25, 29
law of small numbers 6, 204
lead exposure 11, 69, 97, 104
legend 55
life stress hypothesis 114
limited access 290
limited distance functions 187
line samples 150
linear buffer 107
linear regression analysis 193
lines 21
"linked" 26
liquor availability 66
LISA map 178
little family support 272
lives with a smoker 118
local geographic masking 230
local geographic masking methods 232
local indicators of spatial association (LISA) 178
local measures of spatial autocorrelation 181
local Moran's statistic 178
local regression coefficients 194
local regression model 193
local SA 175
local spatial autocorrelation 178
local spatial regression model 194
locally masked maps 234
location allocation routines 67
locational confidentiality 238
logistic regression 194
longitude 25, 29
longitudinal analysis 73
longitudinal birth data 109
Louisiana State University 245
Love Canal 103
low birth weight 2, 39, 56, 66, 147, 155, 250

low maternal weight gain 84
low prepregnancy weight 84
low-birth-weight babies 82
low-birth-weight deliveries 176, 276
lower educational attainment 272
lymphoid leukemia 108

M

macro (city) 271
management information system 68
map classification 2
Mapping and Analysis for Public Safety (MAPS) 229
mapping of vulnerable populations 268
masked point pattern 232
Maternal and Child Health Bureau Division 56
maternal care 45
Maternal Health and Prematurity Group 45
maternal nutrition 88
maximum 63
mean 63
mechanization 294
median clusters 183
mediators 71
Medicaid 8, 99, 245
Medicaid eligibility 253
medical geography 147
medical risks 82
meningitis 272
mental health 252
mental retardation 86
Methyl Parathion (MP) 109, 273
Metropolitan Area (MA) 296
metropolitan areas 32
micro (neighborhood and individual) 271
microcephaly 86
midpoint (center of gravity/center of largest city) 61
mild outliers 191
minimum 63
minimum error 183
missing address range fields 299
mitigation 269
mobile population 205
Modifiable Area Unit Problem (MAUP) 39, 67, 153, 297
Monte Carlo approach 66
Monte Carlo simulation 111, 182
month prenatal care began 214
Moran's I 175
Moran's I statistic 176
Morrison, Dena 247
mortality rate 110
mother medical risk factors 257
mother mobility 205
Mr. Willis of Ohio 149
multilevel modeling 38, 91
multiple linear regression model 192
multiple regression 66
multiple regression analysis 192
multivariate logistic GWR model 194

N

National Ambient Air Quality Standards (NAAQS) 107
national conferences 54
National Healthy Start Association 247
National Institutes of Health (NIH) 294
National Priorities List (NPL) site 104
natural breaks 59
nearest neighbor hierarchical clustering 174, 183
nearest neighbor method 183
near real time GIS 28
negative birth outcomes 271
negative exponential 185
negative global SA 178
negative SA 175
neighborhood 52, 268, 271
neighborhood categorization 209
neighborhood comparisons 71
neighborhood effects 69
neighborhood profiling 71, 260
neighborhood risks 2, 79, 80, 82, 99
neighborhood stressors 262
neighborhood support structure 273
neurotoxicity hypothesis 105
New Orleans 272
newborn care 45
NGO 152
NHSA 247
nitrogen dioxide (NO2) 106

NNHC method 184
no prenatal care (NPC) 208
"no prenatal" information system 254
no prenatal visit 92
nonrandom spatial distributions 107
nonspatial outliers 192
normal kernel function 185
normality 193
north arrow 55
nuclear facilities 108
numerators 153
nutrition 10
nutrition counseling 257
nutritious food 84

O

obesity 10, 69
Office of Public Health (OPH) 224, 248
office-based intervention 88
Oklahoma City 269
one-sample Kolmogorov-Smirnov test 193
Openshaw 108
ordinary least squares (OLS) regression 194
outbreak 270
outliers 66, 191
outreach teams 254
overcrowding 98
overhead power lines 104
overlays 14, 154

P

P.O. boxes 300
particulate pollution 108
partitioning techniques 181
paternal influence 97
patient confidentiality 224, 281
PDAs 28, 264
pediatric accident rates 103
percentages 58
perinatal health care delivery system 253
perinatal mortality 2
perinatal periods of risk (PPOR) 45
pesticide 290
pin map 55
point locations 180
point pattern map 55
point pattern techniques 66, 225
point samples 150
point source of pollution 205
point-in-polygon 58
points 21
police discrimination 101
political conditions 271
political economy approach 79
political fabric 271
political neighborhood 9
political voice 103, 106
polluted water 104
pollution emitters 97
poor 245
population density 297
positional accuracy 29
positive global SA 178
positive SA 175
poverty 7, 70, 245, 251, 290
poverty trap 106, 273
PRAMS 22
pre- and post-growth retardation 86
pregnancy histories 203
pregnancy-related risks 79
prenatal care 8
prenatal risk hot spots 80
prenatal visits 69
Presidential directive #8 269
preterm delivery 82, 155
Privacy Rule 227
profile 256
profiling 208
program participant 263, 278
program participant mobility 73
programming 95
project area 248
project area committee meetings 245
proportional (graduated) point symbol map 61
proportionally stratified samples 151
proximity analysis 68
proximity to toxic locations 271
psychological distresses 97
psychological impact 270
psychosocial services 257

Q

qualitative approaches 80
qualitative data 55
quantiles 59
quantitative approaches 80
quantitative data 55
quarantine 270
quartic 185
queen contiguity 175
query 41

R

race 82
racial conditions 271
racial disparities 249
racial problems 5
racial residential segregation 98
racial segregation 70
racism 97
range 63, 67
Raster GIS 21
rate surface 153
rates 58
recovery 269
redlining 11
reductionism 94
regional patterns 2
regression model 192
relative clusters 180
relocation 270
remotely sensed images 58
reports 262
residence 204
residuals 66
resource deprivation index 43
respiratory problems 9
response 269
response and recovery strategies 273
risk 245
risk factors 79
risk for future pregnancies 86
risk-adjusted nearest neighbor hierarchical clustering 181
risk-based techniques 181
risk-mapping 98
rook contiguity 175
rural 287
rural Delta 95
rural ghettoes 288, 292
rural health 289
rural infant mortality 292
rural routes 29
rural-based clinics 91
“rurality” 296

S

Salmonellosis 269
sample 42
sample strategy 148
SARS 269
SaTScan™ 65
scale 22, 55, 153, 260, 294
scanned in 154
school districts 32
screening 112
Seascale 108
second trimester 8, 155
second-order clusters 181, 183
secondhand smoke 9, 275
secondhand smoking 80
secure Web site 256
secure wireless uploads 264
Sellafield 108
separation 273
September 11th 268
service areas 67
sexual predators 102
short gestation 250
short gestation births 2, 216
short periods between pregnancy 88
significance map 178
single kernel density estimation 187
single kernel density map 187
single kernel density method 181
smallpox 270
smoking 9, 42, 80
smoking cessation programs 80
“smoothed” out 4
smoothed surface 38
social conditions 271
social deprivation 10, 84

social environment 79
social fragmentation 205
social justice 85, 106
social pathway 80
social relevance of actions 79
social risk 9, 80, 271
social variables 69
socially defined neighborhood 205
socially hereditary 96
socioeconomic status 82
socioeconomic data 251
socioeconomic information 25
socioeconomic status of women 88
South Carolina Community Assessment Network 22
Southern Mississippi 109, 273
Southern Regional Climate Center 113
spatial (locational) confidentiality 227
spatial aggregations 57
spatial analysis 62, 146, 248, 256, 294
spatial analyst 64
spatial association 174, 175, 188
spatial association techniques 180
spatial autocorrelation (SA) 174, 193
spatial cluster 279
spatial cohort 98
spatial data 21
spatial decision support system 279
spatial dependency 175
spatial design 148
spatial filter 55, 66, 154, 174, 182, 225
spatial filter analysis 206, 241, 248, 257
spatial filtering (DMAP) 182
spatial fuzzy mode 180
spatial inequality 292
spatial join 38
spatial mode 180
spatial or locational confidentiality 227, 237
spatial outliers 178, 188
spatial pattern 82, 154, 257
spatial queries 42, 248
spatial regression analysis 174, 188, 193
spatial regression model 193
spatial sampling 148
spatial science 70
spatial stability 204
spatially precise targeting 208
spatially specific credo 115
spatially systematic 151
Special Supplemental Nutrition Program for Women, 90
spheres-of-influence 163
spina bifida 294
standard deviation 39, 63, 153
standard deviational ellipses 184
standardized clusters 180
Standards for Privacy of Individually Identifiable Health Information (Privacy Rule) 228
statistical (attribute) confidentiality 226
statistical confidentiality 226
statistical data 238
STD 118
stepwise multiple regression 192
stillbirths 88
stress 11, 66, 80, 274
strong family and religious influence 89
structural functionalism 9, 79
substance abuse 82, 87, 252
substance abuse treatment and counseling 257
sudden infant death syndrome (SIDS) 2, 56, 210, 250
sum 63
Summary Files (SF) numbers 1 and 2 31
Summary Files 3 and 4 31
Super Size Me 90
support network 98
suppression 239
surveillance data 25
symbol form 61
symmetrical kernel function 184
syndromic surveillance 268, 277

T

t-test 160
teenage pregnancies 215
temporal patterns 82
temporal stability 73, 208
thematic maps 57
theoretical research 246
Thiessen polygons 240

third-order clusters 183
Three Mile Island 103
threshold 67
TIGER 29, 296, 299
tobacco use 276
Tobler's First Law of Geography 174
Tobler's Laws of Geography 15, 67
traffic accidents 12
traffic analysis zones 32
transect 151
transport 10, 246
transport-related emissions 104
transportation 257
transportation planning of hazardous material 271
traverse 151
triangular 185
typhoid epidemic 15

U

U.S. Census 57, 228, 229, 296
U.S. Census Bureau 31
U.S. Department of Agriculture 296
U.S. Department of Health and Human Services 227, 247, 276
U.S. Department of Justice 228, 229
U.S. Office of Management and Budget (OMB) 295
UK Index of Multiple Deprivation 43
under-18 pregnancies 169
uniform 185
universal transverse mercator (UTM) 34, 151
unlimited distance function 185
unmasked point patterns 232
unmatched addresses 29
unsafe neighborhoods 97
urban area (UA) 296
urban cluster (UC) 296
"urbanness" 296

V

vaccination 278
vaccines 276
validation 279
validation checks 263
variance 63, 204
vector GIS 21
very low birth weight 45
very low-birth-weight babies 191
very preterm delivery 155
violent crime 80
visualization 153
visualization of disease 14
visualize 21, 52
vital statistics data 25, 53
vulnerability 101
vulnerability analysis 271

W

walk-in pregnancy testing 253
weaponized 270
weighted 159
West Wing 149
window 158
Woburn, Massachusetts 108
Women Infant and Children (WIC) 253
World Health Organization 110, 247
World Trade Center 268
worry 276
www.census.gov 248
www.geographynetwork.com 35, 248

Y

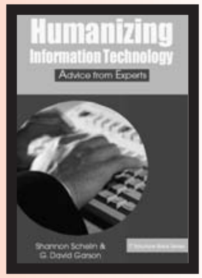
young maternal age 88
YWCA 265

Z

zip code 24, 53, 153, 182, 189, 203, 228, 248, 258
zip code area 149, 188
zone 39

IT Solutions Series

Humanizing Information Technology: Advice from Experts



Authored by:
Shannon Schelin, PhD, North Carolina State University, USA
G. David Garson, PhD, North Carolina State University, USA

With the alarming rate of information technology changes over the past two decades, it is not unexpected that there is an evolution of the human side of IT that has forced many organizations to rethink their strategies in dealing with the human side of IT. People, just like computers, are main components of any information systems. And just as successful organizations must be willing to upgrade their equipment and facilities, they must also be alert to changing their viewpoints on various aspects of human behavior. New and emerging technologies result in human behavior responses, which must be addressed with a view toward developing better theories about people and IT. This book brings out a variety of views expressed by practitioners from corporate and public settings offer their experiences in dealing with the human byproduct of IT.

ISBN 1-59140-245-X (s/c) • US$29.95 • eISBN 1-59140-246-8 • 186 pages • Copyright © 2004

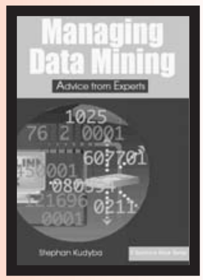
Information Technology Security: Advice from Experts

Edited by:
Lawrence Oliva, PhD, Intelligent Decisions LLC, USA

As the value of the information portfolio has increased, IT security has changed from a product focus to a business management process. Today, IT security is not just about controlling internal access to data and systems but managing a portfolio of services including wireless networks, cyberterrorism protection and business continuity planning in case of disaster. With this new perspective, the role of IT executives has changed from protecting against external threats to building trusted security infrastructures linked to business processes driving financial returns. As technology continues to expand in complexity, databases increase in value, and as information privacy liability broadens exponentially, security processes developed during the last century will not work. IT leaders must prepare their organizations for previously unimagined situations. IT security has become both a necessary service and a business revenue opportunity. Balancing both perspectives requires a business portfolio approach to managing investment with income, user access with control, and trust with authentication. This book is a collection of interviews of corporate IT security practitioners offering various viewpoint on successes and failures in managing IT security in organizations.

ISBN 1-59140-247-6 (s/c) • US$29.95 • eISBN 1-59140-248-4 • 182 pages • Copyright © 2004

Managing Data Mining: Advice from Experts

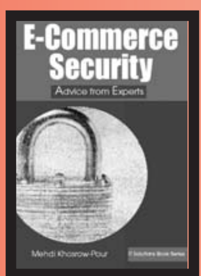


Edited by:
Stephan Kudyba, PhD, New Jersey Institute of Technology, USA
Foreword by Dr. Jim Goodnight, SAS Inc, USA

Managing Data Mining: Advice from Experts is a collection of leading business applications in the data mining and multivariate modeling spectrum provided by experts in the field at leading US corporations. Each contributor provides valued insights as to the importance quantitative modeling provides in helping their corresponding organizations manage risk, increase productivity and drive profits in the market in which they operate. Additionally, the expert contributors address other important areas which are involved in the utilization of data mining and multivariate modeling that include various aspects in the data management spectrum (e.g. data collection, cleansing and general organization).

ISBN 1-59140-243-3 (s/c) • US$29.95 • eISBN 1-59140-244-1 • 278 pages • Copyright © 2004

E-Commerce Security: Advice from Experts



Edited by:
Mehdi Khosrow-Pour, D.B.A.,
Information Resources Management Association, USA

The e-commerce revolution has allowed many organizations around the world to become more effective and efficient in managing their resources. Through the use of e-commerce many businesses can now cut the cost of doing business with their customers in a speed that could only be imagined a decade ago. However, doing business on the Internet has opened up business to additional vulnerabilities and misuse. It has been estimated that the cost of misuse and criminal activities related to e-commerce now exceeds 10 billion dollars per year, and many experts predict that this number will increase in the future. This book provides insight and practical knowledge obtained from industry leaders regarding the overall successful management of e-commerce practices and solutions.

ISBN 1-59140-241-7 (s/c) • US$29.95 • eISBN 1-59140-242-5 • 194 pages • Copyright © 2004

It's Easy to Order! Order online at www.cybertech-pub.com, www.idea-group.com or call 717/533-8845 x10 — Mon-Fri 8:30 am-5:00 pm (est) or fax 24 hours a day 717/533-8661

CyberTech Publishing

CyberTech Publishing

Hershey • London • Melbourne • Singapore

Excellent additions to your library!