The Use of Ecologic Covariates in Cohort Mortality Studies: A Re-analysis of the American Cancer Society Study of the Health Effects of Particulate Air Pollution

Thesis Submitted in Partial Completion of a Masters of Science

Department of Epidemiology and Community Medicine

Faculty of Medicine

University of Ottawa

By Alette Willis

December 2000

Supervisors: Dr. Daniel Krewski

Dr. Richard Burnett

Dr. Robert Spasoff

Abstract

Background: Recent studies of long-term exposure to ambient air pollution suggest that particulate matter in urban air is a risk factor for mortality. However, these studies are potentially biased because they have not taken into account spatial processes. This thesis examines the effect of exposure to sulfate air pollution when place-specific covariates, scale and spatial autocorrelation are controlled for.

Methods: This thesis builds upon the American Cancer Society (ACS) Study by introducing place-specific covariates into the original Cox models. In addition, a new two-stage analysis using both metropolitan and county scale covariates is employed in the analysis of the ACS data. Finally a method of filtering out spatial autocorrelation is applied to the data.

Results: For all-cause mortality, the original ACS Study found a RR of 1.15 (95% confidence interval of 1.09, 1.22) from sulfates. This was reduced to 1.06 (0.99, 1.13) by the inclusion of the place-specific covariate population change. Using the two-stage method to allow for clustering within metropolitan areas, the RR from sulfates increased to 1.25 (1.13, 1.37). When regional trends were controlled for, the RR decreased to 1.19 (1.06, 1.34). When this metropolitan data was filtered to remove spatial autocorrelation, the RR decreased again to 1.09 (1.01, 1.19). However, when counties were used instead the RR increased to 1.50 (1.30, 1.73).

Conclusions: Taking into account spatial processes affects the observed association between ambient sulfate air pollution and mortality. That the sulfate effect is larger and more robust to the inclusion of place-specific covariates at the county scale provides further evidence that ambient sulfate air pollution is a population health hazard.

Acknowledgements

I would like to thank Daniel Krewski for his untiring support of the work undertaken in this thesis. I could not have completed this thesis if it were not for Richard Burnett's enthusiasm for spatial analyses and his willingness to spend time discussing statistical analyses with me. Robert Spasoff was invaluable in ensuring that the thesis stayed on track. I would especially like to take this opportunity to acknowledge the help I received from Michael Jerrett at the School of Geography and Geology at McMaster University. Although he cannot be formally listed as one of my supervisors, because he does not have a graduate appointment at the University of Ottawa, he was nonetheless an integral part of my supervisory committee, giving generously of his time and knowledge.

The work done for this thesis was undertaken as part of a larger re-analysis project funded by the Health Effects Institute (HEI), and it could not have been done without the good-natured cooperation of the American Cancer Society and the Original Investigators, in particular Arden Pope III. I would like to thank HEI for sponsoring this work and being open to the use of spatial analyses. I learned a great deal from being a part of their rigorous peer-review process. I would also like to thank George Wells and Ian McDowell for their insightful and engaging questions at my thesis defence. The discussion they instigated enabled me to solidify my thinking on several issues addressed herein.

Working with the Re-analysis Team, made up of 31 experts, provided a wonderfully supportive and stimulating learning environment. In particular I would like to thank Yuanli Shi and Xia Luo who wrote the SAS programs which were used in the

thesis. My discussions with Mark Goldberg regarding the merits of ecologic covariates were invaluable. Norm Finkelstein and Patrick De Luca at the McMaster University Geographic Information Systems Laboratory provided technical assistance on aspects of the spatial analyses.

My thanks as well for the support I received from all of the faculty members, staff and students at the department of Epidemiology and Community Medicine, especially to Fay Draper who makes sure that all the students are well taken care of.

Table of Contents

List of Tables

(Please note, because of the number of tables in this document, they are located at the end of each chapter, rather than in the middle of the text)

List of Figures

(Please note, because of the number of figures in this document, they are located at the end of each chapter, rather than in the middle of the text)

1 Introduction

Despite having space as one of its central organizing principles (Hennekens and Buring, 1987), epidemiology has not paid close attention to spatial issues such as the role places play in determining health, the tendency for nearby places to have similar values for many variables, or the different scales at which processes operate to influence health. This thesis explores the effect that these aspects of space have on the observed relationship between ambient air pollution and mortality using data from a large cohort assembled by the American Cancer Society (ACS). Pope et al. (1995) used this cohort to link particulate air pollution to all-cause (RR 1.15, 1.09-1.22), cardiopulmonary (RR 1.26, 1.16-1.37) and lung cancer mortality (RR 1.36, 1.11-1.66).

1.1 The original study

That there are acute effects of exposure to particulate ambient air pollution has been acknowledged since the London smog disasters of 1948, 1952 and 1956 in which at least 5,000 people died (Elsom, 1992). Since that time, a great many studies of the health effects of acute exposure to various ambient air pollutants, including sulfates, have been conducted. Time-series analyses have found that periods of short-term elevation in air pollution are related to an exacerbation of respiratory problems (Roemer et al. 1993; Hoek and Brunekreef 1993) leading to an increase in emergency room visits (Delfino et

al. 1997) and hospital admissions (Burnett et al. 1994) for these problems. Daily mortality has also been observed to increase in tandem with short-term increases in air pollution (Schwartz 1991; Schwartz 1994a).

The effects of long-term exposure to particulates have not been subject to such detailed examination. The ACS Study is one of three cohort studies to have been conducted on the effects of chronic exposure to ambient air pollution. The other large studies were the Seventh Day Adventist Health Study (Abbey et al. 1991) and the Harvard Six-cities Study (Dockery et al. 1993). The Seventh Day Adventist Study found statistically significant associations between PM_{10} (particulate matter with an aerodynamic diameter of 10 μm or less) and mortality from respiratory diseases other than cancer in males and females, and between PM_{10} and lung cancer mortality in males. The Harvard Six-cities Study followed a cohort of 8,111 Caucasians between the ages of 25 and 74 at enrollment, who lived in six US cities, over 15 years. The investigators found that all-cause mortality was associated with inhalable particles ($PM_{2.5}$ - PM_{10}), fine particles ($PM_{2.5}$) and sulfate particles, and that cardiopulmonary mortality was associated with particulate air pollution.

Because the investigators in the Harvard Six-cities Study wanted to see if the results obtained with 8,111 subjects would hold over a larger number of cities, they worked with ACS to replicate their results using the Cancer Prevention Study II (CPS-II) cohort. The CPS-II cohort is composed of 1.2 million Americans recruited in 1982 from 50 US states, the District of Columbia and Puerto Rico. To be included in the study, an individual had to be at least 30 years of age and live in a household with at least one

person 45 years or older. All subjects were recruited by 77,000 ACS volunteers. As a result, the study subjects consist mostly of friends, family and acquaintances of the volunteers. Each subject completed a self-administered questionnaire upon enrollment. Their vital status was followed-up by the volunteers, who contacted the participants and their families in 1984, 1986 and 1988. For all subjects lost to follow-up, vital status was ascertained by record linkage to the US National Death Index (1982-1989). All death certificates were obtained from State Health Departments and were coded by a certified nosologist according to the International Classification of Diseases, 9th revision (WHO, 1975).

As the investigators could only obtain sulfate data for 151 American cities and fine particle data for 50, two sub-sets of the CPS-II were used in order to assess the relationship between ambient air pollution exposure and mortality in this cohort. Although it has been difficult to reconstruct the geocoding process used by the investigators, it appears that subjects were assigned to one of the metropolitan areas (MAs) based on the first three numbers of their ZIP codes as of 1982, using ZIP code boundaries from 1989. Altogether 552,138 individuals were used in the ACS study of the relationship between sulfates and mortality and 295,223 subjects were in the fine particles cohort. While the original study examined both fine particle and sulfate particle ambient air pollution, this thesis will focus only on the sulfates because a larger number of cities are required for the spatial analyses than are available for fine particles.

The sulfate particle data used for the ACS Study were collected from several different sources. Most of the data came from Ozkanyak and Thurston (1987). This data

was supplemented with data from high volume samplers in cities which did not meet US Environmental Protection Agency (EPA) annual coverage criteria and from the Inhalable Particle Network. Using this data, sulfate exposure was assigned as the mean level of sulfates in each metropolitan area prior to enrollment of subjects.

ACS assessed a large number of individual-level risk factors, such as smoking and alcohol consumption, for the CPS-II cohort. Pope et al. (1995) fit this data using Cox proportional hazards survival models stratified by sex, race and five-year age groups to determine relative risks from air pollution exposure. These Cox models showed that levels of sulfates were significantly associated with mortality from all causes, lung cancer (ICD-9 162) and cardiopulmonary disease (ICD-9 401 - 440 and 460 - 519). Fine particles were only associated with all-cause and cardiopulmonary mortality.

As other researchers have pointed out (Gamble 1988), policy makers should be cautious in interpreting these results. In particular, the results have been questioned because of possible residual confounding (from both individual and ecological risk factors); the inadequacy of the long-term exposure assessment; possible differential mis-classification errors in assessing the sulfate levels in the cities; and the robustness of the findings to alternative analytic techniques. Moreover, the spatial structure of the data and the underlying phenomena were not taken into account in the analysis conducted by ACS.

In order to address some of these issues, an independent body (the Health Effects Institute) commissioned a re-analysis of the Harvard Six-cities and ACS studies. Part of the work done for this thesis was conducted within the context of this re-analysis (Health Effects Institute, 2000). This work is expanded upon in this thesis to address

issues of space in cohort studies.

1.2 Epidemiologists in Space

Epidemiology is largely the study of patterns, either in time or in space. If two phenomena demonstrate similar patterns, they are said to be associated. According to Hill's criteria for causation, the stronger the association (ie. the overlap in patterns) between the exposure and the outcome, the more likely they are to be causally related.

Notwithstanding recent developments in the field of time-series analysis, the vast majority of environmental epidemiological studies focus on coincidence in space: whether that be the coincidence of asbestos exposure and mesothelioma in an individual or between latitude and multiple sclerosis rates in countries. Indeed, two of Hill's nine criteria for causation can be achieved by pure spatial coincidence alone: the strength of association and the biologic gradient. Phenomena may be coincident in space because they are causally related, but they may also be coincident because they are both caused by related underlying processes, or even because they are both caused by unrelated underlying processes which happen to have similar spatial patterns.

Given the growing interest in the determinants of population health, information about how broad social and environmental factors affect the health of large numbers of people is more important than ever. Questions such as how is health affected by the poverty of a community, or do ambient levels of sulfur dioxide shorten life-expectancy, require the use of data that is measured at an ecological level. Nevertheless, individual-level data such as smoking status or body mass index also greatly impact on

the health of individuals, and as individuals make up communities, on the health of the wider population. To further our knowledge of disease etiology it will become more and more important to conduct multi-level studies. In order to reduce the bias in these studies, it will be important for researchers to take into account spatial processes.

Multi-level studies are relatively new to the field of epidemiology and the majority of these studies use cross-sectional data rather than data from cohorts. Indeed, a Medline search for multi-level cohort studies found only eight published articles (Kee et al 1999; Jarvelin et al 1997; Zahm 1999; Menotti et al 1998; Navidi et al 1994; Fiscella and Franks 1997; Stengard et al 1999; Menotti et al 1997), excluding a few done in psychiatric epidemiology (cf.Driessen et al 1998). This thesis explores the use of data at different scales in cohort analyses. In order to do so it also examines issues inherent in the use of ecologic-level data: the impact of place on health, problems in assuming independence of places and the importance of choosing an appropriate scale for data collection and analysis.

The Importance of Place

Within the discipline of geography, the term place refers to a bounded location to which a set of attributes can be assigned. The boundary for this place can be as ephemeral as the area a person considers home, or as utilitarian as a provincial border.

Epidemiological studies often make reference to the concept of place by using counties or cities as the unit of analysis in ecological studies or by using places to assign characteristics to individuals in individual-level studies. In the latter case, clinical research, with its concept of the influence of the research centre on outcomes, makes the

greatest use of the concept of place.

The Original Investigators in the ACS Study did use place, specifically metropolitan areas, to assign air pollution to individual cohort members. They did not, however, use any other place-specific variables in their analysis. Therefore, the only influence of place which they included in their study was the level of ambient air pollution, which was, itself, aggregated from point-source data.

Places are rarely static, though their borders may stay the same for long periods of time. The instability of a place's boundaries over time and between individuals and institutions can impede the use of place-specific variables in analyses. For example, in 1983 the US Census Bureau discontinued the use of standard metropolitan statistical areas (SMSAs) to characterize metropolitan areas in the US. Instead, they moved to a three-level system of metropolitan statistical areas (MSAs), consolidated metropolitan areas (CMSAs) and primary metropolitan areas (PMSAs). The Original Investigators used the out-of-date SMSA boundaries to define their study areas which created some complications in developing a data-base of place-specific variables for the analyses in this thesis.

Where epidemiologists most commonly make use of the influence of place is in ecologic-level studies. In these studies, the tacit assumption is made that the area in which a person lives has an influence on their health. When used to their advantage, ecologic-level variables measure place-specific constructs such as the availability of hospital beds, the social inequality amongst city residents, or the amount of racial segregation in housing.

In order to measure valid place-specific covariates, places need to be defined in some way that reflects how individuals relate to them. For those determinants of health where government policies have a good deal of influence (such as provision of health care, zoning patterns, and education) political boundaries are the best means of defining research areas. Of course political boundaries are also useful in that they are most commonly used for the collection and reporting of data. In the US, data are also available for other areas such as 5-digit ZIP-code areas. However, data specific to ZIP-code areas is only truly useful to Postmasters who change the boundaries when necessary to make the delivery of mail more efficient.

This thesis takes the metropolitan-scale analysis conducted by the Original Investigators and adds place-specific (ecologic) variables into the models in order to account for the influence that place may have, above and beyond ambient air pollution on mortality in US cities. It also makes use of a technique of two-stage random effects analysis developed by the Re-analysis Team, which treats all individuals within an area as related but treats areas as independent.

Spatial Autocorrelation

Places themselves do not exist in a vacuum. Rather, places exist in space and as such are parts of networks and regions of influence. Both human and physical processes have patterns in space, and therefore the results of these processes show up as patterns in space. The tendency of variables to exhibit patterns is referred to by geographers as spatial autocorrelation (Meade et al, 1988).

Epidemiologists conducting ecological studies are actually exploiting spatial

autocorrelation to develop hypotheses about determinants of health. Using regression methods, these studies look for correlations between two or more variables. If a significant association is found, it is hypothesised that the variables might be related. Usually further studies are then done at the individual level to test the hypothesis. Implicit in most of these studies is the comparison of patterns in space. The problem with such ecologic-level studies is that patterns in space may be coincidentally rather than causally related.

Coincidence is not synonymous with random chance in this case. Rather, it refers to spatial autocorrelation, to the tendency for social and physical processes to exhibit patterns over space. Spatial autocorrelation in regression residuals indicates that an important independent variable is missing from the equation, or that the structure of the model is incorrect or that there is some systematic measurement error present in the data. These problems in turn can lead to biased risk estimates and invalid tests of significance (Miron, 1984).

A method of testing for spatial autocorrelation is outlined in this thesis, and is applied to the residuals of the random effects analysis. Two methods of reducing spatial autocorrelation are explained and applied to metropolitan-scale data. The first, rudimentary, method is to define regions and enter indicator variables into the regression that refer to which region the area is in. In this way, all areas within a region are treated as related. However, the utility of this approach depends upon the appropriateness of the defined regions to the research question. A second, more generalizable though more complicated approach, is to filter out the spatially autocorrelated portions of the variables

of interest. This technique is explained and applied to the metropolitan-scale data.

Level and Scale

Epidemiology acknowledges two levels of analysis: ecological and individual. The individual level is well defined while the ecological category includes all studies which do not use data on individuals. This dichotomization is quite limiting. While it allows for a clear understanding and critique of individual-level studies, the dichotomization results in misconceptions about studies at other scales of analysis by lumping them all together under the label of ecologic-level studies. Assumptions over the utility of ecological studies are often made without acknowledgement of the range and appropriateness of different scales of data. Moreover, by dividing studies into two levels, epidemiology neglects the reality that most health outcomes are caused by processes which operate on many different scales.

Ecological studies are often dismissed because of the ecological fallacy or ecological bias. This bias is most popularly defined as "the failure of expected ecologic effect estimates to reflect the biologic effect at the individual level," (Morgenstern, 1998, p. 469, who also gives credit for this definition to 6 other sources). As this definition implies, the ecological fallacy only results if findings at the ecological level are used to make inferences about effects at the individual level. However, there is by no means a consensus amongst investigators that the goal of epidemiology is to assess biologic effects at the individual level. In a fairly recent series of articles in the American Journal of Public Health, the Sussers (1996a, 1996b) and Pearce (1996) have called for epidemiologists to return to an interest in public health and its focus on the health of

populations rather than on the individuals who make them up.

Not all investigators share the goal of making inferences about individual-level relationships between exposures and outcomes. Instead, an equally legitimate goal of epidemiology is to make inferences about the relationships between exposures and outcomes measured on a group or ecological level. With respect to population health, it is important to note that the ecological fallacy also holds true when making inferences about group-level relations based on individual-level studies.

In the ACS Study, on which this thesis is based, exposure to sulfates and to fine particles was averaged over all the ambient air pollution monitors in the metropolitan area over the entire year of 1980 (or in some cases 1981). This data was then treated as an individual-level variable in a Cox proportional hazards model which used information regarding individual-level covariates and time to death, for a prospectively followed cohort of 552,138 (sulfates) or 295,223 (fine particles) people.

In part, the ACS Study was conducted because previous studies showing a relationship between this type of pollutant and mortality were limited by their ecologic design. However, the study by Pope et al. (1995) is arguably weakened by treating an ecologic exposure measurement (annual city-wide average air pollution levels) as a characteristic of the individuals in the cohort. When variables measured on one scale are applied to another scale researchers risk having ecologically biassed results.

Beyond issues related to applying data measured at one scale to another scale of analysis, the choice of scale of measurement itself is not one that should be taken lightly. Different health-determining processes operate on different scales. An analysis

conducted on the appropriate scale can reveal associations that are truly related to causal processes. One conducted at an inappropriate scale can lead to erroneous conclusions.

Many health-related processes operate on the scale of individuals. Happily this scale is also a simple one to define, and one for which a great deal of data are routinely collected. It is perhaps this clarity of scale which has led to the appeal of the epidemiological study conducted on the individual scale. Another advantage of the individual scale is that a good number of health interventions are aimed at individuals. Physicians, nurses and surgeons all deal with individuals and require accurate information about health processes on this scale. However, individuals cannot be considered autonomous units, they are related to each other by social and economic processes, not to mention shared cultural heritage.

Moreover, governments and health departments also have a role to play in the prevention and treatment of adverse health outcomes. These institutions do not deal with individuals but with large groups of people. The approach taken to provide these players with appropriate information is the population health approach. In order to take actions to improve the health of groups of people these institutions need information at scales larger than individuals. Through legislation and education, these institutions can intervene in processes which operate at wider scales, improving the health of populations through altering social processes or changing the physical environment. Unfortunately, the scale at which these processes operate is not always clear. Also, in many cases data regarding these processes are not available at the appropriate scale, due to cost or political jurisdictions. Nevertheless, if population health interventions are to have an impact on

people's health, larger scales of analysis cannot be overlooked.

This thesis uses multi-level techniques to incorporate data from both individual and ecological levels. The Original Investigators arbitrarily used the metropolitan scale for their study, but did not use any place-specific variables in their analysis other than ambient air pollution. The two-level approach allows for the incorporation of air pollution and place-specific covariates at the ecologic-level. It does not, in and of itself, solve the question of what scale the data should be measured at. In order to examine the impact of the scale of the data, a two-stage analysis was conducted first with metropolitan-scale ecologic data and then with county-scale ecologic data. (In the United States metropolitan statistical areas are made up of counties.)

Given the recent upsurge of interest in population health approaches to risk management, there will be more and more interest in studying those determinants of health that operate at the ecologic-level. In order to obtain unbiased results using ecologic-level covariates, it is essential that investigators take into account spatial processes. Through the different approaches to incorporating spatial considerations into the ACS Study on ambient air pollution and mortality, a greater understanding will be achieved of the spatial issues involved in incorporating ecologic-level variables into cohort studies.

1.3 Outline of Thesis

The objectives of this thesis are twofold. First, to examine the impact of three spatial concepts – place-specific variables, scale, and spatial autocorrelation – on the

results of cohort mortality studies. Second, to contribute to an understanding of the effect that exposure to ambient sulfate air pollution has on all-cause, cardiopulmonary and lung cancer mortality.

There are two broad steps to take in order to incorporate spatial processes into a cohort mortality study. The first step is data acquisition. Chapters 2 and 3 both deal with spatial issues in data. Chapter 2 presents a theoretical discussion regarding place-specific variables as well as an examination of the epidemiological literature on the effect of various place-specific ecologic covariates on mortality. The place-specific covariates selected for the research presented in this thesis are then outlined. These variables are used in all of the statistical analyses presented in Chapters 4, 5 and 6.

Chapter 3 also deals with data issues. In this chapter I outline how the places cohort members resided in were defined. I address problems related to using ZIP codes to assign individuals to places and I outline my attempt to recreate the boundaries employed by the Original Investigators of the ACS Study.

Once the places were defined and the place-specific covariates selected, statistical analyses were undertaken. Three types of analyses were employed in this thesis: Cox proportional hazards regression models, a two-stage random effects analysis, and a spatial filtering method. This thesis presents each of the methods and the results in a separate chapter. Chapter 4 explains the Cox method used by the Original Investigators and the results of using this method with place-specific covariates. Chapter 5 outlines a two-stage method developed by the Re-analysis Team (Health Effects Institute, 2000) in their re-analysis of the ACS Study that incorporates data at both individual and ecological

levels. The results of both metropolitan-scale and county-scale analyses are presented at the end of chapter 5. Finally, chapter 5 outlines a rudimentary way of dealing with broad regional trends using EPA airshed regions.

Chapter 6 discusses methods used by geographers to deal with spatial autocorrelation. It outlines the Moran's I test for spatial autocorrelation and presents the results of applying this test to the residuals from the two-stage models in chapter 5. Finally, a method of filtering spatial autocorrelation out of the data is explained and the results of using these filtered variables in the two-stage analysis are presented.

The results of the three types of analyses are discussed in chapter 7 in light of what they tell us about place-specific attributes, scale and spatial autocorrelation in cohort mortality studies. A discussion of what the results tell us about the effect of sulfate air pollution exposure on mortality is also presented in chapter 7. The thesis concludes, in chapter 8, with a proposed agenda for further research.

A large portion of the work done for this thesis was conducted as part of the Health Effects Institute-funded re-analysis of the ACS Study. As a member of the Re-analysis Team, I defined the places in which the cohort members resided, selected place-specific covariates, and compiled the data from various sources.

Yuanli Shi wrote the SAS Cox regression program used here, based on the one supplied by the Original Investigators, and I used the output from that program to calculate the relative risks and confidence intervals presented in chapter 4. Xia Luo and Rick Burnett wrote the SAS program for the metropolitan area two-stage random effects models which I modified to analyse the county-scale data. Once again I calculated all of

the rate ratios and confidence intervals using the SAS output from these programs.

I performed all of the spatial autocorrelation tests using Moran's I statistics. The spatial filtering was done by McMaster's Geographic Information Systems Laboratory but I produced the semivariogram and all of the maps and figures included in this thesis. The analyses using the spatial data were done using Xia Luo's SAS program. Once again, all rate ratios and confidence intervals were calculated by the author from output obtained from the SAS program.

The focus of the re-analysis was to determine what happens to the effect of sulfates on mortality when the ACS data were subjected to various alternative analyses. The thesis takes the re-analysis one step farther by investigating the effects of sulfates at the county scale. Moreover, this thesis makes a unique contribution, beyond the re-analysis, by systematically comparing the results of some of these analyses within a spatial framework.

2 Selection of Place-specific Ecologic-level Covariates

2.1 Introduction

The ACS Study can be considered as a hybrid design in that detailed individual information was collected but the primary exposure variables (sulfates and fine particles) were based on measurements taken at the metropolitan scale. Thus, although the study provided data on the joint distribution of many covariates across the study population, exposure was assessed on an ecological level. This is not a serious difficulty if metropolitan-scale ambient air pollution, as estimated from regularly collected data at the beginning of the study period, represents the relevant exposure metric for individuals. On the other hand, if these exposure metrics also represent certain characteristics inherent to the metropolitan areas, and these are correlated with other metropolitan place-specific characteristics, then it is possible that there could be residual confounding by these other characteristics.

This thesis seeks to assess the potential bias that could result if the latter situation pertained. An attempt was made to select variables that could be construed as measuring characteristics of metropolitan areas: ecologic covariates that measure social or other constructs at the group level. Although many of the selected variables were aggregated from census data (e.g. mean family income), in this case they estimate an ecologic-level construct, such as the type of social milieu that subjects lived in, with the

presumption that the place in which one lives can significantly affect one's health, after accounting on the individual level for personal, familial, or genetic risk factors. In other words, following Evans and Stoddarts (1990), amongst others, this thesis assumes that it is not merely the biology and behaviour of individuals that affects their health, but the physical and social environment that individuals find themselves in. It is also assumed that these factors act on the local or community scale; for example, if one lives in a disadvantaged neighbourhood then, despite one's own personal risk factors, the influence of the community through direct (e.g. local sources of air or water pollution) or indirect means (e.g. income inequality) may adversely affect one's health (for a review of these latter issues see Wilkinson 1996).

Ecological Bias

It is worth examining here the effect of the ecological bias on multi-level studies. According to Morgenstern, the ecological fallacy arises from confounding of the individual-level relationship due to a heterogeneity of amount of exposure to the variable of interest and other covariates within groups (1998). He outlines three sources of ecologic bias: within-group bias, confounding by group, and effect modification by group. If there is confounding of the relationship of interest in each group (from selection bias or misclassification), the resulting ecological study of the groups themselves will also be biased. This first issue cannot be addressed with multi-level analyses per se, as it has to do with the measurement of the exposure variable. In other words, multi-level studies are not immune from using an inappropriate scale of measurement. An appropriate scale of measurement can be determined, in part, by assessing whether the

exposure of the individuals within each group is more homogenous than among individuals of different groups.

Multi-level analyses are better able to deal with the remaining issues. Morgenstern has a further explanation for why effect modification by group itself results in non-communicable disease studies (1998). First, effect modification by group can arise from the differential distribution of other risk factors across groups. This can be dealt with in a multi-level analysis by controlling for risk factors at other scales, such as the individual. Second, it can arise because "the ecologic exposure variable has a contextual effect on risk separate from the biologic effect of its individual-level analogue" (1998 p. 470).

On this latter point, if we take the goal of multi-level studies to be the study of the effect of processes that operate on different scales on individual outcomes, then the sources of the ecologic bias actually become objects of study. For example, a legitimate study question is whether, and how, background disease rates (or in the case of cohort studies, relative risks) differ between groups. Moreover, the only way of studying ecologic-level risk factors or determinants of health, what Morgenstern calls "contextual effects" (1998) is to use covariates that are measured at the appropriate, non-individual, scale.

In fact, it is irresponsible to dismiss ecological covariates. They represent the only way of studying the relation between place-specific variables and population health outcomes (Schwartz 1994b). Studies done at the individual-level cannot provide valid measures of such relationships because it is equally fallacious to infer group-level effects

from individual-level studies. In particular, many social determinants of health operate at the ecologic level, such as the distribution of wealth, political stability, religion, and laws.

Ecological covariates offer the only means of studying these "contextual effects" (Morgenstern 1998). Susser differentiates two types of variables that represent characteristics of groups: integral and contextual variables (1994a). Variables such as the occurrence of a natural disaster, a law, latitude, or barometric pressure that affect virtually everyone within an area are integral variables. A contextual variable, while describing a place-specific characteristic, can be obtained from measurements at the individual-level. Contextual variables derived from measurements at the individual level may represent a different construct than the individual-level variable from which it is derived or to which it initially appears to correspond (Schwartz 1994b). For example, while the Gini coefficient is calculated from each individual's income in a city, the construct it measures is the distribution of income amongst city residents.

Ecologic Covariates and the ACS Study

In interpreting the use of place-specific variables in the context of the ACS Study, it should be recalled that the ACS enlisted a highly selected cohort whose members are not necessarily representative of the general population of the area in which they reside. Moreover, these individuals likely live in specific locations within the city for which metropolitan-scale data may not be representative. Thus, one potentially serious limitation of adjusting for place-specific ecological variables is that these data do not represent the study subjects, thus leading to misclassification. In addition, if air pollution also follows a similar geographic pattern as the ecological covariates, then it is possible

that effects from air pollution could not be distinguished from other effects measured at the same scale.

Misclassification could be differential, in that smaller metropolitan areas may be more internally homogeneous for these factors than larger ones. In any event, it has been shown that nondifferential misclassification in ecological studies will always lead to an underestimation of effects in ecological studies as compared to those derived from individual-level studies, even if there are no other sources of bias (Brenner et al. 1992).

Air pollution is determined, to a large extent, by the same economic processes which influence other risk factors. For example, heavy industry is one major source of particulate air pollution. However, cities in which such industry is prevalent are prone to other risk factors such as a depressed economy in the 1980s, and a blue-collar culture which is related to higher smoking, alcohol consumption, less exercise and a poorer diet (Garrison et al., 1993). The only chance for distinguishing between the effects of these different risk factors is if they operate on different scales. For instance, through using the CPS-II prospective cohort, some of the behavioural risk factors, such as smoking, drinking and exercise, can be controlled for at the individual-level.

There are confounding issues that are particular to ecologic covariates (Morgenstern 1998; English 1992). As Morgenstern points out, an ecologic confounder does not have to be associated with the exposure in every group, rather, across groups it may be 'ecologically' associated with the exposure (Morgenstern 1998). Ecological confounding results when the spatial pattern of the covariate and of the exposure are such that the covariate confounds the relationship between the exposure and the outcome. In

other words, unlike individual-level confounders, an ecologic covariate need only be associated with the outcome, not the exposure of interest (Morgenstern 1998). This is because regression at this level is merely a measure of spatial coincidence. For this reason, the ecologic covariates outlined below were only examined for their relationship to mortality and not to air pollutants. Moreover, due to the paucity of multi-level cohort studies, in most cases, evidence for the inclusion of these covariates was obtained from studies which were conducted entirely on the ecological level.

2.2 The Selection of Place-specific Ecologic Covariates

There is a great deal of literature on the determinants of population health which emphasizes the importance of integral and contextual risk factors on individual health. For example, in their important paper on the subject, Evans and Stoddart (1994) outline three categories of integral and contextual risk factors for health: the social environment, the physical environment and the health care system. These three categories guided the search for possible ecological confounders (see Figure 2.1).

A Medline search was done to find evidence of links between specific contextual variables within these categories and mortality. For those variables for which the literature indicated a possible health risk, data was sought from US government sources that had been collected in the early 1980s for metropolitan areas or counties. Needless to say, data was not available at this geographic resolution for all possible contextual variables.

Determinants of Health

One of the key contributions of the population health perspective to health risk management is the consideration of a wide range of determinants of health (Evans and Stoddart, 1994). This perspective goes beyond the simple disease models which focus on single factor etiology to incorporate broader social and environmental processes into more complex models of disease causation. Inherent to the determinants of health approach is the need to acknowledge that different risk factors operate on different scales of space.

The ecological fallacy is not specifically a problem in using ecologic-level data to infer individual-level relationships, rather it is a problem inherent in using results obtained at any scale of research to infer relationships at any other scale. Unfortunately, there are important determinants of health (or health risks) which exist at a number of different scales. Ambient air pollution occurs over large areas and its relationship to individual mortality may well be confounded by other processes which operate at a similar scale.

The Physical Environment

Many aspects of the physical environment have been shown to be associated with mortality and can be considered specific to places. These include climate variables, such as daily maximum temperature, daily minimum relative humidity, daily average barometric pressure, and the geographical variables latitude, longitude and altitude.

Weather: Ambient temperature is an attribute of a place that affects almost all residents. Both extreme heat and extreme cold are associated with increased mortality (Mannino and Washburn 1989; Smoyer 1998; Greenburg et al. 1983; Larsen 1990;

Kinney et al. 1991), primarily from myocardial infarction, stroke and pneumonia (Bull and Morton 1978; Danet et al. 1999). Daily maximum temperature values were obtained for all the cities in the ACS Study that had weather stations that report to the National Oceans and Atmospheric Administration (NOAA). Two variables were derived from this data: average daily maximum temperature and average monthly variation in maximum temperature. The first measure provides an indication of overall temperature while the latter measure provides an indication of the extremeness of a climate.

While temperature variables can, for the most part, be considered to be integral variables, some heterogeneity of exposure will be found between individuals within an area through activity patterns and the use of air conditioners and heaters. There is some evidence that air conditioning has a protective effect against heat-related mortality (Marmor 1975; Ramlow and Kuller 1990; Rogot et al. 1992; Seretakis et al. 1997); however, at least one study found no effect of air conditioning (Smoyer 1998). Unfortunately, no individual-level data was available for the cohort members on these variables.

Barometric pressure is an integral variable that affects everyone within a region and at least one study has found that barometric pressure is related to coronary and myocardial infarction deaths (Danet et al. 1999). In time-series studies, changes in daily barometric pressure were found to have a strong effect on daily mortality (Pope et al., 1999). Unfortunately, only data on station pressure were available from the NOAA. Station pressure is specific to the place at which it is measured and must be indexed to elevation in order to be comparable to measurements taken in other places. As no data

were readily available on the elevation of the weather stations, barometric pressure could not be calculated.

Relative humidity is another contextual variable for which there is evidence of an impact on death rates (Tselepidaki et al. 1995). In a study conducted in Greece, humidity was found to have a synergistic effect on the association between air pollution and mortality (Katsouyanni et al. 1993). However, this study applied to very hot conditions which are not relevant to many parts of the US. Minimum daily relative humidity was available from the NOAA, monthly averages were derived from these data and an average of those means was calculated to give a general indication of the level of humidity within that city. Again to provide an indication of the extremeness of relative humidity, average monthly variation was also calculated.

Altitude: Altitude appears to have a complicated relationship with mortality. While several studies have found positive associations between altitude and death rates from emphysema (Moore et al.1982), insignificant increases were found between altitude and asthma (Sly and O'Donnell 1989). Evidence of a link between altitude and chronic obstructive pulmonary disease is more controversial; whereas in one study a significant positive association (Cote et al. 1993) was found, in another study a significant negative association (Coultas et al. 1984) was observed. Negative associations between altitude and arteriosclerotic heart disease (Mortimer et al. 1977) and cancers of the tongue, mouth, oesophagus, larynx, lung and melanoma (Amsel et al. 1982) have also been found.

While altitude effects everyone who lives at that distance above sea-level, there are problems involved with taking an average measure of elevation throughout a

metropolitan area. For example, while there is little variation in elevation in some of the metropolitan areas in the mid-west, within the metropolitan area of San Diego altitude varies by 1,600 metres. Thus, metropolitan area may be an inappropriate scale of measurement for this variable.

Geographical position: Latitude and longitude are two other contextual variables that were considered as potential integral confounders. In terms of biological mechanisms, however, melanotic skin cancer may be one of the few causes of death for which latitude can actually be considered to be causally related (Lee and Scotto 1993; Decarli and La Vecchia 1986). One cohort study which incorporated latitude into its analysis did find a significant relationship between more southerly latitudes in Sweden and non-Hodgkin's lymphoma, malignant melanoma and squamous cell skin cancer (Adami et al 1999). Latitude and longitude may serve as surrogates for other variables that, for cultural or geological reasons, vary along those geographic lines (Peter et al. 1996; Fabsitz and Feinleib 1980). Because latitude and longitude are implicitly dealt with in the spatial analysis of ecological variables discussed in chapter 6, these variables will not be considered further here.

Gaseous Co-pollutants were also considered as potential confounders at the ecological level at the request of the Health Effects Institute. Daily average concentrations of nitrogen dioxide, sulfur dioxide, and carbon monoxide were obtained from 1980 to 1989, in addition to the daily one-hour maximum concentrations of ozone. Using the Atmospheric Information Retrieval System (AIRS) database, city-specific annual averages of these gaseous co-pollutants (restricted to monitors with land-use coded as

residential, commercial, and mobile) for the year 1980 were used. In this analysis, monitors were restricted to those with land-use coded as residential, commercial, or mobile, in order to avoid data from monitoring stations specifically located to monitor large point sources of pollution which may be not be reflective of general population exposures.

I had misgivings over considering these variables as variation in exposure to some of them within a city may be higher than between cities. However, considering these pollutants as potential confounders provides the opportunity to examine the utility of scale in studying the health effects of air pollutants. Different air pollutants are distributed differently in space. Sulfate particulates extend over wide distances, while ozone, nitrogen dioxide and carbon monoxide can range in concentration over small distances (Elsom, 1992). Sulfur dioxide has an intermediate distribution (Elsom, 1992). The scale of analysis may, in fact, prove to be the easiest means for differentiating between these different elements of combustion source-air pollution in terms of their impact on health outcomes.

Other Airborne Toxic Substances: The presence or absence of toxic substances in the environment can also be related to mortality. For example, exposure to benzo(a)pyrene from domestic and industrial coal burning, has been found to be associated with an increased risk of lung and stomach cancer (Lawther and Waller 1976; Nakanishi et al. 1997; Xu et al. 1996 and He et al. 1991). Although people are also exposed to benzo(a)pyrene from car exhaust (Lawther and Waller 1976), an American study concluded that the impact of such an exposure on lung cancer was most likely

negligible compared to other more potent risk factors such as smoking (Higgins 1977). Unfortunately, there is little consistent data on toxic substances such as benzo(a)pyrene available for metropolitan areas in the US during the early 1980s, so it was not included.

Water Hardness: The evidence as to whether the hardness of drinking water (as measured by the concentration of calcium carbonate, $CaCO_3$) is related to mortality is inconclusive. Most studies used cities or counties as the basis for determining exposure to water hardness (Sharrett 1981 and Comstock 1979). However, discrepancies have been found between studies that compared cities across nations, for which inverse associations have been found, and studies comparing areas within smaller regions showing no relationships (Sharrett 1981 and Comstock 1979). While $CaCO_3$ may be the agent responsible for protecting people against heart disease, it has also been suggested that magnesium in drinking water may have a protective effect against cardiovascular mortality (Borgman 1985; Sharrett 1981; Comstock 1979; Neri and Johansen 1978).

The scale of the analysis of water hardness affects the observed results. At the national-scale there is probably more heterogeneity between the areas than within them. However within regions where the geology might be similar, differences in exposure between cities may be less than those within a city (due to the use of household water softeners, filters or bottled water). As the ACS Study was a nation-wide one, it was thought that water hardness could be a confounding factor and US National Institute of Health data on water hardness were obtained from Feinleib et al. 1979 (cited in Lipfert et al. 1988).

Radon Gas: In individual-level studies, residential exposure to radon gas has

been shown to be related to lung cancer mortality (NRC 1999). At the ecological level, however, average county radon has a negative relationship to county lung cancer mortality rates (Cohen 1995). This discrepancy is almost certainly due to ecological confounding in the latter study as radon levels vary dramatically between houses within areas (see Greeland and Robbins, 1994). Radon gas does not qualify as either an integral or contextual variable at the MA or county scale and was, therefore, not included in these analyses.

The Social Environment

With the emergence of the determinants of health model of population health (Evans and Stoddart 1990) has come a concurrent interest in the effect of social factors on health outcomes. There is evidence that place-specific contextual or integral social, economic and demographic factors are related to mortality (cf. Marmot and Wilkinson, 1999; Amick 3rd et al., 1995). For this reason, data were collected from various sources on a number of these potential covariates.

Population Dynamics: Increases in urban density appear to be associated with increased mortality from heart disease (Waddell 1983). Urban density itself is closely associated with the total size of the community or city as well as with traffic levels and hence with exposure to automobile-related air pollution. However, density varies greatly over a metropolitan area, from partly rural and suburban areas to congested inner cities. The average density of a metropolitan area or a county is, therefore, not an appropriate measurement as it does not approximate exposures of the residents accurately. For this reason, density was not used.

Population change, like unemployment rates, is an indicator of economic growth or decline. It is also a variable that has no direct correlate at the individual-level and is, therefore, a contextual variable. An area that is experiencing an economic boom will likely experience a concurrent growth in its population. Conversely, an area that is in an economic recession will shrink as people leave to seek employment elsewhere. In this study, population change was used as a crude indicator of what part of the boom-and-bust cycle the metropolitan area experienced in the early 1980s. It is also possible that those who are in better health are more likely to migrate between cities in search of work. Therefore, although population change is a surrogate for a host of social, economic, behavioural and cultural attributes that are associated with recessions, it may also be an indicator of the movement of healthy people.

Unemployment: It is generally accepted that, at the individual level, unemployment is inversely related to health (Brenner and Mooney 1983; Sorlie and Rogot 1990). At the ecological level, the unemployment rates of communities have also been associated with increased mortality (Colledge 1982; Mansfield et al. 1999; Guest et al, 1998). This ecologic effect may be due to the impact of the anticipation of unemployment on health and health related behaviour (Ferrie, 1995). Thus, unemployment rate was included as a covariate.

Social Inequality: There is a growing literature linking social inequality to health outcomes (cf. Wilkinson 1996). One ecological study showed that income inequality at the state level was related to overall mortality, even after controlling for poverty (Kennedy et al. 1996). While this study used the Robin Hood Index, I prefer the

Gini coefficient which is a more widely used measure of income disparity and has also been related to increased all-cause mortality (Kawachi and Kennedy 1997).

The Gini coefficient is based on a Lorenz curve of the distribution of income in a population (Shryock et al. 1976). The Gini coefficient is the ratio of the area between the Lorenz curve and the line of equality, and the area below the Lorenz curve (Kawachi and Kennedy, 1997), and was calculated as outlined in Shryock et al. (1976). When income is completely equally distributed, each person has exactly the same income, the "curve" is a straight line and the Gini coefficient would have a value of 0. As the distribution moves away from equality, the Gini coefficient moves towards 1. It has been hypothesized that both behavioural and psycho-social factors may be responsible for the association between income disparity and mortality (Wilkinson 1996).

One cohort study has shown that when household poverty is controlled for, the income inequality of area of residence is no longer significantly associated with mortality (Fiscella and Franks 1997). Nevertheless, Gini coefficients were calculated for each metropolitan area and for each county and these measures were used as ecologic covariates in the analyses. The measure of income inequality used in this thesis is not a true Gini coefficient as census data on income groups, rather than individuals, was used to calculate it.

Poverty: While poverty is generally thought of as an individual attribute, it is also relevant to think of communities as experiencing poverty. While it has been shown repeatedly that poverty is related to mortality at the individual level (cf. Hahn et al. 1996), less research has been done into the health effects of poverty on the community level.

Nevertheless, the US Census Bureau has recently conducted important research on "poverty areas" (www.census.gov/socdemo/www/povarea.html). They found that poverty areas were significantly different from other areas in a number of essential ways, including racial composition, household composition, unemployment rates, average earnings, and levels of education. Of particular relevance to this thesis is the finding that people who lived in poverty areas were more likely to have had a debilitating health problem.

A number of epidemiological studies have also found significant detrimental health effects of living in a poverty area, even after controlling for individual variables such as race, income, education, and smoking (Haan et al. 1987; Carr et al. 1992; Polednak 1998; Waitzman and Smith 1998; Yen and Kaplan 1998 and 1999). In this thesis, both poverty rates and per capita income variables were used to investigate the impact of poverty on the findings. It is important to note that poverty areas are usually measured by the Census Bureau at the scale of tracts, which is smaller than the metropolitan or even county scale at which poverty was measured in this study.

Race: Living in a poor neighbourhood seems to have a more significant effect on African Americans than on whites (Geronimus et al. 1999). In the US, race is clearly related to mortality at the individual level. Because the ACS cohort is predominantly (94%) white (Pope et al. 1995), the impact of this relationship on the present analysis is not expected to be great.

However, there are large differences between areas in terms of the proportion of African Americans. As an ecological variable, there is at least one study that examined

proportion of African Americans in a community as a contextual variable. This study found that living in a predominantly black, poor neighborhood presented significantly higher risk of mortality than living in a predominantly white, poor neighborhood (Geronimus et al. 1999). Again, this study used measures at a smaller scale, the neighbourhood, than were used here.

Studies of environmental justice have also shown that African American communities are more likely to be burdened with toxic waste sites and industries that produce toxic pollution than are white communities (for an overview of this research see the Committee on Environmental Justice, 1999). As many of the pollutants from such sites are related to health outcomes, it seems reasonable to assume that the proportion of African Americans in a community might impact mortality rates. Because other studies have shown that Hispanic and Native American communities are often also burdened in this way (Bullard 1993), the proportion of white Americans in a community was also considered as an indirect measure of the proportion of minority group members residing in a city or county.

Education: A great deal of research has shown differences in mortality according to education attainment at the individual level (Feldman et al. 1989; Christenson and Johnson 1995; Elo and Preston 1996). It is possible that these differential mortality patterns may be related to differences in income or in health related behaviours. Studies have shown that most of the known risk factors for heart disease are more prevalent amongst less educated people (Garrison et al. 1993).

In another individual-level study, Clay et al. (1988) found that higher mortality

rates among the undereducated were attributable to higher levels of smoking-related deaths. However, even when smoking was controlled for at the individual level in the re-analysis of the ACS Study (Health Effects Institute, 2000) different relationships were found between air pollution and mortality in the different strata of those without a high school education, those with a high school education and those with more than a high school education (Health Effects Institute, 2000). The difference between the three strata was so marked that while the lower two groups had a significantly elevated risk of mortality from sulfate or fine particle exposure, those in the highest group did not.

While a cursory Medline search did not reveal any health articles which used education as a contextual ecologic variable, because education was found to be an important variable at the individual level, I chose to also examine it at the ecologic level. Because this variable could be controlled for at the individual level, any effect of education at a larger scale can be considered to be purely contextual.

Education attainment was calculated as the percentage of persons 25 years of age and over in counties who had responded on the census that they had completed four years of secondary schooling or had completed some post-secondary education. Counties were then aggregated up to metropolitan areas. It was postulated that the education levels of a metropolitan area might act as a surrogate measure of the type of economy of that city and of the predominance of certain health related behaviours in that city.

Health Care Services

Health Insurance: At the individual level, persons without health care insurance are at higher risk of mortality than those with insurance (Brown et al. 1998).

Unfortunately, medical insurance rates are only available at the state level, but not at the metropolitan level. These state-wide medical insurance rates have been shown to be associated with mortality rates (Kaplan et al. 1996). The Kaplan study also found that the prevalence of uninsured persons is highly correlated with income inequality, therefore, the inclusion of income inequality may control in part for health insurance prevalence.

Hospital Beds and Physicians: Health services indicators that are readily available at both the metropolitan and county scales include the number of hospital beds and the number of physicians per 100,000 residents. At least one study has found that the number of hospital beds is positively associated with hospitalization rates for respiratory disease, yet the number of physicians is negatively associated with this outcome (Morris and Munasinghe 1994). The cause-effect relationship between health service provision and health outcomes is complex. While access to health care can have a large impact upon the health of individuals within a community, health care services also tend to locate in areas which have greater health care needs.

While it is not clear whether the number of physicians or hospital beds may be a confounder for mortality rates, these variables were included in the analyses.

Infant Mortality Rates: Infant mortality rates (IMR) are often used as measures of development. Within the US, IMR has been used as an outcome measure when examining racial differences in reproductive health (Geronimus 1986). IMR in the US is highly correlated to both the percentage of African Americans in a community and to the level of poverty within that community. Most of the influence of these two factors may be captured through some of the other ecologic covariates we used, such as poverty

rates, mean income and percent African American.

Most importantly, acute exposure to air pollution has been found to be related to IMR (Woodruff et al. 1997; Loomis et al. 1999), therefore, controlling for IMR could over adjust the association between air pollution and mortality. For all of these reasons, IMR was not used as an ecologic covariate.

Summary of Ecologic Covariates Included

Table 2.1 shows those ecologic variables that were used in this thesis. This table lists the variable, key studies that provide evidence that the variable may or may not be linked to health outcomes, whether the variable was used, the source of data for the variable and how the variable was derived. The actual values used for the study can be found in Appendix A.

Table 2.1. Ecologic Covariates Considered with Variable Description and Source of Data

Potential Covariates		Studies Showing Link With Mortality	Derivation of Variable	Source of Data
Demographic Data				
Change in population between 1980 and 1986	✔	Waddell 1983	The percent net change in number of residents between 1980 and 1986.	This figure is based on the US Bureau of the Census, *1986 Population Estimates by County with Components of Change*, cited in the County City Data Book, 1988.
Population density	✘	There is little evidence that the density of a metropolitan region contributes to health outcomes.		
Total population	✘	This variable was obtained to provide information about the relative size of cities. It was not used as an ecologic covariate as there is little evidence of a causal role for this variable	The total population for 1980.	from the 1980 Census of Population and Housing as of April 1, 1980, cited in the County City Data Book, 1988.
Racial Composition of Community				
Percent African American	✔	Geronimus et al 1999; Polednak 1993	The percent of persons in 1980 who classify themselves as being of black race.	Geolytics software from 1980 Census of Population and Housing (April 1, 1980), STF3 data.
Percent White	✔	Geronimus et al 1999; Polednak 1993	The percent of persons in 1980 who classify themselves as being of Caucasian race.	From the US Bureau of the Census County Population Estimates (experimental) By Age, Sex, and Race: 1980-1984, cited in the County City Data Book, 1988.

Potential Covariates		Studies Showing Link With Mortality	Derivation of Variable	Source of Data
Health Services Data				
Physicians per 100,000 residents	✔	Morris and Munasinghe 1994	The number of professionally active, non-Federal physicians with known addresses per 100,000 resident population as of July 1, 1985.	American Medical Association, Chicago, IL, Physician Characteristics and Distribution in the US, 1986 edition cited in the County City Data Book, 1988
Hospital beds per 100,000 residents	✔	Morris and Munasinghe 1994	The number of hospital beds per 100,000 resident population as of July 1, 1985. From a survey (September 30, 1985) of all hospitals (registered and un-registered), excluding old-age homes, convalescent homes and sanatoriums.	Source: American Hospital Association, Chicago, IL, Hospital Statistics, 1986, cited in the County City Data Book, 1988
Percentage with medical insurance	✘	Brown et al 1998 and Kaplan et al 1996	Data on health insurance rates is only available at the state level.	
Access to medical care by age, sex, race and level of education	✘		No suitable data available at the ecologic level.	
Infant mortality rate	✘	IMR is correlated with both poverty level and percentage of African Americans. IMR may also be correlated with air pollution, raising the possibility of overadjustment (Woodruff et al 1997; Loomis et al 1999).		
Iron serum levels	✘	Meyers 1996 finds inconclusive evidence of a link between iron intake and mortality	No suitable ecologic level data is available.	

Potential Covariates		Studies Showing Link With Mortality	Derivation of Variable	Source of Data
Socio-economic Aspects of the Environment				
Mean annual income in 1979	✔	Silver 1973; Kitagawa and Hauser 1973; Pappas et al. 1993	Per capita income for 1979.	Geolytics software from 1980 Census of Population and Housing (April 1, 1980), data. Geolytics
Median annual income in 1979			Median income in a county for 1979.	from the 1980 Census of Population and Housing as of April 1, 1980, cited in the County City Data Book, 1988.
Unemployment rate	✔	Brenner and Mooney 1983; Colledge 1982; Catalano et al. 1993; Ferrie et al. 1995	Percent of total civilian labour force who were unemployed in 1986.	US Bureau of Labor Statistics, Employment and Unemployment in States and Local Areas, annual, cited in the County City Data Book, 1988
Individual poverty rate	✔	Geronimus et al 1999; Haan et al 1987; Kaplan 1996; Polednak 1998; Polednak 1993; Menchick 1993; Hadley 1982	The percent of individuals in 1979 who were classified as living below the poverty level specific to their family size, age, and number of dependents	US Bureau of the Census, Current Population Reports, Series P-26, Nos. 86-NE-SC, 86-S-SC, 86-ENC-SC, 86_WNC-SC, and 86-W-SC; and 1980 Census of Population and Housing, Summary Tape File 3C, cited in the County City Data Book, 1988
Percent with high school education	✔	Christenson 1995; Clay et al 1988; Elo and Preston 1996; Feldman et al 1989; Garrison et al 1993	The number of persons 25 years of age or older who indicated that they had completed 4 years of high school or some years of college divided by the total number of persons 25 years and over	Geolytics software from 1980 Census of Population and Housing, (April 1, 1980), STF3, data.
Gini coefficient	✔	Wilkinson 1996; Kawachi and Kennedy 1997	Calculated from income group data for 1979 as outlined in Shryock et al. 1976.	Geolytics software from 1980 Census of Population and Housing (April 1, 1980), STF3 data.

Potential Covariates		Studies Showing Link With Mortality	Derivation of Variable	Source of Data
Crime rate	✘	While at least one study has shown the potential relevance of crime rates to health outcomes (Kawachi, Kennedy and Wilkinson 1999). The problems with the data available made it impossible to use crime data in the Re-analysis.	The reported incidences of the following crimes in 1985: murder, nonnegligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny-theft and motor vehicle theft. The total numbers were divided by the population for 1986. These numbers are not adjusted for underreporting which may make it difficult to compare across geographical areas	
Sedentary lifestyle	✘		No reliable data available at the ecologic level; however individual level data is available for the ACS cohort	
Diet	✘		No suitable data available at the ecologic level	
Climate Data				
Mean maximum daily temperature	✔	Bull and Morton 1978; Danet et al 1999; Gorjanc et al 1999; Greenberg et al 1983; Larsen 1989; Rogot and Padgett 1976	Maximum daily temperature in Fahrenheit for the years 1980 to 1989 inclusive was averaged by month. The average of all monthly averages was used as the ecologic covariate.	The data was provided to us by the US National Climatic Data Center of the National Oceanic and Atmospheric Administration.
Mean monthly variation in maximum temperature	✔		The monthly variation in maximum daily temperature in Fahrenheit for the years 1980 to 1989 inclusive was obtained. The average of this monthly variation was used as the ecologic covariate	The data was provided to us by the US National Climatic Data Center of the National Oceanic and Atmospheric Administration.

Potential Covariates		Studies Showing Link With Mortality	Derivation of Variable	Source of Data
Mean daily minimum relative humidity	✔	Katsouyanni et al 1993; Dalib 1997; Tselepidaki et al 1995	The minimum daily relative humidity in whole percent for the years 1984 to 1989 inclusive was averaged by month. The average of all monthly averages was used as the ecologic covariate.	The data was provided to us by the US National Climatic Data Center of the National Oceanic and Atmospheric Administration.
Mean monthly variation in minimum relative humidity	✔		The monthly variation in minimum daily relative humidity in whole percent for the years 1984 to 1989 inclusive was obtained. The average of this monthly variation was used as the ecologic covariate	The data was provided to us by the US National Climatic Data Center of the National Oceanic and Atmospheric Administration.
Mean and variation of barometric pressure	✘	While there is some evidence that barometric pressure may affect health (Danet et al 1999) We were unable to obtain reliable barometric pressure data.	Because no reliable measure of altitude could be derived for metropolitan areas we were unable to calculate an appropriate value for barometric pressure out of the daily station pressure data available from the NOAA	US National Climatic Data Center of the National Oceanic and Atmospheric Administration.
Residential use of air conditioning	✘	Seretakis et al 1997; Rogot et al 1994; Ramlow and Kuller 1990 and Marmor 1975	Data on air-conditioning use is only available at the state level	

Potential Covariates		Studies Showing Link With Mortality	Derivation of Variable	Source of Data
Physical Environment Data				
Altitude	✔	Amsel et al 1982; Cote et al 1993; Coultas et al 1984; Moore et al 1982; Mortimer et al 1977	It is measured in metres above sea-level. Only one measure of elevation was provided for each city. We do not have a measure of the variation of elevation within all cities, however, we know that in at least one metropolitan area used in the ACS study, the elevation varies by as much as 1600m.. Therefore usefulness of a single measure of elevation may be limited	Environmental Systems Research Institute US Places (24000+) file provided with their ArcView GIS software. Nassau was derived from monitor location data using spatial interpolation. Data for El Paso, TX was obtained from their official website
Latitude and longitude	✔	Peter et al 1996; Fabisitz and Feinleib 1980	Implicitly controlled for in the spatial analysis, see appendix **	
Water hardness	✔	Borgman 1985; Comstock 1979; Neri and Johansen 1978; Sharrett 1981	Data on concentration of CaCo3 (ppm) in drinking water, measured ca. 1970.	National Institute of Health data cited in Feinleib et al. 1979.
Radon	✘	Ecological studies contradict analytical studies		
Benzo(a)pyrene	✘	Higgins 1977	No comprehensive data base was located for the US	
Gaseous Co-pollutants				
Sulfur Dioxide	✔		All gaseous co-pollutants were measured as annual average of daily one-hour maximum concentrations for 1980 from residential, commercial or mobile monitors.	All data on gaseous co-pollutants came was compiled by Center for Air Pollution Impact and Trend Analysis (CAPITA) from the EPA AIRS database.
Carbon Monoxide	✔			
Nitrogen Dioxide	✔			
Ozone	✔			

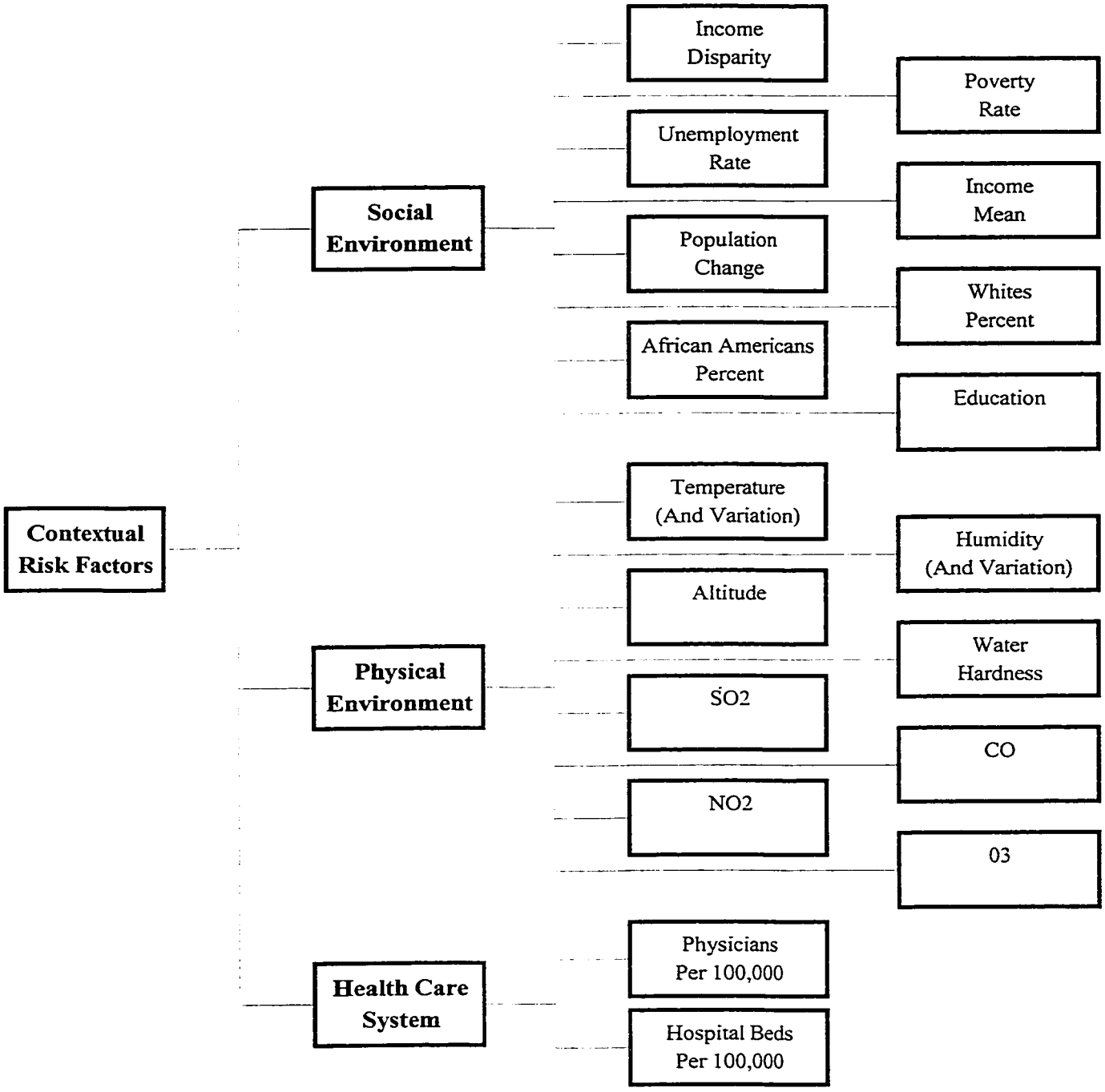


Figure 2.1. Selection of Place-specific Covariates Using a Determinants of Population Health Framework

3 Construction of Place-specific Ecologic Covariate Database

When integrating place-specific variables into a cohort study, assignment of cohort members to places must be done carefully. Therefore, this section outlines how the places used in this thesis were defined. ACS used the first three digits of the respondents' ZIP codes to match them with metropolitan areas. Because the basic unit for which US Census data are available is the county and because counties comprise metropolitan areas, the counties in which ACS respondents resided had to be identified so that the ecologic covariates could accurately reflect the places in which people are exposed. For the purposes of this thesis, an analysis was conducted at the scale of metropolitan areas and counties.

3.1 Definition of Metropolitan Areas in the American Cancer Society Study

Census Metropolitan Areas

Before describing the process by which the geographical units for the place-specific covariates were determined, it is necessary to define certain terms used by the US Census Bureau. The term "standard metropolitan statistical area" (SMSA), historically used to characterize metropolitan areas in the US, is no longer used by the Bureau. It was supplanted in 1983 by the terms "metropolitan statistical area" (MSA), "consolidated

metropolitan statistical areas"(CMSAs), and their component "primary metropolitan statistical areas" (PMSAs) (see Figure 3.1). An MSA must have at least one city with 50,000 or more inhabitants, and a metropolitan region which contains a total of at least 100,000. MSAs are comprised of the county which includes the largest city, as well as any adjacent counties in which at least 50% of the population is in an urbanized region contiguous with that city. If an MSA contains a minimum of one million inhabitants or more and can be broken into recognizable component areas (using statistical criteria as well as local opinion), it may be designated as a CMSA, with the components becoming PMSAs. Together, CMSAs, PMSAs and MSAs are referred to as "metropolitan areas" (MA) in this thesis.

The Original Investigators' Approach

The Original Investigators used a file to link the ZIP codes of the Cancer Prevention Study II respondents to 151 metropolitan areas in the US for which sulfate data were obtainable. No documentation was available regarding the construction of this file. The file matched the first three digits of ZIP codes to "SMSA" ID numbers, a four-letter code for the city, and a two-letter code for the state (see Appendix B). As discussed below, the use of the obsolete term SMSA instead of the more specific terms MSA, PMSA or CMSA presented some difficulty in reconstructing the ACS file.

The ZIP codes were obtained from the questionnaires which the ACS CPS-II respondents filled out in the fall of 1982 (Pope et al. 1995). The Original Investigators used MA boundaries, as defined on June 30, 1983, and ZIP code maps provided in the 1989 National Five Digit ZIP Code and Post Office Directory (Pope, personal

correspondence).

In the process of obtaining air pollution data from the AIRS database using the SMSA codes provided by the Original Investigators (see Appendix B), it became apparent that the geographical areas to which these codes referred did not match, in all cases, the areas covered by the ZIP codes of participants in the CPS-II study.

Taking the ACS location "SLOU, MO" as an example, the SMSA number it was assigned (7040) only refers to St. Louis proper—the PMSA on the Missouri side of the border (see Appendix B). The ZIP codes assigned to SLOU, MO, however, are from both sides of the border, as indicated by the differences in the first two digits of the three-digit forward sortation codes (620, 622, 630, 631). Moreover, ACS obtained air pollution data for SLOU, MO at a time when the state of Illinois monitored air pollution but the state of Missouri did not. While there is no air pollution data available for the PMSA of "St. Louis" (7040), the larger CMSA of "St. Louis-East St. Louis-Alton, MO-IL" (7042) includes counties in Illinois for which AIRS data are available. Therefore, it seems likely that the Original Investigators were referring to the CMSA 7042, rather than to the PMSA 7040.

It could not be assumed, however, that all of the MAs used by ACS were at the level of CMSAs. Three California MAs included by ACS—Los Angeles, Anaheim and Riverside—are of necessity PMSAs that comprise the larger CMSA of Los Angeles-Anaheim-Riverside. The ID codes provided by the Original Investigators for these three places are all PMSA codes, as they should be. Given these discrepancies, it became necessary to attempt some sort of validation of the mapping of ZIP codes onto

metropolitan area identification numbers.

ZIP Codes and the Census

The following information regarding ZIP codes was obtained through consultation with Rose Quatero, the ZIP code expert in the Geography Division, US Bureau of the Census; although similar issues are documented by Dr. Michael Frank Goodchild at the National Center for Geographical Information and Analysis (www.ncgia.ucsb.edu/~good/176b/p04.html#SEC29.3.1).

ZIP codes are derived from postal carriers' routes. ZIP codes and their boundaries are the purview of local Post Masters who may change the local boundaries at any time in order to provide more efficient service. For the most part, these changes affect only the last two digits of a ZIP code; however, a local Post Master may have jurisdiction over an area which crosses state boundaries in which case changes in the first three digits may occur. As well, since 1980, three- digit codes which were previously reserved for the US military were added to civilian ZIP codes.

Because of this fragmented jurisdiction, it would have been necessary to contact each local Post Master and obtain his/her archives for 1982 in order to compile a comprehensive file of ZIP codes. This was obviously outside the scope of this thesis. Alternatively, there are products available which are based on the 1980 census and which contain ZIP codes. These ZIP codes were taken directly from the census questionnaires. Unfortunately, although there were approximately 40,000 ZIP codes in existence at the time, the census only published information for about 20,000 (conversation with Rose Quatero).

In the late 1980s, commercial companies such as Rand McNally began to provide atlases of ZIP codes. The Original Investigators used ZIP code maps from this era: the 1989 National Five Digit ZIP Code and Post Office Directory. However, these products were not always reliable and obviously applied to 1989 not 1982. For the purpose of verifying the mapping of ZIP codes to SMSAs, 1999 ZIP codes available on-line were used (www.usps.com/ncsc/lookups/lookup_zip+4.html).

Using this web site, the acceptable ZIP codes for cities and towns falling within the 1983 MSA boundaries, as shown in the County City Data Book (US Bureau of the Census 1988), were obtained. The first three digits of these ZIP codes were compared with those provided by the Original Investigators. The results of this examination are included on the right half of Appendix B. It was determined that the ZIP codes provided for STLO MO did include the PMSAs in Illinois. Thirty-nine other discrepancies were found between the SMSA code used and the ZIP codes. STLO MO was not the only area for which a PMSA code was used when the ZIP codes actually applied to an entire CMSA (see for example, DALL, TX). In other cases, ZIP codes which applied to one or some counties in a metropolitan area were left out (see for example, OKLA, OK or DULU, MN). At this stage, without additional information, it was assumed that the missing ZIP codes were due to the addition of ZIP codes since 1989, which had been reserved for the military, and these counties were included in the MAs to which the Original Investigators had assigned them.

In other cases, the ZIP codes listed actually applied to counties outside of the boundaries to which the ID code referred. Occasionally these counties were in other

metropolitan areas (see, for example, CRAP IO); in other cases the ZIP codes referred to rural counties adjacent to metropolitan areas (see for example, CHAR WV).

Aggregation of Metropolitan-scale Covariates

The US Postal Service web site was used to determine the counties which were predominantly made up of ZIP codes used to assign ACS cohort members to places (see Appendix B). For the purposes of this thesis, the largest relevant geographical area was used to represent each metropolitan region in the ACS Study. Thus, if the ZIP boundaries encompassed two adjacent metropolitan areas, data for both were aggregated into one value. On the other hand, if the ZIP codes for one of the counties in the central metropolitan area were missing (such as OKLA OK), that county was still included because it is known that new three-digit codes have been added to the US since 1989.

Depending upon the type of data provided (see Table 2.1), weighted averages (according to the number of residents in each county) or simple aggregations, were calculated from component counties for each metropolitan area. In some of the Eastern Seaboard States (CT, ME, MA and NH), metropolitan areas were aggregated from towns and cities, not from counties. The list of component places contained in the 1983 US Census Bureau definition of metropolitan areas (http://www.census.gov/population/estimates/metro-city/83mfips.txt) was used to aggregate data for metropolitan areas in these states.

The possibility of using 1982 ZIP code boundaries to determine the geographical statistical areas represented by the ACS cohort was considered. As noted above, however, a complete ZIP code file does not appear to exist for that time period.

Further, the metropolitan area boundaries used by the Original Investigators remain unclear. Consequently, it may be difficult to significantly improve upon the approach taken here to decide likely geographic boundaries for the 151 cities included in the ACS Study. The lack of clearly defined areas could lead to misclassification biases if the ecologic "exposure" variables calculated do not refer to the same area in which the ACS respondents reside. This type of bias generally results in an underestimation of the relative risks associated with the ecologic covariates (Hennekens and Buring, 1988).

3.2 County-scale Place-specific Data

Sulfate Data

In order to investigate the influence of scale, data were also acquired for counties. Some problems of comparability were uncovered with the Original Investigators' use of air pollution data from different sources (see Health Effects Institute, 2000). For simplicity's sake, the county-scale analysis done for this thesis used only sulfate data from the AIRS database which used the same type of monitor in all jurisdictions. Monitor data from the AIRS database was obtained from the Washington University's Centre for Air Pollution Impact and Trend Analysis (CAPITA). All those counties which contained at least one sulfate monitor in 1980 or 1981 in the AIRS network and for which there were at least 10 observations available during these years were identified. Sulfate exposure was calculated as the mean of all the sulfate measurements taken in the county in 1980 or 1981. Sulfur dioxide exposure was obtained similarly.

Identification of Counties

There are 513 counties which make up the metropolitan areas used in the original study. Only 139 counties met the above criteria and were, therefore, considered for the county-level analysis (see Table 3.1). Using the Std3b census files from the US National archives, a list of 5-digit ZIP-codes was derived for each county. Data were then obtained from the US Census and other sources for each of these counties (see Table 2.1).

The advantage of 5-digit ZIP codes is that they have a finer granulation and therefore can be matched more closely to the desired geographic area. This method of determining place of residence for the ACS county cohort is also more accurate than that used by the Original Investigators because it matches the 1982 ZIP codes of the subjects to 1980 ZIP-code boundaries, rather than to 1989 boundaries.

The main disadvantage is that the census file only contains about half of all the ZIP codes which existed in 1980. As a result, some of the ACS participants who were included in the metropolitan-scale analysis were excluded from the county-level analysis. Other metropolitan-scale cohort members were excluded from the county-scale analysis because the 3-digit ZIP code areas often extend beyond county and metropolitan borders (see Figure 3.2). For example, the metropolitan areas of Great Falls, MT and Billings MT are made up of only one county each, Cascade and Yellowstone respectively. Using 3-digit ZIP codes, 600 respondents were identified for Great Falls and 855 for Billings. When 5-digit ZIP codes were used instead, only 452 and 795 respondents were selected for each of the respective areas, a loss of only about 25% and 7% respectively.

Because of the greater conciseness of matching cohort members to counties,

the lack of sulfate data for a number of the component counties of the original metropolitan areas, and the limited availability of 5-digit ZIP code information, only 245,315 members of the ACS sulfate cohort were included in the county analysis. In other words, 346,395 members of the metropolitan-scale analysis were excluded because they lived in a county without sufficient sulfate data, because they did not actually reside in a component county, or because their ZIP code did not match any of those in the Census Bureau database.

During the process of mapping ACS cohort members to counties, six more counties were discarded from the study because no cohort members resided there. A seventh county was also discarded because only one cohort member lived there, leaving a total of 132 counties in the county analysis (see Figure 3.3). Of the remaining counties, the average number of subjects was 1858 (with a close median of 1329), while the smallest number was 79 (Merrimack, NH) and the largest number was 9527 (Hennepin, MN).

3.3 Values of the Ecologic Covariates

In Appendix A, the actual values of the ecologic covariates used are listed for each of the 151 metropolitan areas in the ACS Study and the 132 counties used in the county analysis. These data were derived from many sources as outlined in Table 2.1. Most of the socio-economic data comes from the US Census Bureau. The climate data were obtained from the US National Oceans and Atmospheric Administration from their database of reporting sites.

Data from these sources was then aggregated to the appropriate geographical area, (see Tables 3.1 and 3.2). For several reasons, it was not possible to obtain a full set of ecologic covariates for all of the metropolitan area. The NOAA only archives data for a very specific set of meteorological stations in the US. Not all of the ACS cities are NOAA reporting stations. The values for the hardness of water were taken from a previous report (Feinleib et al. 1979, cited in Lipfert et al. 1988). Not all of the ACS cities were included in this other study.

The gaseous covariate data were obtained from the AIRS database through CAPITA. As different monitoring programs applied to each gaseous pollutant, data are not available for all of the ACS metropolitan areas for these covariates.

Some 12 or 13 ACS MAs are missing metropolitan area-specific data on population change, doctors per 100,000 and hospital beds per 100,000 (see Appendix A). There are two reasons for this. First, some of these MAs did not correspond to metropolitan areas as defined by the Bureau. Second, some of the MAs in New England are defined in terms of towns not counties. As metropolitan-specific data had to be aggregated up from counties, and not all of the data available for counties are available at the town level, many of the New England MAs are missing data. For example the number of physicians was not available for towns and therefore, these MAs are missing a value for this variable.

For all of these reasons, data for all of the covariates were not available for every MA, Table 3.2 lists the number of metropolitan areas for which each ecologic covariate was available as well as the mean, minimum and maximum values of those

variables. In the following chapters, all confidence intervals are at 95% confidence and are calculated over the range of the data.

Table 3.2 also lists the descriptive statistics for the county-scale data. The only data that were readily available at the county-scale was census data. Therefore, variables concerning the physical environment were left out of the county-scale analysis.

Descriptive Statistics for the Ecologic Covariates

As noted above, it was not possible to obtain data on all of the ecologic covariates of interest for each metropolitan area in the ACS Study. While the socio-economic, race and demographic data were available for almost all cities, data on the physical environment, climate and gaseous co-pollutants was readily available only for certain cities. The variables with the most missing values are relative humidity and nitrogen dioxide, which were available from a reliable national source for only 95 and 74 out of 151 MAs respectively. No models were available for estimating the values of missing data points with the information that was available, therefore, cities with missing data were left out of the analyses that used that variable.

For the most part, data were available for all the counties. The exception is that population change data were not published for Middlesex, MA. No explanation is given as to why the data are missing (US Census Bureau, 1988).

Interestingly, the average values of the MA variables are very similar to the average values of the county variables. Therefore, the two sets of data are representing a similar cross-section of the US. The range in the data are also quite similar in most cases. Exceptions to this are a slightly wider range in poverty rates, in percent racial

composition and number of doctors between the counties than was seen between the MAs. This might indicate that within MAs there is a greater segregation of individuals and services than between MAs.

The range in sulfates at the MA level is from 3.6 $\mu g/m^3$ in Great Falls, MT to 23.5 $\mu g/m^3$ in Steubenville, OH. A similar spread in sulfates is seen amongst the counties which range from 3.8 $\mu g/m^3$ in Cascade MT to 24.1 $\mu g/m^3$ in Carbon, PA. While Cascade, MT is in Great Falls, MT, Carbon, PA is approximately 400 km away from Steubenville, OH. Nevertheless, when the study covers the entire US, 400 km is not that far.

Correlations Amongst Place-specific Covariates

Collinearity can be a problem in ecological studies since socio-economic and environmental risk factors are often more correlated with each other between groups than they are at the individual level (Morgenstern 1998). If such covariates are not sufficiently controlled for in the design, it may be nearly impossible to differentiate their effect through statistical analysis. Luckily, other than the variables percent Whites and percent African Americans there do not appear to be any cases where the collinearity between two variables is uncomfortably high in the MA data set (see Table 3.3).

Within the counties (see Table 3.4), however, the Gini coefficient and the poverty rate were also highly correlated (87%) as were poverty and median income (-76%). As might be expected the correlation between the Gini and income was also high (-67%). Poverty was also strongly correlated with race (69%).

In general the correlations between variables in the MA data set are similar to

those in the county data set. However, the Pearson coefficients do seem to be a bit more extreme (larger or smaller) in the counties than in the MAs. In the county analyses, median income was used instead of mean income and it appears to relate to the other place-specific variables in different ways. The correlations between poverty and physicians and population change differ between the two data sets.

Table 3.1. Counties Used in the County-scale Analysis[1]

County Name	State Name	Fips Code
Counties with monitors in which ACS cohort members reside		
Broward	Florida	12011
Duval	Florida	12031
Hillsborough	Florida	12057
Orange	Florida	12095
Pinellas	Florida	12103
Fulton	Georgia	13121
Muscogee	Georgia	13215
Lake	Indiana	18089
Monroe	Indiana	18105
St. Joseph	Indiana	18141
Vanderburgh	Indiana	18163
Vigo	Indiana	18167
Black Hawk	Iowa	19013
Dubuque	Iowa	19061
Linn	Iowa	19113
Polk	Iowa	19153
Sedgwick	Kansas	20173
Shawnee	Kansas	20177
Fayette	Kentucky	21067
Kenton	Kentucky	21117
East Baton Rouge	Louisiana	22033
Orleans	Louisiana	22071
Androscoggin	Maine	23001
Cumberland	Maine	23005
Penobscot	Maine	23019
Prince Georges	Maryland	24033
Bristol	Massachusetts	25005
Hampden	Massachusetts	25013
Middlesex	Massachusetts	25017
Plymouth	Massachusetts	25023
Worcester	Massachusetts	25027
Genesee	Michigan	26049
Ingham	Michigan	26065
Wayne	Michigan	26163
Hennepin	Minnesota	27053
Ramsey	Minnesota	27123
Hinds	Mississippi	28049
Cascade	Montana	30013
Yellowstone	Montana	30111

Table 3.1. Counties Used in the County-scale Analysis[1]

County Name	State Name	Fips Code
Douglas	Nebraska	31055
Lancaster	Nebraska	31109
Clark	Nevada	32003
Washoe	Nevada	32031
Hillsborough	New Hampshire	33011
Merrimack	New Hampshire	33013
Rockingham	New Hampshire	33015
Hudson	New Jersey	34017
Mercer	New Jersey	34021
Passaic	New Jersey	34031
Bernalillo	New Mexico	35001
Albany	New York	36001
Bronx	New York	36005
Chemung	New York	36015
Dutchess	New York	36027
Erie	New York	36029
Kings	New York	36047
Monroe	New York	36055
Nassau	New York	36059
New York	New York	36061
Oneida	New York	36065
Onondaga	New York	36067
Queens	New York	36081
Rensselaer	New York	36083
Richmond	New York	36085
Schenectady	New York	36093
Suffolk	New York	36103
Westchester	New York	36119
Durham	North Carolina	37063
Forsyth	North Carolina	37067
Mecklenburg	North Carolina	37119
Butler	Ohio	39017
Clark	Ohio	39023
Cuyahoga	Ohio	39035
Franklin	Ohio	39049
Hamilton	Ohio	39061
Jefferson	Ohio	39081
Montgomery	Ohio	39113
Richland	Ohio	39139
Stark	Ohio	39151
Summit	Ohio	39153

Table 3.1. Counties Used in the County-scale Analysis[1]

County Name	State Name	Fips Code
Trumbull	Ohio	39155
Warren	Ohio	39165
Wood	Ohio	39173
Oklahoma	Oklahoma	40109
Multnomah	Oregon	41051
Allegheny	Pennsylvania	42003
Berks	Pennsylvania	42011
Blair	Pennsylvania	42013
Bucks	Pennsylvania	42017
Cambria	Pennsylvania	42021
Carbon	Pennsylvania	42025
Centre	Pennsylvania	42027
Chester	Pennsylvania	42029
Cumberland	Pennsylvania	42041
Dauphin	Pennsylvania	42043
Delaware	Pennsylvania	42045
Erie	Pennsylvania	42049
Fayette	Pennsylvania	42051
Lackawanna	Pennsylvania	42069
Lancaster	Pennsylvania	42071
Lebanon	Pennsylvania	42075
Lehigh	Pennsylvania	42077
Lycoming	Pennsylvania	42081
Mercer	Pennsylvania	42085
Montgomery	Pennsylvania	42091
Northampton	Pennsylvania	42095
Perry	Pennsylvania	42099
Philadelphia	Pennsylvania	42101
Washington	Pennsylvania	42125
York	Pennsylvania	42133
Kent	Rhode Island	44003
Newport	Rhode Island	44005
Providence	Rhode Island	44007
Charleston	South Carolina	45019
Greenville	South Carolina	45045
Pickens	South Carolina	45077
Richland	South Carolina	45079
York	South Carolina	45091
Davidson	Tennessee	47037
Hamilton	Tennessee	47065
Shelby	Tennessee	47157

Table 3.1. Counties Used in the County-scale Analysis[1]

County Name	State Name	Fips Code
Bexar	Texas	48029
Dallas	Texas	48113
Harris	Texas	48201
Hampton	Virginia	51650
Norfolk	Virginia	51710
Roanoke	Virginia	51161
King	Washington	53033
Spokane	Washington	53063
Dane	Wisconsin	55025
Eau Claire	Wisconsin	55035
Milwaukee	Wisconsin	55079
Racine	Wisconsin	55101
Counties with monitor data but with no ACS cohort members residing in them		
Essex	Massachusetts	25009
St. Louis[2]	Minnesota	27137
Camden	New Jersey	34007
Gloucester	New Jersey	34015
Lawrence	Ohio	39087
Douglas	Wisconsin	55031

[1] In mapping ACS cohort members onto counties, 12 701 5-digit ZIP codes were used, too many to be listed here.

[2] One ACS cohort member resided in St. Louis, Minnesota. As this is too small a sample size for the purposes of the analyses, this county was also discarded.

Table 3.2. Descriptive Statistics for the Place-specific Covariates

	Metropolitan-scale					County-scale				
	N	min	max	mean	cor†	N	min	max	mean	cor
Sulfates	151	3.6	23.5	10.6	100	132	3.8	24.1	11.2	100
Demographic Data										
Population Change	139	-9.7	28.4	5.7	-40	131	-7.5	23.0	3.4	-43
Race Data										
Whites %	151	57.3	99.2	87.4	-22	132	39.3	99.6	86.7	-15
African Americans %	151	0.1	46.6	10.1	28	132	0.01	55.3	11.1	20
Health Services Data										
Doctors per 100000	138	97	512	206	-13	132	50	931	227	-6
Hospital Beds per 10000	139	222	1424	609	-5	132	0	1567	644	-0
Socio-Economic Data										
High School Educated %	151	41.8	82.7	68.4	-44	132	50.5	83.7	68.3	-49
1979 Income (mean for MAs, median for counties)	151	5700	10360	7363	-13	132	10947	26090	17338	-14
Poverty Rate	151	5.69	20.02	11.12	8	132	4.7	27.6	11.0	18
Income Inequality	151	0.34	0.45	0.39	2	132	0.34	0.49	0.39	9
Unemployment Rate	151	2.7	14.5	6.9	16	132	2.0	12.7	6.2	35

Table 3.2. Descriptive Statistics for the Place-specific Covariates

	Metropolitan-scale					County-scale				
	N	min	max	mean	cor†	N	min	max	mean	cor
Climate Data										
Maximum Temperature mean	135	49.3	87.0	65.4	-12					
Variation in Max. Temp.	135	14.0	144.7	80.1	-6					
Relative Humidity mean	95	17.0	56.4	46.3	59					
Variation in Relative Humidity	95	76.3	312.6	213.0	23					
Physical Environment Data										
Water Hardness	109	9.0	422.0	106.2	-9					
Elevation	110	2	1831	279.9	-49					
Gaseous Co-pollutants										
Carbon Monoxide	107	194.2	3952.3	1561.3	-7					
Nitrogen Dioxide	74	7.8	51.1	25.5	-1					
Ozone	117	10.4	41.1	23.5	2					
Sulfur Dioxide	113	0.02	29.32	9.27	48	132	0.03	27.22	10.7	56

† "Cor" stands for Pearson correlation coefficients, expressed here as percentages.

Table 3.3. Pearson correlations (%) between metropolitan-scale place-specific covariates

	Pop. Change	White	Black	Doctors	Hospital Beds	High School	Mean Income	Poverty rate	Gini Unemploy ment.	Temperature	Var. in Temp.	R. Humidity	Var. in R.H.	H2O Hardness	Altitude	Sulfates	CO	NO_2	O_3	SO_2
Pop. Change		-17	11	8	-26	25	14	19	28	-28	70	-36	-60	-15	-6	27	-40	10	10	27	-43
Whites			-96	-35	-12	24	-12	-45	-51	3	-46	35	-13	-22	27	28	-22	-3	-4	-3	12
Blacks				24	17	32	2	50	49	4	46	27	17	23	-24	-29	28	-7	-7	4	-6
Doctors					32	26	41	-2	20	-36	51	-28	-42	-13	-12	-13	-13	12	21	-14	-7
Hospital Beds						-12	-19	10	5	-9	-7	26	7	26	-9	-14	-5	-12	-20	-8	7
High School							55	-44	-27	-31	-19	17	-23	-35	18	38	-44	15	22	-0	-22
Income								-58	-20	-22	-2	-10	-7	-26	5	1	-13	16	42	-20	-12
Poverty									76	25	50	-24	-1	22	-15	-6	8	-7	-22	6	-16
Gini										9	60	36	12	27	-22	-8	2	10	-1	1	-22
Unemploy.											4	4	15	-4	5	9	16	-27	-23	-13	3
Temperature												-50	-48	-9	5	0	-12	-3	-6	9	-45
Var. in Temp.													-6	10	23	46	-6	-12	-16	-1	34
R. Humidity														32	-16	-73	59	-21	-8	-8	25
Var. in R.H.															-27	-30	23	-7	8	28	12
H2O Hardness																20	-9	-17	-16	3	4
Altitude																	-49	6	7	7	-8
Sulfates																		-7	-1	2	48
CO																			66	5	12
NO_2																				10	21
O3																					-6
SO_2																					

Table 3.4. Pearson correlations (%) between metropolitan-scale place-specific covariates and between county-scale place-specific covariates

MA / County	Pop. Change	White	Black	Doctors	Hospital Beds	High School	Mean Income	Poverty rate	Gini	Unemploy ment.	Sulfates	SO2
Pop. Change		-17	11	8	-26	25	14	19	28	-28	-40	-43
Whites	-4		-96	-35	-12	24	-12	-45	-51	3	-22	12
Blacks	2	-99		24	17	32	2	50	49	4	28	-6
Doctors	3	-43	41		32	26	41	-2	20	-36	-13	-7
Hospital Beds	-14	-31	32	53		-12	-19	10	5	-9	-5	7
High School	19	24	-29	17	4		55	-44	-27	-31	-44	-22
Income	-1	27	-29	-10	-27	47		-58	-20	-22	-13	-12
Poverty	-1	-69	69	31	27	-42	-76		76	25	8	-16
Gini	5	-65	64	51	37	-26	-67	87		9	2	-22
Unemploy.	-41	-12	13	-21	-9	-33	-14	15	15		16	3
Sulfates	-44	-15	20	-7	-0	-49	-14	18	9	35		48
SO_2	-38	-2	-38	14	7	-29	-14	15	17	15	56	

Figure 3.1 US Census Bureau Definitions for Metropolitan Areas

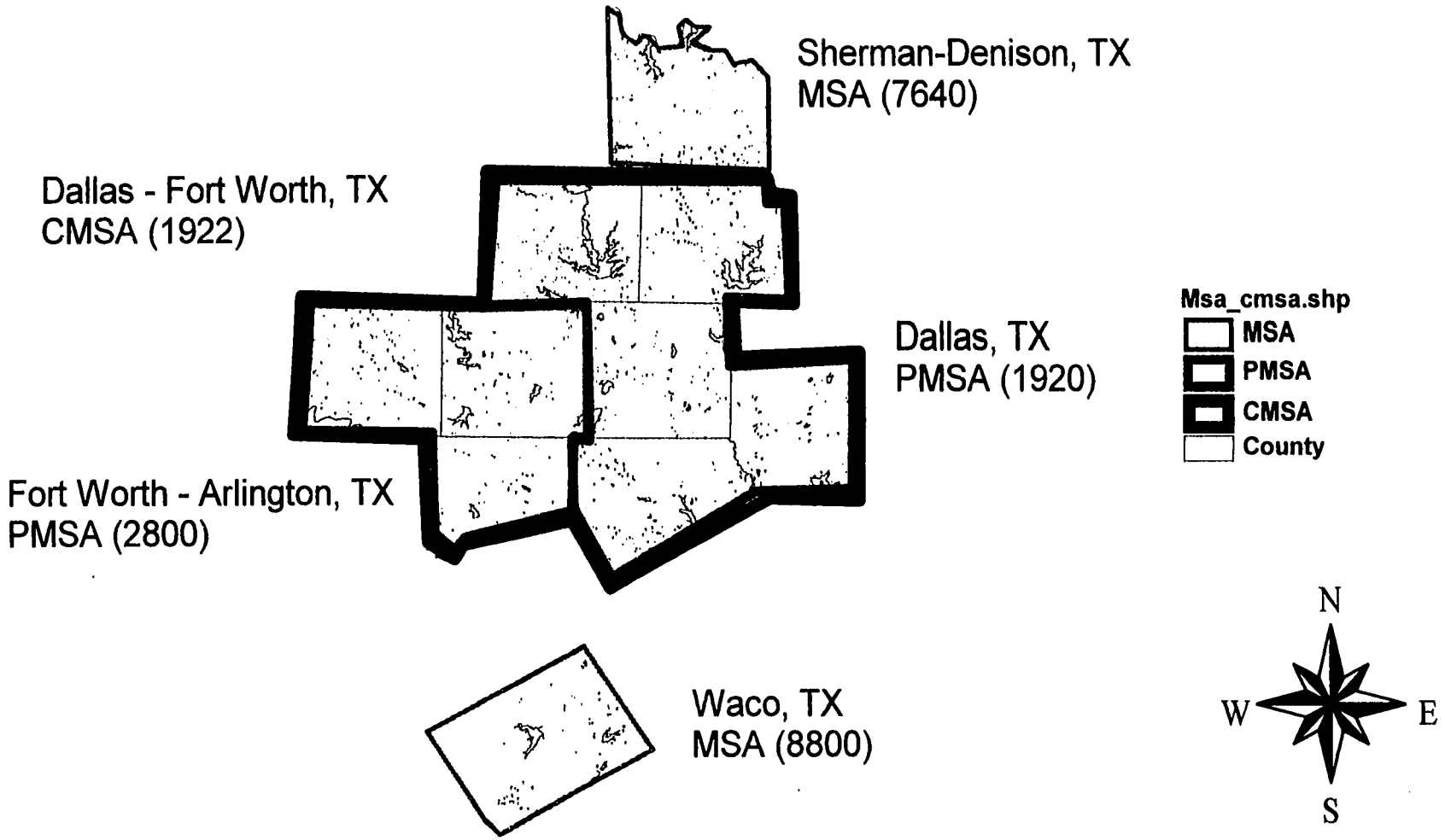
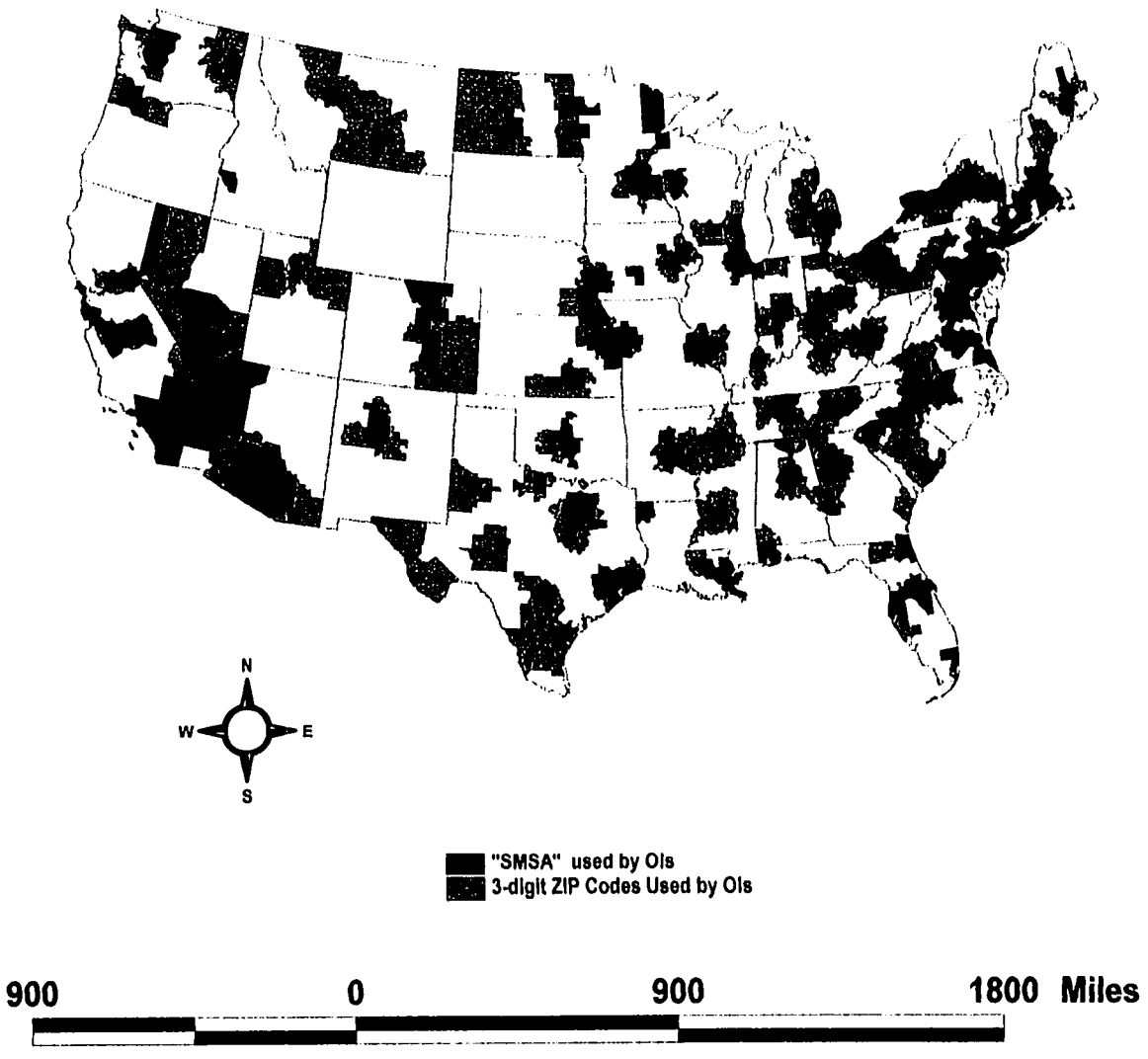


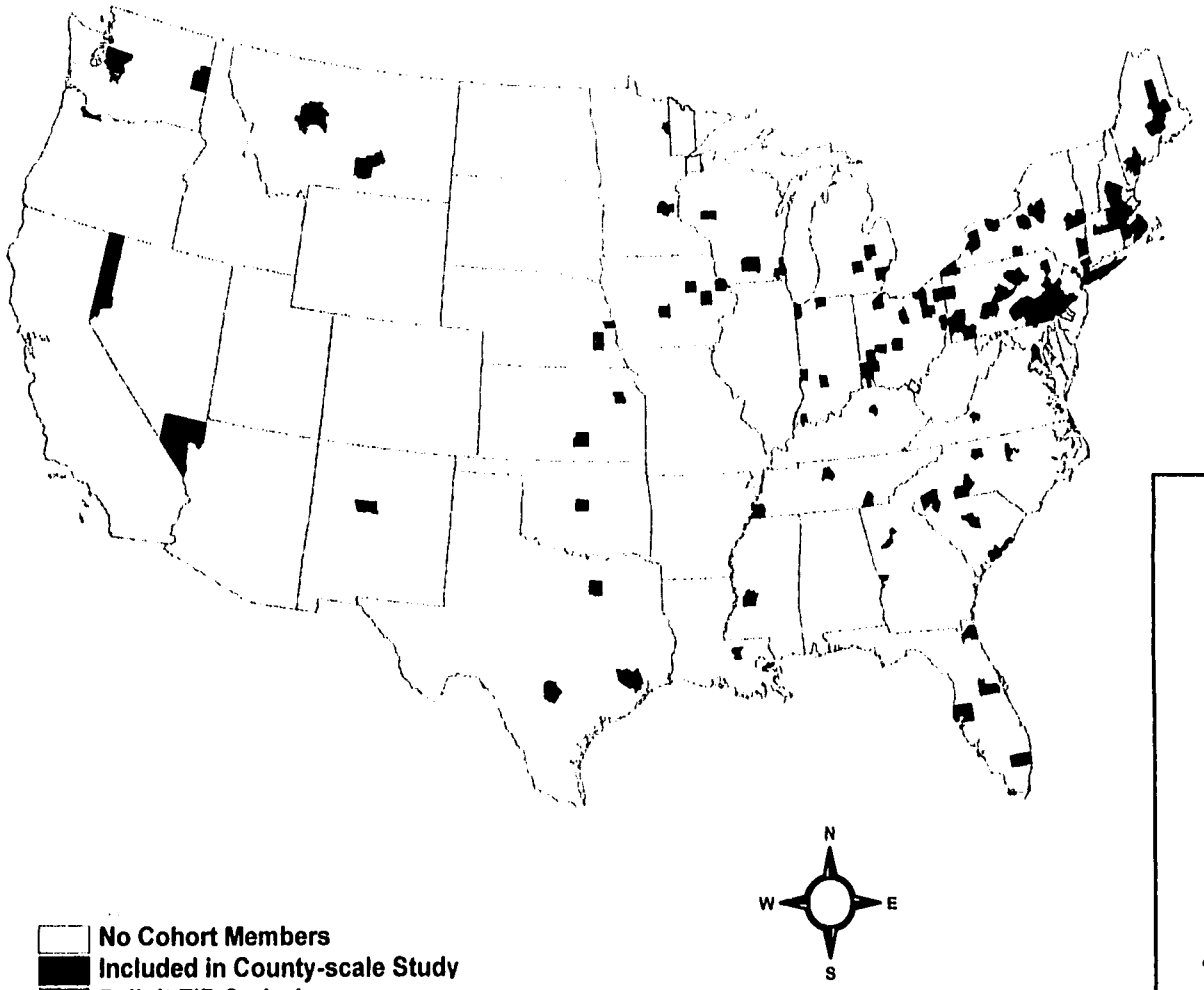
Figure 3.2. The Original Investigators' Mapping of 3-Digit ZIP Codes Onto MAs



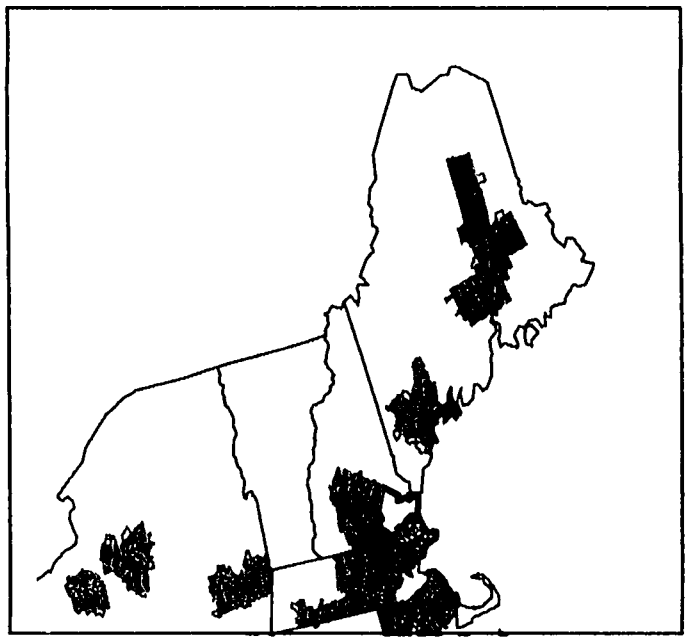
(Note, this map was created on ArcView GIS using data supplied by ESRI, 1999. The MSAs are based on 1990 data and the ZIP boundaries on 1998 data. The boundaries are, therefore, only approximate to those used in the analyses.)

Maine

Figure 3.3. Mapping ZIP Codes Onto Counties



(Note, this map was created on ArcView GIS using data supplied by ESRI, 1999. The counties are based on 1990 data and the ZIP boundaries on 1998 data. The boundaries are, therefore, only approximate to those used in the analyses.)



Maine

4 Incorporation of Place-specific Ecologic Covariates Into Cox Regression

4.1 Cox Proportional Hazards Regression Modelling

While all of the subjects in the CPS-II (and consequently the ACS air pollution study) were recruited in the autumn of 1982, they were followed up for unequal lengths of time. While the investigators were interested in a number of health outcomes, mortality from different causes was of primary interest. For this reason, the Cox proportional hazards model was used in the original analysis of the ACS Study. Because Cox modelling uses a mathematical formula to relate both the occurrence of an event and the time to that event to individuals and their exposure profiles, it is commonly used in longitudinal studies (Kahn and Sempos, 1989) such as the ACS Study.

The Cox proportional hazards model is a form of multivariate regression. Similarly to logistic regression, it is used for binary outcomes (in this case mortality). However, it has a more complicated structure as it is not simply the occurrence of an event which is of interest, but the time it takes to reach that event (Hennekens and Buring, 1987: 319). The coefficients of a Cox model are comparable to those from a logistic model, although they are technically estimates of the incidence density ratio rather than an odds ratio (Hennekens and Buring, 1987). The general form of the Cox proportional hazards model can be expressed as:

$$\ln[\text{incidence rate } (t)] = a(t) + b_1X_1 + \ldots + b_n X_n,$$

where a(t) is the baseline incidence rate, X_i are the independent variables and b_i are the coefficients for each independent covariate.

4.2 Cox Modelling of the ACS Cohort

CPS-II Variables

Information on a number of individual-level risk factors was collected by ACS at the time of enrolment through self-administered questionnaires. Not all of these variables were used by the Original Investigators. A subsequent data-audit found minimal errors in the coding of this information and work done for the Re-analysis (Health Effects Institute, 2000) and for this thesis made use of the majority of available individual-level risk factors. Cox models were stratified by 1-year age groups, race and gender. The following tobacco consumption variables were used (* indicates an indicator variable): current smoker*, years of smoking, cigarettes per day, current smoker pack-years, former smoker*, former smoker pack-years, and age current smokers started smoking (at 18 or older or below 18)*. An indicator variable was also used to distinguish those people who had a high school diploma, and those who had more than a high school education. Exposure to dust or fumes at work*, body mass index (and body mass index squared), marital status (married, separated, or widowed versus single)*, and any alcohol consumption (beer, wine and liquor consumption)* were also included in the Cox regression model. Finally, interaction terms between gender and the following variables were included: indicator variables for marital status, and indicators for alcohol consumption (interaction terms between gender and all other variables were run by the

Re-analysis Team, however, these variables were the only ones with statistically significant interaction terms, Health Effects Institute, 2000). All of the Cox models evaluated relative risk over the range of the variables (see Table 3.2) and used calendar years as the time axis.

Place-specific Ecologic Covariates

Although Pope et al (1995) did assign individual members of the ACS cohort to specific metropolitan areas, the only metropolitan-specific variable they used was ambient air pollution. For the sake of comparison, in order to investigate whether the relationship the Original Investigators observed between sulfate levels and mortality in the 151 MAs was confounded by other metropolitan-specific ecologic covariates which varied between cities in a similar way, the ecologic covariates outlined in chapter 2 were entered into the Cox model. Since the Cox model can only accommodate individual-level data, the ecologic covariates had to be treated as individual attributes. This was done by assigning values for each variable to individuals in the cohort according to which MA they had been identified as residing within. The Original Investigators used this approach to incorporate the pollution values into their Cox models. Following the Original Investigators, the rate ratios for each of the ecologic covariates, including sulfates, were evaluated over the range of that covariate[1] (Table 3.2 lists the minimum

---

1 The risk ratios were calculated as follows:

$RR = e^{(Range)*\beta}$ The 95% confidence intervals were calculated as follows:

$CI = e^{(Range*(\beta \pm 1.96*SE)}$

and maximum values for each variable).

Multiple Covariate Analyses

In order to examine whether groups of place-specific covariates had an effect beyond that of single covariates, three multiple covariate analyses were conducted. First, a model was run with all the gaseous co-pollutants: carbon monoxide, nitrogen dioxide, ozone and sulfur dioxide. Second, all of the socio-economic place-specific covariates were run along with population change to verify that population change had an effect on its own beyond that of socio-economic factors. Third, all of those place-specific covariates which individually changed the effect of sulfates by 25% or more were entered into a Cox model. Unfortunately, because data were missing in different metropolitan areas for different covariates, the multiple covariate analyses often were run with a reduced subset of MAs (for example, data were only available for 58 out of 151 MAs for all of the gases). Because a different effective sample of cities was used, these results are not directly comparable. However, they do still provide an indication of the amount of the sulfate effect that can be accounted for by other place-specific covariates.

4.3 Results

Before examining the impact the place-specific covariates had on the association between sulfates and mortality, a series of analyses were run with the place-specific covariates alone. Using the Cox model, it can be seen that ambient sulfate levels are not the only place-specific variables to be significantly associated with all-cause, cardiopulmonary or lung cancer mortality (see Tables 4.1, 4.2 and 4.3). Looking at the

results from the sulfate cohort of 151 cities, increase in population, high school education, income inequality, all appear to be significantly protective of both all-cause and cardiopulmonary mortality (see Tables 4.1 and 4.2). These associations reflect those reported in the literature (see chapter 2). Also as expected from the literature, an increase in the unemployment rate is associated with higher risks of both all-cause and cardiopulmonary mortality (see Tables 4.1 and 4.2).

However, contrary to expectation, an increase in income inequality appears to lower the risk of either all-cause or cardiopulmonary mortality (see Tables 4.1 and 4.2); an increase in $CaCO_3$ concentrations appears to increase the risk of cardiopulmonary or all-cause mortality (see Tables 4.1 and 4.2); and both NO_2 and O_3 appear to be protective from all-cause mortality (see Table 4.1). Other discrepancies, such as the increased risk of all-cause or cardiopulmonary mortality associated with increases in hospital beds (see Table 4.2) may be explained by reverse causation: more hospital beds may be opened up to serve sicker areas.

Interestingly, none of the demographic or socio-economic covariates was found to be associated with an increased risk of mortality from lung cancer in the Cox model that controls for smoking at the individual level (see Table 4.3). However, the physical environment did seem to have an influence on lung cancer mortality. Specifically, areas with increased elevation or increased O_3 demonstrated significantly lower risk of lung cancer mortality, while areas with higher relative humidity experienced a significantly raised risk of lung cancer mortality (see Table 4.3).

When sulfates were added to the Cox models, most of the ecologic covariates

remained significant. However, only two reduced the risk from sulfates to any great degree: population change and SO_2. Both of these variables reduced the sulfate effect to insignificance for all-cause mortality (see Table 4.1). The effect of sulfates on cardiopulmonary mortality was slightly more robust, being halved by the incorporation of population change or SO_2 but remaining significant (see Table 4.2). Altitude was the only variable to reduce the sulfate effect on lung cancer mortality (see Table 4.3).

As was mentioned above, full ecologic data were not available for all of the metropolitan areas. Consequently, different subsets of cities had to be used when analysing different ecologic covariates (the number of cities involved in each analysis are listed on the left of Tables 4.1, 4.2 and 4.3). Simply changing the set of cities being examined had an impact on the size of the risk presented by ambient sulfate pollution. For example, in the full set of 151 cities there was an all-cause mortality RR of 1.15 from sulfates alone, whereas in the subset of 110 cities with altitude data, sulfates only had an RR of 1.10.

Problems With Cox Models and Ecologic-level Variables

Although confidence intervals have been mentioned, it bears noting that because of the nature of the Cox model, achieving statistical significance is not that difficult. The Cox model used here treats each of the ecologic covariates as individual attributes. That is to say, rather than treating the sulfate cohort as having 151 observations, it treats it as having 552,138, essentially implying a much larger effective sample size than is actually the case.

By treating place-specific ecologic covariates as individual-level attributes,

these Cox models violate the assumption of independence. In other words, by treating everyone in the same metropolitan area as having the same exposure to a place-specific variable, the Original Investigators are indicating that the individuals in the metropolitan areas are not really independent. While the Cox model is an ideal way to incorporate individual-level data and the domain of time, it does so by sacrificing the domain of space. In order to accommodate the spatial elements inherent in the study of ambient air pollution, investigators must go beyond the simple Cox model. In the next chapter, I outline a two-stage approach which takes into account the dependence of observations on individuals within the same metropolitan area.

Table 4.1 Relative risk of all-cause mortality associated with sulfates after adjustment for metropolitan-specific ecological covariates using the Cox proportional hazards model.

Ecologic Covariate	Relative Risk from Sulfate		Relative Risk from Ecologic Covariate	
	Extended Model	Extended Model With Ecologic Covariate	Extended Model Without Sulfates	Extended Model With Sulfates
Sulfates Alone		1.15 *(1.08,1.21)*		
Demographic Data				
Population Change	1.15	1.06 *(0.99, 1.13)*	0.85 *(0.81, 0.89)*	0.87 *(0.82, 0.91)*
Race Data				
Whites(%)	1.15	1.18 *(1.11, 1.25)*	1.02 *(0.98, 1.06)*	1.06 *(1.02, 1.11)*
African Americans (%)	1.15	1.17 *(1.10, 1.24)*	1.01 *(0.96, 1.06)*	0.96 *(0.91, 1.01)*
Health Services Data				
Doctors per capita	1.14	1.14 *(1.08, 1.21)*	0.95 *(0.89, 1.01)*	0.96 *(0.90, 1.01)*
Hospitalbeds	1.14	1.13 *(1.07, 1.20)*	1.13 *(1.06, 1.21)*	1.12 *(1.04, 1.19)*
Socio-Economic Data				
High School Educated (%)	1.15†	1.13 *(1.06, 1.20)*	0.91 *(0.86, 0.96)*	0.96 *(0.90, 1.02)*
Income 1979 (mean)	1.15	1.15 *(1.08, 1.21)*	0.93 *(0.88, 0.97)*	0.93 *(0.88, 0.97)*
Povertyrate	1.15	1.15 *(1.09, 1.22)*	0.95 *(0.91, 1.00)*	0.94 *(0.90, 0.99)*
GiniCoefficient	1.15	1.15 *(1.09, 1.21)*	0.88 *(0.84, 0.93)*	0.88 *(0.83, 0.93)*
Unemployment rate	1.15	1.12 *(1.06, 1.19)*	1.12 *(1.06, 1.19)*	1.09 *(1.03, 1.16)*
Climate Data				
Temperature (max.)	1.14	1.10 *(1.04, 1.17)*	0.88 *(0.85, 0.92)*	0.90 *(0.86, 0.94)*
Variation in temperature (max.)	1.14	1.10 *(1.04, 1.17)*	1.18 *(1.11, 1.24)*	1.16 *(1.09, 1.22)*
Relative humidity	1.13	1.14 *(1.05, 1.24)*	1.05 *(0.99, 1.12)*	0.98 *(0.91, 1.06)*
Variation in relative humidity	1.13	1.16 *(1.08, 1.25)*	0.96 *(0.90, 1.02)*	0.92 *(0.86, 0.98)*
Physical Environment Data				
WaterHardness	1.11	1.12 *(1.05, 1.19)*	1.08 *(1.02, 1.13)*	1.08 *(1.03, 1.14)*
Altitude	1.10	1.16 *(1.08, 1.24)*	1.05 *(0.99, 1.12)*	1.12 *(1.04, 1.19)*
Gaseous Pollutants				
CO	1.14	1.14 *(1.08, 1.21)*	0.98 *(0.92, 1.03)*	0.97 *(0.92,1.03)*
NO_2	1.11	1.14 *(1.06, 1.22)*	0.93 *(0.89, 0.98)*	0.91 *(0.87,0.96)*
O_3	1.15	1.15 *(1.09, 1.22)*	0.93 *(0.87, 0.99)*	0.92 *(0.86,0.98)*
SO_2	1.16	1.04 *(0.98, 1.11)*	1.30 *(1.23, 1.38)*	1.28 *(1.20,1.37)*
Multiple Covariate Analyses				
Gases	1.09	1.00 *(0.93, 1.09)*		
SES	1.15	1.10 *(1.02, 1.18)*		
25% Change‡	1.10	0.99 *(0.90, 1.09)*		

‡ population change, variation in maximum temperature, altitude SO_2

Table 4.2. Relative risk of cardiopulmonary mortality associated with sulfates evaluated after adjustment for metropolitan-specific covariates using the Cox proportional hazards model

Ecologic Covariate	Relative Risk from Sulfate			Relative Risk from Ecologic Covariate		
	Extended Model	Extended Model With Ecologic Covariate		Extended Model Without Sulfates	Extended Model With Sulfates	
Sulfates Alone		1.25	*(1.16, 1.36)*			
Demographic Data						
Population Change	1.24	1.12	*(1.03, 1.23)*	0.80 *(0.75, 0.85)*	0.83	*(0.77, 0.90)*
Race Data						
Whites (%)	1.25	1.30	*(1.20, 1.42)*	1.02 *(0.97, 1.08)*	1.09	*(1.03, 1.16)*
African Americans (%)	1.25	1.28	*(1.17, 1.40)*	1.03 *(0.96, 1.10)*	0.95	*(0.88, 1.03)*
Health Services Data						
Doctors per capita	1.24	1.23	*(1.14, 1.34)*	0.86 *(0.79, 0.94)*	0.87	*(0.80, 0.95)*
Hospital beds	1.24	1.22	*(1.13, 1.33)*	1.18 *(1.07, 1.30)*	1.15	*(1.04, 1.26)*
Socio-Economic Data						
High School Educated (%)	1.25	1.21	*(1.11, 1.32)*	0.84 *(0.77, 0.92)*	0.92	*(0.83, 1.00)*
Income 1979 (mean)	1.25	1.25	*(1.16, 1.36)*	0.87 *(0.81, 0.93)*	0.87	*(0.81, 0.94)*
Poverty rate	1.25	1.26	*(1.16, 1.36)*	0.99 *(0.93, 1.05)*	0.97	*(0.91, 1.04)*
Gini Coefficient	1.25	1.25	*(1.16, 1.36)*	0.90 *(0.83, 0.97)*	0.90	*(0.83, 0.97)*
Unemployment rate	1.25	1.20	*(1.11, 1.30)*	1.27 *(1.16, 1.38)*	1.21	*(1.11, 1.32)*
Climate Data						
Temperature (max.)	1.22	1.19	*(1.09, 1.29)*	0.87 *(0.46, 1.64)*	0.89	*(0.83, 0.95)*
Variation in temperature (max.)	1.22	1.17	*(1.08, 1.28)*	1.26 *(1.16, 1.36)*	1.22	*(1.12, 1.32)*
Relative humidity	1.21	1.20	*(1.06, 1.35)*	1.11 *(1.01, 1.21)*	1.01	*(0.91, 1.13)*
Variation in relative humidity	1.21	1.28	*(1.15, 1.42)*	0.93 *(0.85, 1.01)*	0.86	*(0.79, 0.95)*
Environmental Data						
Water Hardness	1.19	1.20	*(1.09, 1.31)*	1.11 *(1.03, 1.20)*	1.12	*(1.04, 1.21)*
Altitude	1.20	1.26	*(1.14, 1.39)*	1.02 *(0.93, 1.12)*	1.12	*(1.02, 1.23)*
Gaseous Co-pollutants						
CO	1.27	1.27	*(1.17, 1.38)*	0.94 *(0.87, 1.02)*	0.94	*(0.86, 1.01)*
NO_2	1.25	1.29	*(1.17, 1.42)*	0.93 *(0.87, 1.01)*	0.89	*(0.83, 0.97)*
O_3	1.26	1.27	*(1.16, 1.38)*	0.98 *(0.90, 1.08)*	0.96	*(0.88, 1.05)*
SO_2	1.28	1.14	*(1.04, 1.25)*	1.40 *(1.29, 1.52)*	1.33	*(1.22, 1.46)*
Multiple Covariate Analyses						
Gases	1.22	1.11	*(1.00, 1.25)*			
SES	1.24	1.18	*(1.06, 1.31)*			
25% change†	1.20	1.17	*(0.97, 1.41)*			

† population change, variation in relative humidity

Table 4.3. Relative risk of lung cancer mortality associated with sulfates after adjustment for selected metropolitan-specific covariates using the Cox proportional hazards model.

Ecologic Covariate	Relative Risk from Sulfate			Relative Risk from Ecologic Covariate			
	Extended Model	Extended Model With Ecologic Covariate		Extended Model Without Sulfates		Extended Model With Sulfates	
Sulfates Alone		1.33	*(1.10, 1.61)*				
Demographic Data							
Population Change	1.31	1.30	*(1.05, 1.61)*	0.91	*(0.78, 1.06)*	1.00	*(0.84, 1.18)*
Race Data							
Whites (%)	1.33	1.34	*(1.10, 1.64)*	0.96	*(0.84, 1.10)*	1.03	*(0.89, 1.19)*
African Americans (%)	1.33	1.35	*(1.10, 1.66)*	1.05	*(0.89, 1.23)*	0.95	*(0.80, 1.13)*
Health Services Data							
Doctors per capita	1.31	1.30	*(1.07, 1.57)*	0.91	*(0.74, 1.11)*	0.93	*(0.76, 1.13)*
Hospital beds	1.31	1.32	*(1.09, 1.60)*	0.92	*(0.73, 1.16)*	0.89	*(0.71, 1.13)*
Socio-economic Data							
High School Educated (%)	1.33	1.34	*(1.09, 1.65)*	0.90	*(0.74, 1.10)*	1.02	*(0.82, 1.27)*
Income 1979 (mean)	1.33	1.33	*(1.10, 1.61)*	1.01	*(0.86, 1.19)*	1.02	*(0.86, 1.20)*
Poverty rate	1.33	1.33	*(1.10, 1.61)*	0.97	*(0.83, 1.13)*	0.96	*(0.82, 1.12)*
Gini Coefficient	1.33	1.32	*(1.09, 1.60)*	0.88	*(0.73, 1.07)*	0.89	*(0.73, 1.08)*
Unemployment rate	1.33	1.31	*(1.08, 1.60)*	1.13	*(0.92, 1.38)*	1.05	*(0.86, 1.29)*
Climate Data							
Temperature (max.)	1.37	1.36	*(1.11, 1.67)*	0.94	*(0.81, 1.09)*	0.98	*(0.84, 1.15)*
Variation in temperature (max.)	1.37	1.39	*(1.13, 1.71)*	1.00	*(0.83, 1.21)*	0.94	*(0.77, 1.14)*
Relative humidity	1.53	1.39	*(1.04, 1.86)*	1.37	*(1.10, 1.72)*	1.17	*(0.90, 1.53)*
Variation in relative humidity	1.53	1.49	*(1.16, 1.92)*	1.19	*(0.96, 1.47)*	1.09	*(0.88, 1.37)*
Environmental Data							
Water Hardness	1.31	1.31	*(1.05, 1.62)*	0.94	*(0.78, 1.12)*	0.94	*(0.79, 1.13)*
Altitude	1.24	1.14	*(0.91, 1.44)*	0.72	*(0.56, 0.93)*	0.76	*(0.58, 0.99)*
Gaseous Co-pollutants							
CO	1.26	1.26	*(1.03, 1.54)*	0.82	*(0.68, 1.00)*	0.82	*(0.68, 1.00)*
NO_2	1.28	1.33	*(1.05, 1.69)*	0.91	*(0.76, 1.09)*	0.87	*(0.73, 1.05)*
O_3	1.27	1.30	*(1.07, 1.59)*	0.74	*(0.59, 0.92)*	0.72	*(0.58, 0.90)*
SO_2	1.31	1.36	*(1.08, 1.72)*	1.06	*(0.87, 1.30)*	0.93	*(0.74, 1.17)*
Multiple Covariate Analyses							
SES	1.31	1.20	*(0.94, 1.54)*				
gases	1.42	1.61	*(1.21, 2.14)*				
25% change†	1.53	1.39	*(1.04, 1.86)*				

† relative humidity

5 Using Random Effects Regression Models to Incorporate Different Scales

The Cox regression outlined in chapter four incorporates place-specific ecologic-level covariates into the analysis as individual attributes. As each individual within a metropolitan area is given the same attribute for the ecologic covariate, it is apparent that these observations are not independent.

Most epidemiological studies in the past have remained at one scale, either at the individual level or at some ecological level. Ecologic-level studies cannot take advantage of individual-level data properly, nor can individual-level studies incorporate ecologic-level data without the risk of biased results. In this chapter, I introduce a method developed by the Re-analysis Team (Health Effects Institute, 2000) to incorporate both individual-level and ecologic-level data into the same analysis while treating each kind of data appropriately.

The two-stage approach, outlined below allows the use of both individual and ecologic data; however, it does not solve the issue of what scale the ecologic-level data should be measured on. In the US, census data are collected for individuals but are usually only available to the public for census tracts. Figure 5.1 shows Jefferson County, Alabama broken down into its component census tracts. These census tracts can then be aggregated up to counties. The US Census Bureau then defines metropolitan statistical

areas as being made up of one or more counties; Huntsville, AL is made up of one county, Huntsville County, while Birmingham, AL is comprised of five counties. Counties could be further aggregated into states (the full extent of the state of Alabama is not shown in Figure 5.1). Census tracts, counties, metropolitan areas, and states all represent different scales of analysis. The availability of air pollution data limited which of these scales could be used to examine the relationship between sulfates and mortality in the ACS Study. For example, census tracts are too fine a geographic unit for the number of sulfate monitors. Here, counties, metropolitan areas and airshed regions were explored.

Most importantly, the two-stage regression allows for clustering of mortality within cities, therefore it includes a city-specific effect on mortality, treating observations within cities as dependent.

5.1 Two-stage Regression Methods

The Re-analysis Team (Health Effects Institute, 2000) developed a two-stage method to integrate an individual-level Cox regression model with an ecologic-level linear regression model (see Table 5.1). The objective of the first stage is to estimate city-specific mortality rates (in logarithm form), which will be used as the dependent variable in stage two, and to estimate within-city estimation errors. Stage one assumes that the individual observations are independent given the random location effects. In stage two, after adjusting for place-specific ecologic covariates, the logarithms of the city-specific mortality rate are assumed to be distributed statistically so that expectation is

given by a linear regression model consists of place-specific variables, with variance defined as the sum of the within-place estimation error and the true variation in the adjusted logarithms of the city-specific mortality rates.

In the first stage, individual-level data (see section 4.2) and an indicator function for the subject's city of residence (compared to an index city) were entered into a Cox proportional hazards model. The index city used in this case was Greenville, SC which had a sulfate concentration of 11 $\mu g/m^3$ which was close to the mean of all 151 metropolitan areas. The Re-analysis Team conducted runs with different index cities and found that the choice of index city, did not alter the results (Health Effects Institute, 2000).

The use of an index MA does mean that the estimates of the place-specific mortality rates are by definition correlated which causes certain problems: the covariance between the MA-specific estimates increases, and the variance in each estimate is inflated (Health Effects Institute, 2000). Finally, the Cox model in stage one is not able to estimate the variance of the index city (Health Effects Institute, 2000). To deal with these problems, the Re-analysis Team used a simplified version of an adjustment outlined by Easton et al. (1991). This method calculates an adjusted variance for each MA as the difference between the variance and the average of the covariances between estimates of the logarithms of the city-specific adjusted mortality rates. The variance of the index city is simply estimated by the average of the covariances of the other cities.

Because the actual mortality rates for cohort members in each MA are unknown, there is uncertainty inherent in the estimates obtained in Stage one. This

uncertainty is incorporated into the second stage by estimating the variance of the logarithm of each city-specific mortality rate as the sum of the true variation in rates between cities and the estimation error for each individual city. Using a method outlined by DerSimonian and Laird (1986), an initial estimate of the true variation in mortality between MAs can be estimated in stage two. This is done before any adjustment is made for city-specific effects, depending only on the mean and variance structure of the city-specific mortality rates estimated in stage one.

Stage two involves two iterations of a weighted least squares regression model (see Table 5.1). The first set of weights are calculated for each city as the inverse of the sum of the initial estimate of true variance between MAs (obtained using DerSimonian and Laird's technique with no place-specific covariates) and the MA-specific estimation variances (obtained from stage one and adjusted according to Easton). This weighting takes into account the statistical uncertainty in the estimation of the mortality rates. The logarithms of these weighted mortality rates are then regressed on sulfates and other appropriate ecologic-level covariates using weighted least squares regression methods.

Taking the residuals from this weighted linear regression, an updated estimate of the true variance between cities is calculated. A second set of weights are calculated as the inverse of the sum of the second estimate of the true variance and the MA-specific estimation variances from stage one, and a second weighted linear regression model is fit.

Using this weighting scheme, the two-stage regression model allows for mortality rates to cluster by metropolitan area, therefore acknowledging that within-area observations are not independent. The two-stage method does, however, treat the

metropolitan area-specific mortality rates as statistically independent. Therefore, the two-stage method, in and of itself, cannot account for processes that operate on a broader regional scale, this is discussed further in section 5.2.

County Analysis

Because of the inherent problems in using a Cox model to study the effects of place-specific ecologic covariates, no Cox models were run with the county level data. However, for the sake of comparison to the metropolitan Cox models outlined in chapter 4, a two-stage analysis was conducted with the county data which assumed independence of the within-county observations. This was simply done by assigning a zero to the between county variation and the residual variance to the errors of the within-county estimations. This method has been shown to produce results similar to the Cox model outlined in chapter 4 (Health Effects Institute, 2000). Like the Cox regression model using ecologic data with this method also violates the assumption that the exposures of the people within an county are not related.

A proper two-stage random effects analysis was then run with the county data allowing for clustering of mortality within the counties.

Metropolitan-scale Results

First, ecologic-level regression analyses were run with the metropolitan-specific variables as independent variables and the metropolitan-specific relative risks obtained through the first stage Cox regression which did not allow for within area clustering (see the results for independent observations in Tables 5.2a&b, 5.3a&b and 5.4a&b). As expected, these results were very similar to those obtained through entering

the ecologic variables into the Cox regression (see Tables 4.1, 4.2, and 4.3).

Using the model outlined above, rate ratios were produced which treat observations within metropolitan areas as dependent but those between different metropolitan areas as independent (see independent MAs in Tables 5.2a&b, 5.3a&b and 5.4a&b). Consistently larger rate ratios were produced using this method than when using the Cox models outlined in chapter 4. For example, the risk of all-cause mortality from sulfates with no other ecologic covariates increased from 1.15 to 1.25 when the random effects model was used. Population change and sulfur dioxide continued to influence the relationship between sulfates and all-cause mortality. However, as the rate ratios were higher to begin with, these two variables no longer decreased the relative risk from sulfates to the point of insignificance. When SO_2 and population change were included together, the confidence interval for sulfates did include 1. Likewise, including all of the gases as covariates diminished the relative risk from sulfates to the point of insignificance.

A similar pattern is seen in the relative risk presented by sulfate exposure to mortality from cardiopulmonary causes. The rate ratio from sulfates alone is 1.29 and population change and SO_2 exposure diminish that rate ratio substantially but not to insignificance. While mean relative humidity might appear to have a dramatic effect on sulfates, an examination of Table 5.3a shows that the effect of sulfates alone in that particular subset of 95 cities is already quite a bit lower than the rate ratio in the full 151 cities (1.21 and 1.29 respectively). Again, when all of the gases are included as covariates the relative risk from sulfates becomes insignificant. The same is true when

population change, SO_2 and education (all of which changed the sulfate effect by at least 25%) are entered into a model with sulfates: in this case, the mortality rate ratios for sulfates drops from 1.34 to 1.07 and is not significant.

The RR from sulfate exposure for lung cancer is 1.39 in the independent MAs model. The only covariate which has an impact on this effect is altitude which, when included with sulfates, leads to a sulfate effect of 1.13 that is not significant. When altitude and relative humidity are both included, the RR is also 1.39, however it is no longer significant. In the case of lung cancer, including all the socio-demographic variables decreases the RR from sulfates to an insignificant 1.23. Including all the gas data actually increases the RR from sulfates to 1.63 from 1.48.

Because data were missing for some variables in some metropolitan areas, different subsets of cities were used for different groups of variables. In part, the variation in rate ratios outlined above is due to the different metropolitan areas being included. The cardiopulmonary deaths rate ratio from sulfates alone between the different subsets used varied from 1.30 (1.07, 1.58), for the subset of 110 areas with altitude data, to 1.49 (1.23, 1.81), for the 74 MAs with nitrogen dioxide information. The sulfate effect for all 151 metropolitan areas was 1.38 (1.17, 1.62). Similar variation was seen for all-cause mortality and for lung cancer mortality.

County-scale Results

When the relative risks are calculated for counties rather than for metropolitan areas and the ecologic-level data are collected for counties as well, the rate ratios for mortality from sulfate exposure become higher than those for metropolitan areas (see

Tables 5.5 and 5.6). Using the random effects model, the RR of all-cause mortality from sulfate exposure alone is 1.50 (1.30, 1.73), as compared to 1.25 (1.13, 1.37) for the metropolitan-scale analysis. A similar pattern is seen for cardiopulmonary mortality with a sulfate RR of 1.75 (1.48, 2.08) at the county scale compared to 1.29 (1.15, 1.46) at the metropolitan scale. Because lung cancer mortality is quite low in some counties (with no deaths observed in three counties), the two-stage random effects model could not be used as it became unstable.

The results of the county-scale independent observations model (in which the variance was not modified) were quite similar to those resulting from the model which allows for clustering within the counties: 1.47 (1.32, 1.65) for all-cause mortality and 1.72 (1.47, 2.01) for cardiopulmonary mortality.

At the county scale, the risk posed by sulfates appears to be higher; it also appears to be more robust to the inclusion of county-specific ecologic covariates. No ecologic covariates reduced the RR for either all-cause or cardiopulmonary mortality from sulfate exposure to the level of insignificance. Population change was the only variable to reduce the RR of all-cause mortality from sulfates by more than 25 % in the independent observations model. No ecologic covariates had this much of an impact on the RR for all-cause mortality from sulfates in the independent counties model. Population change also had a larger than 25% effect on the RR for cardiopulmonary mortality, along with unemployment rate. Even when these two variables were run in a model together with sulfates, the relative risk, though reduced, was still significant: 1.41 (1.17, 1.70) in the independent observations analysis, and 1.44 (1.18, 1.75) in the

independent counties analysis.

Overall, at the county scale, the observed effect of sulfates on both all-cause and cardiopulmonary mortality is higher and more robust against the inclusion of ecologic covariates in the analysis than the effect observed at the metropolitan scale. However, the two scales are not directly comparable since data at the county scale was not available for every county that comprised the metropolitan areas, therefore, the two analyses represent slightly different data sets.

5.2 Seven-Region Analysis

The random effects model in and of itself is only able to accommodate dependence amongst observations within each of the metropolitan areas. In order to deal with another scale of dependence, namely that due to broad regional trends, the US was split into 7 regions using the EPA's airsheds (as interpreted by Samet et al., 2000). Using ArcView, the metropolitan areas were matched up with regions and indicator variables were then assigned to each MA (see Figure 5.2). These indicator variables were entered into the analysis at the ordinary least squares stage.

The regional analysis could not be conducted with the county data because of the paucity of counties in the western regions (see Figure 5.3). Only one county is located in the Southern California region, three in the Southwest, and six in the Northwest, which is not sufficient for a stable analysis.

Results

Overall, when broad regional trends are controlled for, the all-cause mortality

RR from sulfates decreases to 1.19 (see Table 5.2). In the regional analysis, while including SO_2 still reduces the RR from sulfates down to insignificance, population change no longer has an impact. In fact, when a variable is introduced which controls for region, population change no longer has an effect on the relationships between sulfates and either all-cause or cardiopulmonary mortality (see Tables 5.2 and 5.3). Controlling for region, however, does not decrease the impact of SO_2 on the relationship between sulfates and mortality. Although the other gases have no observable effect on their own, when they are entered into a model together, they reduce the RR from sulfates on all-cause mortality from 1.25 to 1.06 (0.90, 1.26). Likewise, with cardiopulmonary mortality all the gases combined led to a reduction from 1.36 to 1.15 (0.93, 1.42).

The regional adjustment models do offer some control for spatial processes operating on the scale of the somewhat arbitrarily determined EPA airsheds. The use of broad regions also represents a rudimentary way of dealing with spatial autocorrelation. More sophisticated methods of controlling for spatial autocorrelation and of choosing a more appropriate scale of analysis are outlined in the following chapter.

Table 5.1. The Two-Stage Method of Random Effects Modeling.

	Stage	
	One	Two
Unit of Analysis	• Individual	• Area
Assumption	• Independence of Observations	• Clustering of mortality by area
Dependent Variable	• Deaths and times to death	• ln area-specific mortality rates
Independent Variable	• Individual risk factors	• Place-specific risk factors
Method of Analysis	• Cox proportional hazards	• Weighted linear regression
Output	• Place-specific mortality rate estimates • Adjusted within area variance estimates	• Rate ratio estimates of the relative risk associated with place-specific covariates

Table 5.2a. Impact of metropolitan area-specific covariates on the all-cause mortality rate ratio associated with sulfate exposure, using a two-stage random effects approach.

		Independent MAs			Regional Adjustment		
Ecologic Covariate	n	Sulfate Alone	Sulfate with Covariate	Covariate and Sulfates	Sulfate Alone	Sulfate with Covariate	Covariate and Sulfate
Sulfates Alone	151		1.25** (1.13, 1.37)			1.19 (1.06, 1.34)	
Demographic							
Population Change	139	1.24 (1.13, 1.37)	1.16** (1.05, 1.29)	0.88 (0.80, 0.96)	1.18 (1.04, 1.34)	1.17 (1.02, 1.33)	0.94 (0.84, 1.05)
Race							
Whites %	151	1.25 (1.13, 1.37)	1.27* (1.15, 1.39)	1.06 (0.99, 1.14)	1.19 (1.06, 1.34)	1.2 (1.06, 1.36)	1.03 (0.95, 1.10)
African Americans %	151	1.25 (1.13, 1.37)	1.26** (1.15, 1.39)	0.96 (0.88, 1.04)	1.19 (1.06, 1.34)	1.19 (1.06, 1.35)	0.99 (0.90, 1.09)
Health Services							
Doctors per 100000	138	1.24 (1.13, 1.37)	1.23** (1.12, 1.36)	0.92 (0.84, 1.02)	1.18 (1.04, 1.34)	1.17 (1.04, 1.33)	0.93 (0.85, 1.02)
Hospital Beds per 10000	139	1.24 (1.13, 1.37)	1.24** (1.13, 1.36)	1.13 (1.02, 1.25)	1.18 (1.04, 1.34)	1.19 (1.05, 1.35)	1.13 (1.02, 1.25)
Socio-Economic							
High School Educated %	151	1.25 (1.13, 1.37)	1.20** (1.08, 1.33)	0.92 (0.82, 1.02)	1.19 (1.06, 1.34)	1.16 (1.02, 1.33)	0.91 (0.80, 1.03)
1979 Mean Income	151	1.25 (1.13, 1.37)	1.23** (1.12, 1.35)	0.91 (0.83, 0.99)	1.19 (1.06, 1.34)	1.18 (1.05, 1.33)	0.9 (0.83, 0.97)
Poverty Rate	151	1.25 (1.13, 1.37)	1.25** (1.14, 1.37)	0.97 (0.89, 1.05)	1.19 (1.06, 1.34)	1.19 (1.06, 1.34)	0.99 (0.91, 1.08)
Income Inequality	151	1.25 (1.13, 1.37)	1.24* (1.13, 1.36)	0.92 (0.83, 1.01)	1.19 (1.06, 1.34)	1.19 (106, 1.34)	0.95 (0.86, 1.06)
Unemployment Rate	151	1.25 (1.13, 1.37)	1.22** (1.11, 1.34)	1.11 (1.01, 1.22)	1.19 (1.06, 1.34)	1.18 (1.05, 1.33)	1.06 (0.97, 1.17)
Climate							
Mean Maximum Temperature	135	1.25 (1.13, 1.38)	1.23** (1.11, 1.36)	0.94 (0.87, 1.01)	1.2 (1.06, 1.37)	1.2 (1.05, 1.37)	0.98 (0.86, 1.13)
Variation in Max. Temp.	135	1.25 (1.13, 1.38)	1.22** (1.11, 1.35)	1.13 (1.03, 1.25)	1.2 (1.06, 1.37)	1.18 (1.08, 1.34)	1.12 (0.99, 1.26)
Mean Relative Humidity	95	1.18 (1.05, 1.33)	1.17** (1.01, 1.35)	1.02 (0.90, 1.16)	1.16 (0.96, 1.40)	1.15 (0.95, 1.39)	1.06 (0.90, 1.24)
Variation in Relative Humidity	95	1.18 (1.05, 1.33)	1.21** (1.07, 1.37)	0.92 (0.83, 1.03)	1.16 (0.96, 1.40)	1.16 (0.96, 1.40)	0.99 (0.87, 1.12)
Physical Environment							
Water Hardness	109	1.24 (1.10, 1.38)	1.24 (1.11, 1.39)	1.07 (0.98, 1.17)	1.17 (1.01, 1.35)	1.17 (1.03, 1.33)	1.05 (0.94, 1.16)
Altitude	110	1.2 (1.07, 1.35)	1.24** (1.09, 1.40)	1.07 (0.95, 1.20)	1.12 (0.96, 1.30)	1.12* (0.96, 1.31)	1.06 (0.94, 1.21)
Gaseous Co-pollutants							
Carbon Monoxide	107	1.25 (1.14, 1.37)	1.25** (1.14, 1.37)	0.99 (0.90, 1.09)	1.18 (1.04, 1.34)	1.17 (1.01, 1.36)	1.06 (0.97, 1.16)
Nitrogen Dioxide	74	1.27 (1.12, 1.43)	1.27** (1.13, 1.44)	0.93 (0.83, 1.04)	1.2 (1.01, 1.43)	1.21 (1.02, 1.44)	0.94 (0.82, 1.07)
Ozone	117	1.25 (1.14, 1.38)	1.26** (1.15, 1.38)	0.93 (0.84, 1.02)	1.16 (1.03, 1.30)	1.16 (1.03, 1.30)	1.02 (0.92, 1.12)
Sulfur Dioxide	113	1.28 (1.16, 1.42)	1.13 (1.02, 1.25)	1.27 (1.15, 1.40)	1.21 (1.07, 1.37)	1.1 (0.97, 1.24)	1.23 (1.11, 1.36)

* The residuals from the bivariate models are spatially autocorrelated at 0.10 confidence using a Moran's I test statistic.
** The residuals from the bivariate models are spatially autocorrelated at 0.05 confidence using a Moran's I test statistic. (See chapter 6)

Table 5.2b. Impact of multiple metropolitan area-specific covariates on the all-cause mortality rate ratio associated with sulfate exposure, using a two-stage approach.

Ecologic Covariate	Variables Included	Number of MAs	Relative Risk from Sulfates Alone	Adjusted Relative Risk From Exposure to	
				Sulfate	SO_2
Socio-economic and Demographic Covariates					
Independent MAs	Population change, Education, Unemployment,	139	1.24 (1.13, 1.37)	1.17 (1.05, 1.31)	
Regional Adjustment	Poverty, Income, Income inequality	139	1.18 (1.04, 1.34)	1.21 (1.06, 1.38)	
All Gases					
Independent MAs	SO_2, NO_2, CO, O_3	58	1.25 (1.10, 1.43)	1.05 (0.93, 1.18)	1.39 (1.24, 1.55)
Regional Adjustment		58	1.25 (1.03, 1.51)	1.06 (0.90, 1.26)	1.28 (1.12, 1.46)
All Covariates Which Changed the RR from Sulfates by 25% on their Own					
Independent MAs	SO_2 and Population Change	103	1.28 (1.16, 1.42)	1.10 (0.99, 1.22)	1.10 (0.97, 1.24)
Regional Adjustment	SO_2	113	1.21 (1.07, 1.37)	1.23 (1.11, 1.37)	1.23 (1.11, 1.36)

* The residuals are spatially autocorrelated at 0.10 confidence using a Moran's I test statistic.
** The residuals are spatially autocorrelated at 0.05 confidence using a Moran's I test statistic. (See Chapter 6)

Table 5.3a. Impact of metropolitan area-specific covariates on the cardiopulmonary mortality rate ratio associated with sulfate exposure, using a two-stage random effects approach.

Ecologic Covariate	n	Independent MAs			Regional Adjustment		
		Sulfate Alone	Sulfate with Covariate	Covariate with Sulfate	Sulfate Alone	Sulfate	Covariate
Sulfates Alone	151		1.29 (1.15, 1.46)			1.19 (1.06, 1.34)	
Demographic							
Population Change	139	1.28 (1.13, 1.45)	1.17** (1.03, 1.33)	0.84 (0.76, 0.94)	1.19 (1.01, 1.40)	1.16 (0.98, 1.37)	0.91 (0.78, 1.05)
Race							
Whites %	151	1.29 (1.15, 1.46)	1.32** (1.17, 1.50)	1.07 (0.98, 1.18)	1.19 (1.06, 1.34)	1.24 (1.05, 1.46)	1.03 (0.93, 1.13)
African Americans %	151	1.29 (1.15, 1.46)	1.31** (1.15, 1.49)	0.96 (0.86, 1.07)	1.19 (1.06, 1.34)	1.22 (1.04, 1.44)	1.01 (0.89, 1.14)
Health Services							
Doctors per 100000	138	1.28 (1.13, 1.45)	1.25** (1.11, 1.41)	0.85 (0.75, 0.96)	1.19 (1.01, 1.40)	1.17 (1.00, 1.37)	0.85 (0.76, 0.96)
Hospital Beds per 10000	139	1.28 (1.13, 1.45)	1.27** (1.13, 1.44)	1.15 (1.01, 1.30)	1.19 (1.01, 1.40)	1.20 (1.02, 1.41)	1.14 (0.99, 1.30)
Socio-Economic							
High School Educated %	151	1.29 (1.15, 1.46)	1.22** (1.06, 1.39)	1.50 (1.14, 1.98)	1.19 (1.06, 1.34)	1.17 (0.99, 1.38)	0.85 (0.72, 1.00)
1979 Mean Income	151	1.29 (1.15, 1.46)	1.27** (1.13, 1.43)	0.87 (0.78, 0.97)	1.19 (1.06, 1.34)	1.21 (1.04, 1.42)	0.86 (0.77, 0.96)
Poverty Rate	151	1.29 (1.15, 1.46)	1.29** (1.14, 1.46)	1.00 (0.91, 1.09)	1.19 (1.06, 1.34)	1.22 (1.03, 1.43)	1.04 (0.93, 1.17)
Income Inequality	151	1.29 (1.15, 1.46)	1.29** (1.14, 1.46)	0.94 (0.83, 1.07)	1.19 (1.06, 1.34)	1.22 (1.04, 1.44)	1.00 (0.87, 1.16)
Unemployment Rate	151	1.29 (1.15, 1.46)	1.24** (1.10, 1.40)	1.21 (1.07, 1.36)	1.19 (1.06, 1.34)	1.20 (1.02, 1.40)	1.15 (1.01, 1.30)
Climate							
Mean Maximum Temperature	135	1.27 (1.11, 1.45)	1.25** (1.09, 1.42)	0.93 (0.84, 1.03)	1.21 (1.01, 1.44)	1.21 (1.01, 1.44)	0.97 (0.81, 1.17)
Variation in Max. Temp.	135	1.27 (1.11, 1.45)	1.23** (1.08, 1.40)	1.19 (1.05, 1.36)	1.21 (1.01, 1.44)	1.17* (0.98, 1.39)	1.19 (1.01, 1.39)
Mean Relative Humidity	95	1.21 (1.03, 1.43)	1.17** (0.96, 1.43)	1.05 (0.88, 1.26)	1.17 (0.90, 1.51)	1.15 (0.88, 1.39)	1.11 (0.89, 1.38)
Variation in Relative Humidity	95	1.21 (1.03, 1.43)	1.25** (1.06, 1.48)	0.89 (0.76, 1.03)	1.17 (0.90, 1.51)	1.17 (0.91, 1.52)	0.97 (0.81, 1.16)
Physical Environment							
Water Hardness	109	1.26 (1.10, 1.44)	1.26* (1.10, 1.44)	1.11 (1.00, 1.24)	1.13 (0.95, 1.36)	1.14 (0.96, 1.37)	1.08 (0.95, 1.22)
Altitude	110	1.26 (1.08, 1.47)	1.28** (1.09, 1.52)	1.04 (0.89, 1.22)	1.23 (0.99, 1.52)	1.23 (0.99, 1.53)	1.01 (0.84, 1.21)
Gaseous Co-pollutants							
Carbon Monoxide	107	1.33 (1.19, 1.48)	1.32** (1.18, 1.48)	0.94 (0.84, 1.05)	1.23 (1.06, 1.44)	1.24 (1.06, 1.45)	0.98 (0.88, 1.10)
Nitrogen Dioxide	74	1.34 (1.17, 1.54)	1.35** (1.18, 1.55)	0.91 (0.81, 1.02)	1.31 (1.06, 1.61)	1.32 (1.07, 1.62)	0.92 (0.79, 1.07)
Ozone	117	1.32 (1.18, 1.48)	1.32** (1.18, 1.48)	0.96 (0.85, 1.09)	1.18 (1.02, 1.37)	1.18 (1.01, 1.37)	1.08 (0.96, 1.23)
Sulfur Dioxide	113	1.35 (1.19, 1.53)	1.18 (1.04, 1.34)	1.30 (1.16, 1.47)	1.25 (1.07, 1.47)	1.12 (0.96, 1.32)	1.27 (1.12, 1.44)

* The residuals from the bivariate models are spatially autocorrelated at 0.10 confidence using a Moran's I test statistic.
** The residuals from the bivariate models are spatially autocorrelated at 0.05 confidence using a Moran's I test statistic. (See chapter 6)

Table 5.3b. Impact of multiple metropolitan area-specific covariates on the cardiopulmonary mortality rate ratio associated with sulfate exposure, using a two-stage approach.

Ecologic Covariate	Variables Included	Number of MAs	Relative Risk from Sulfates Alone	Relative Risk From Exposure to	
				Sulfate	SO_2
Socio-economic and Demographic Covariates					
Independent MAs	Population change, Education, Unemployment,	139	1.28 (1.13, 1.45)	1.18** (1.02, 1.37)	
Regional Adjustment	Poverty, Income, Income inequality	139	1.19 (1.01, 1.40)	1.21 (1.01, 1.44)	
All Gases					
Independent MAs	SO_2, NO_2, CO, O_3	58	1.31 (1.13, 1.52)	1.11 (0.97, 1.27)	1.42 (1.25, 1.61)
Regional Adjustment		58	1.36 (1.08, 1.72)	1.15 (0.93, 1.42)	1.30 (1.11, 1.52)
All Covariates Which Changed the RR from Sulfates by 25% on their own					
Independent MAs	SO_2, Population Change and Education	103	1.34 (1.18, 1.51)	1.07 (0.93, 1.24)	1.29 (1.14, 1.46)
Regional Adjustment	SO_2	113	1.25 (1.07, 1.47)	1.12 (0.96, 1.32)	1.18 (1.08, 1.28)

* The residuals are spatially autocorrelated at 0.10 confidence using a Moran's I test statistic.

** The residuals are spatially autocorrelated at 0.05 confidence using a Moran's I test statistic. (See chapter 6)

Table 5.4a. Impact of metropolitan area-specific covariates on the lung cancer mortality rate ratio associated with sulfate exposure, using a two-stage random effects approach.

		Independent MAs†		
Ecologic Covariate	n	Sulfate Alone	Sulfate with Covariate	Covariate with Sulfate
Sulfates Alone	151		1.39 (1.09, 1.75)	
Demographic				
Population Change	139	1.36 (1.07, 1.74)	1.36 (1.04, 1.77)	0.99 (0.80, 1.23)
Race				
Whites %	151	1.39 (1.09, 1.75)	1.39 (1.09, 1.78)	1.02 (0.85, 1.22)
African Americans %	151	1.39 (1.09, 1.75)	1.40 (1.10, 1.80)	0.96 (0.78, 1.18)
Health Services				
Doctors per 100000	138	1.36 (1.07, 1.73)	1.35 (1.06, 1.71)	0.92 (0.72, 1.19)
Hospital Beds per 10000	139	1.36 (1.07, 1.74)	1.37 (1.08, 1.75)	0.87 (0.66, 1.15)
Socio-Economic				
High School Educated %	151	1.39 (1.09, 1.75)	1.38 (1.06, 1.80)	1.00 (0.75, 1.31)
1979 Mean Income	151	1.39 (1.09, 1.75)	1.39 (1.09, 1.76)	1.01 (0.82, 1.24)
Poverty Rate	151	1.39 (1.09, 1.75)	1.39 (1.09, 1.76)	0.99 (0.81, 1.20)
Income Inequality	151	1.39 (1.09, 1.75)	1.38 (1.09, 1.75)	0.90 (0.71, 1.15)
Unemployment Rate	151	1.39 (1.09, 1.75)	1.34 (1.06, 1.71)	1.15 (0.90, 1.46)
Climate				
Mean Maximum Temperature	135	1.44 (1.40, 1.47)	1.43 (1.12, 1.84)	0.99 (0.82, 1.21)
Variation in Max. Temp.	135	1.44 (1.40, 1.47)	1.46 (1.14, 1.88)	0.92 (0.72, 1.18)
Mean Relative Humidity	95	1.51 (1.14, 2.00)	1.34 (0.96, 1.87)	1.21 (0.90, 1.63)
Variation in Relative Humidity	95	1.51 (1.14, 2.00)	1.47 (1.10, 1.97)	1.08 (0.84, 1.40)
Physical Environment				
Water Hardness	109	1.42 (1.08, 1.86)	1.41 (1.08, 1.86)	0.94 (0.75, 1.18)
Altitude	110	1.24 (0.95, 1.62)	1.13 (0.86, 1.50)	0.75 (0.56, 1.01)
Gaseous Co-pollutants				
Carbon Monoxide	107	1.29 (1.03, 1.62)	1.29 (1.03, 1.61)	0.83 (0.66, 1.03)
Nitrogen Dioxide	74	1.34 (1.03, 1.74)	1.36 (1.04, 1.76)	0.88 (0.71, 1.10)
Ozone	117	1.32 (1.05, 1.64)	1.33 (1.07, 1.65)	0.72 (0.57, 0.91)
Sulfur Dioxide	113	1.35 (1.08, 1.68)	1.39 (1.08, 1.81)	0.94 (0.73, 1.20)

† Neither the raw lung cancer mortality rates nor the residuals of the independent cities models incorporating spatially autocorrelated sulfate and covariate values were found to be spatially autocorrelated. Therefore, it was not necessary to run regionally adjusted or spatially filtered models. (See chapter 6)

Table 5.4b. Impact of multiple metropolitan area-specific covariates on the lung cancer mortality rate ratio associated with sulfate exposure, using a two-stage approach.

Ecologic Covariate	Variables Included	Number of MAs	Relative Risk from Sulfates Alone	Relative Risk From Exposure to	
				Sulfate	SO_2
Socio-economic and Demographic Covariates					
Independent MAs	Population change, Education, Unemployment, Poverty, Income, Income inequality	139	1.36 (1.07, 1.74)	1.23 (0.90, 1.68)	
All Gases					
Independent MAs	SO_2, NO_2, CO, O_3	58	1.48 (1.12, 1.96)	1.63 (1.19, 2.23)	0.90 (0.67, 1.21)
All Covariates Which Changed the RR from Sulfates by 25% on their Own					
Independent MAs	Mean Relative Humidity and Altitude	68	1.62 (1.21, 2.16)	1.39 (0.97, 2.01)	

Table 5.5. Impact of county-specific covariates on the all-cause mortality rate ratio associated with sulfate exposure, using a two-stage approach.

Ecologic Covariate	n	Independent Observations		Independent Counties	
		Sulfates	Ecologic	Sulfates	Ecologic
Sulfates Alone	132	1.47 (1.32, 1.65)		1.50** (1.30, 1.73)	
Demographic Data					
Population Change	131	1.35 (1.19, 1.54)	0.89 (0.81, 0.98)	1.38** (1.20, 1.59)	0.89 (0.80, 0.99)
Race Data					
Whites %	132	1.54 (1.37, 1.72)	1.11 (1.02, 1.22)	1.55** (1.37, 1.76)	1.11 (1.01, 1.21)
African Americans %	132	1.55 (1.38, 1.75)	0.90 (0.82, 0.98)	1.57** (1.38, 1.78)	0.90 (0.82, 0.99)
Health Services					
Doctors per 100,000	132	1.46 (1.31, 1.64)	0.93 (0.81, 1.06)	1.49** (1.32, 1.68)	0.92 (0.80, 1.06)
Hospital Beds per 100,000	132	1.47 (1.31, 1.64)	1.03 (0.92, 1.15)	1.49** (1.32, 1.69)	1.03 (0.92, 1.16)
Socio-Economic Data					
High School Educated	132	1.44 (1.26, 1.64)	0.97 (0.88, 1.07)	1.46** (1.26, 1.68)	0.96 (0.87, 1.07)
Median Income	132	1.48 (1.32, 1.66)	1.03 (0.94, 1.13)	1.50** (1.33, 1.70)	1.02 (0.92, 1.14)
Poverty Rate	132	1.50 (1.34, 1.68)	0.93 (0.84, 1.04)	1.52** (1.34, 1.73)	0.94 (0.84, 1.05)
Gini Coefficient	132	1.50 (1.34, 1.67)	0.87 (0.78, 0.97)	1.52** (1.34, 1.71)	0.88 (0.78, 0.99)
Unemployment Rate	132	1.41 (1.25, 1.59)	1.08 (0.98, 1.19)	1.42** (1.25, 1.63)	1.11 (1.00, 1.23)
Co-pollutants					
Sulfur Dioxide	132	1.36 (1.16, 1.61)	1.11 (0.97, 1.27)	1.38** (1.16, 1.64)	1.12 (0.97, 1.28)
Multiple Covariate Analyses					
Socio-economic data and population change[1]		1.30 (1.11, 1.52)		1.32** (1.12, 1.56)	
25%[2]		1.35 (1.19, 1.54)		N/A	

** The residuals are spatially autocorrelated at 0.05 confidence using a Moran's I test statistic.

[1] Population change, % with high school education, median income, individual poverty rate, Gini coefficient, unemployment rate.

[2] Population change was the only variable in the independent observations model to effect the RR from sulfates by 25% or more. No variables effected the sulfate RR by 25% or more in the independent counties model.

Table 5.6. Impact of county-specific covariates on the cardiopulmonary mortality rate ratio associated with sulfate exposure, using a two-stage approach.

Ecologic Covariate		Independent Observations		Independent Counties	
	n	Sulfates	Ecologic	Sulfates	Ecologic
Sulfates Alone	132	1.72 (1.47, 2.01)		1.75** (1.48, 2.08)	
Demographic Data					
Population Change	131	1.52 (1.27, 1.82)	0.85 (0.75, 0.97)	1.55** (1.28, 1.88)	0.84 (0.73, 0.97)
Race Data					
Whites %	132	1.86 (1.59, 2.17)	1.22 (1.08, 1.38)	1.87** (1.58, 2.22)	1.22 (1.07, 1.39)
African Americans %	132	1.89 (1.61, 2.21)	0.82 (0.73, 0.93)	1.90** (1.60, 2.27)	0.82 (0.72, 0.93)
Health Services					
Doctors per 100,000	132	1.68 (1.44, 1.96)	0.78 (0.64, 0.94)	1.71** (1.45, 2.02)	0.76 (0.62, 0.93)
Hospital Beds per 100,000	132	1.73 (1.48, 2.03)	0.93 (0.79, 1.09)	1.76** (1.48, 2.09)	0.92 (0.79, 1.09)
Socio-Economic Data					
High School Educated	132	1.66 (1.37, 2.00)	0.95 (0.83, 1.10)	1.67** (1.36, 2.04)	0.93 (0.80, 1.08)
Median Income	132	1.74 (1.49, 2.04)	1.08 (0.95, 1.24)	1.77** (1.49, 2.11)	1.07 (0.92, 1.24)
Poverty Rate	132	1.78 (1.52, 2.08)	0.88 (0.75, 1.02)	1.81** (1.52, 2.15)	0.88 (0.75, 1.03)
Gini Coefficient	132	1.77 (1.52, 2.05)	0.78 (0.66, 0.91)	1.79** (1.51, 2.12)	0.80 (0.68, 0.95)
Unemployment Rate	132	1.52 (1.29, 1.80)	1.25 (1.09, 1.43)	1.54** (1.29, 1.85)	1.29 (1.12, 1.48)
Co-pollutants					
Sulfur Dioxide	132	1.53 (1.22, 1.92)	1.13 (0.93, 1.36)	1.54** (1.22, 1.95)	1.13 (0.93, 1.37)
Multiple Covariate Analyses					
Socio-economic data and population change[1]		1.40 (1.14, 1.72)		1.42 (1.14, 1.76)	
25%[2]		1.41 (1.17, 1.70)		1.44** (1.18, 1.75)	

** The residuals are spatially autocorrelated at 0.05 confidence using a Mooran's I test statistic.

[1] The variables included in this analysis are: population change, % with high school education, median income, individual poverty rate, Gini coefficient, unemployment rate.

[2] In both the independent observations and the independent counties models population change and unemployment rate effected the RR from sulfates by more than 25%.

Figure 5.1
Scales of Analysis

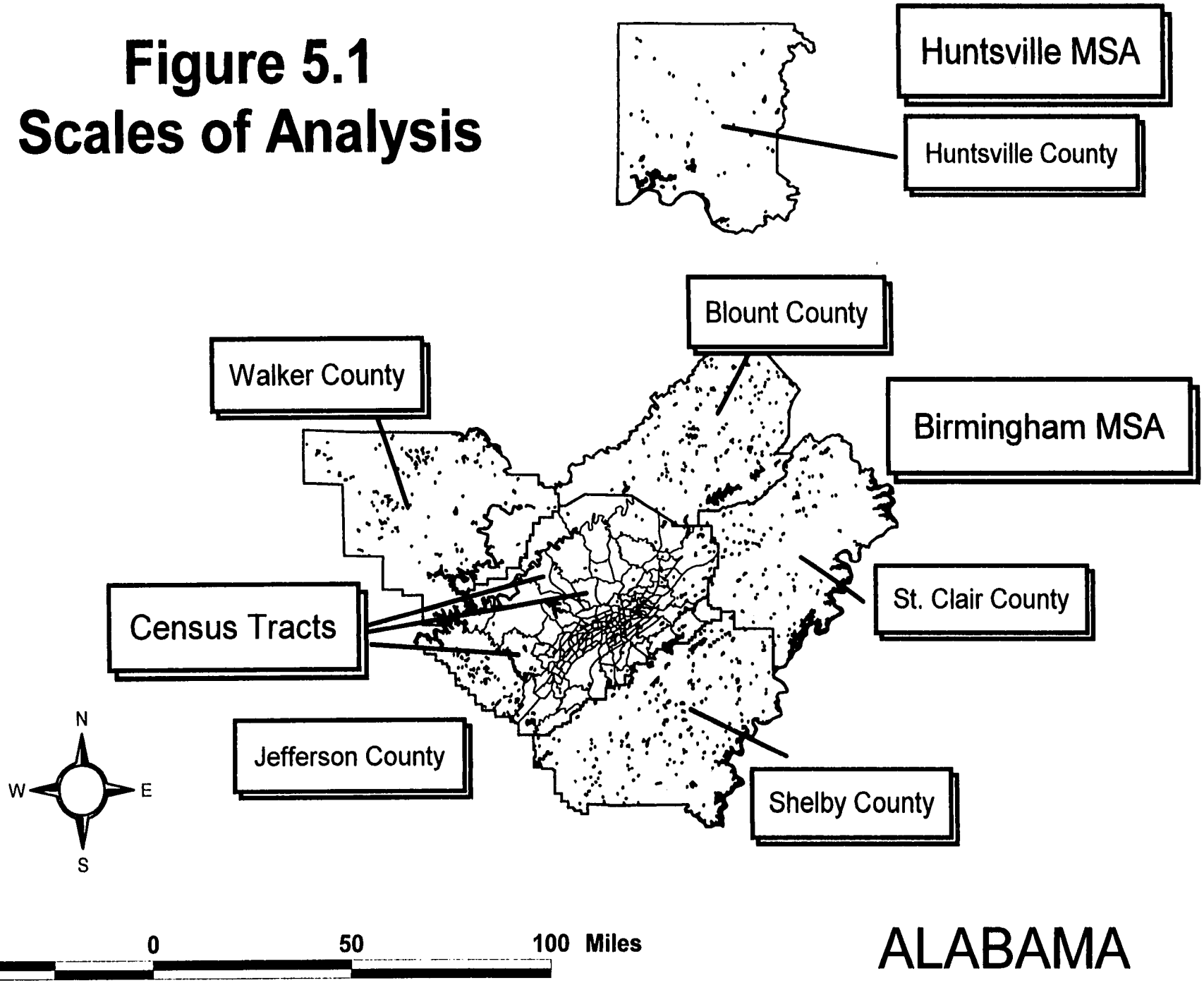


Figure 5.2 Assignment of MAs to EPA Airshed Regions

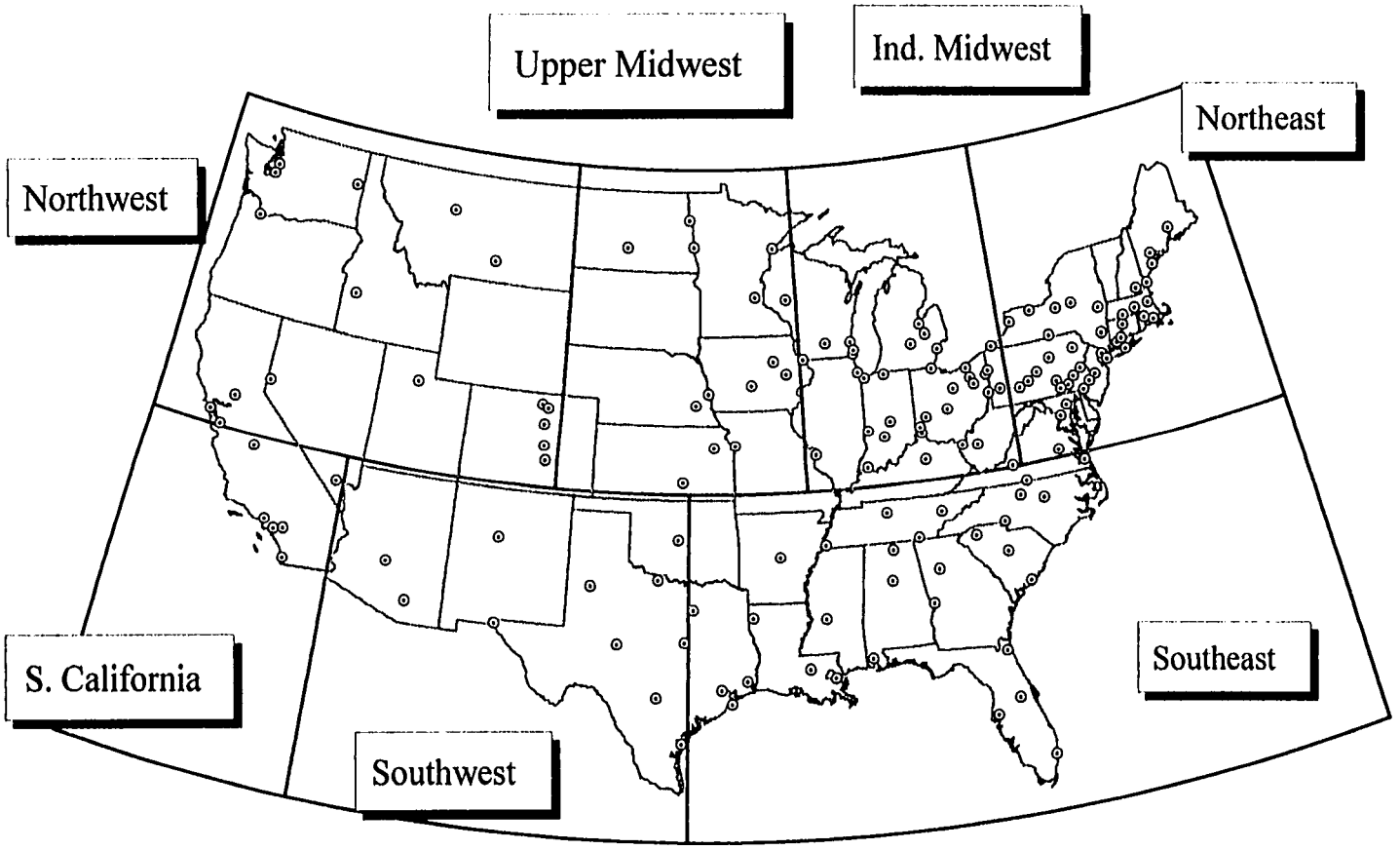
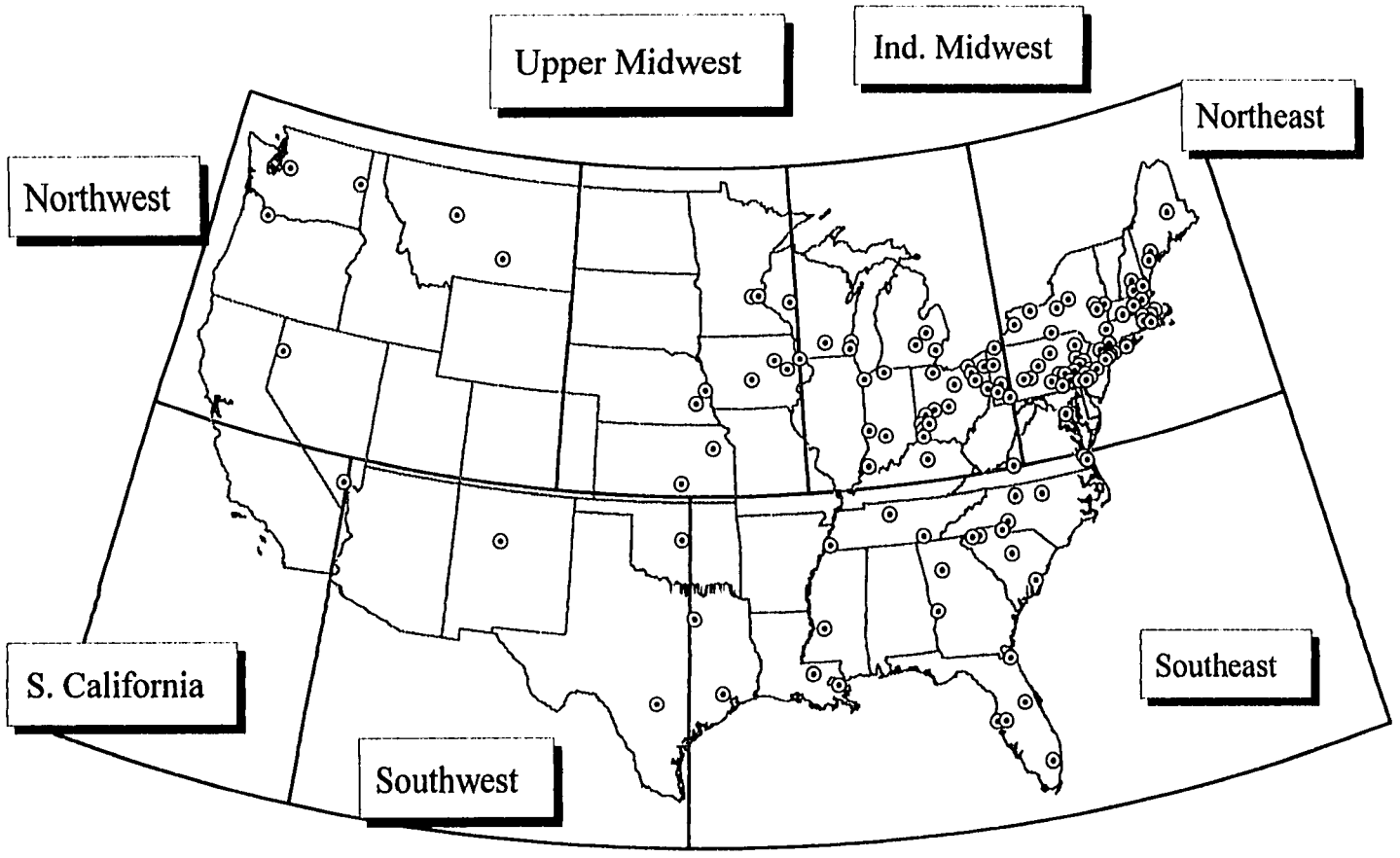


Figure 5.3 Assignment of Counties to EPA Airshed Regions



6 Spatial Analyses

6.1 Spatial Patterns

While ecologic data may be collected for groups who do not have a specific location in space, most routinely collected ecologic-level data are collected for specific areas: census tracts, counties, metropolitan areas, states and countries. To a certain extent, therefore, studies utilizing ecologic-level covariates simply measure spatial coincidence, or the degree to which two or more covariates vary together across space. It is for this reason that an ecologic covariate only needs to be associated with the outcome of interest and not the exposure to be a potential confounder. Multi-level analyses do not completely overcome this problem since they still rely on spatial coincidence to establish associations. However, they have the advantage of controlling for confounding on different scales through measuring the degree to which covariates vary together at more than one scale.

Most human processes – economic, industrial, social, cultural – are dispersed in space in a particular pattern. Because most of these processes are interdependent, they tend to have similar broad spatial patterns. Physical environmental processes also exhibit clear spatial patterns. For example, locations within an airshed region will have more similar air pollutant composition than locations in different airsheds. Whenever a variable is dispersed over space in a somewhat regular pattern, it is said to be spatially

autocorrelated (Odland, 1988). In the words of Tobler's First Law of Geography, "everything is related to everything else, but near things are more related than distant things" (cited in Sawada, 1999). In terms of data analysis, when values of a variable are dispersed in space in a non-random pattern they are considered to be spatially autocorrelated (see figure 6.1).

Spatial Autocorrelation

It is, in fact, spatial autocorrelation which epidemiologists exploit in ecologic-level studies. An association is observed between two variables if they vary in space in a similar way. In other words, if the spatial patterns of each of the variables are similar, then the variables are thought to be related. Understanding spatial autocorrelation in ecologic variables can aid investigators in developing hypotheses. If a variable has no spatial autocorrelation, with no discernable pattern in its spatial distribution, it is unlikely to be related to a variable with spatial autocorrelation. Spatial autocorrelation tests can be conducted at different scales. One variable is more likely to be related to another if they both exhibit spatial autocorrelation at the same scale. However, knowledge of the spatial autocorrelation of one's data does not solve all problems. Two unrelated variables could also have similar patterns as a result of a third unmeasured variable. In this case, the observed association would be a false one, and the results would be confounded.

While testing individual variables for spatial autocorrelation may help a researcher to generate hypotheses regarding which variables are related to each other, it is more common to see residuals of models tested. Finding spatial autocorrelation in the residuals of an ordinary least squares regression indicates one of three things: that an

important variable is missing from the model, that the wrong structure has been used, or that there is systematic measurement error in one or more variables (Miron, 1984). If the missing variable is a confounding one, its exclusion may lead to biased estimates of the relative risk from the exposure variable (Miron, 1984). By definition, if spatial autocorrelation is present, there is spatial dependency in the random error term (Miron 1984). If dependence is present in the random error term some tests of significance may become unreliable (Miron 1984).

Statistical tests for spatial autocorrelation come from the discipline of geography, which is centrally concerned with the organization of objects and processes in space (Odland, 1988). There are two general categories of spatial autocorrelation tests: global and local. Local tests are not useful to the particular set of research questions addressed in this thesis as they merely provide a means of determining "hotspots," or clusters of high or low values. In ecologic-level studies, which focus on broader spatial patterns, global spatial autocorrelation tests are appropriate.

A formal statistical test of global spatial autocorrelation is able to answer the question: is this set of values randomly distributed in space (Odland, 1988)? Spatial autocorrelation statistics are descriptive statistics which describe the distribution of a variable in space (Odland 1988). Spatial autocorrelation tests are concerned with the random error term. They conceptualise the error term as being made up of two parts: one which varies randomly and is independent and the other which varies across space in a systematic fashion and is, therefore, dependent on the error terms of spatially adjacent observations (Miron 1984). Epidemiologists may be more familiar with autocorrelation

statistics in time series analyses (cf Burnett and Krewski, 1994); however, while the concepts are similar, some important difference arise due to the two-dimensional coordinate system used in spatial analyses (Odland 1988).

Spatial autocorrelation statistics are functions of both the values of the data and the distribution in space of the sample points. In the simplest situation, data will have been collected in a regular lattice. For example, a soil scientist might take samples of topsoil at regular intervals in space, forming a grid of known values. However, in the case of the ACS Study, the data were collected for a set of cities which are situated in space in an irregular lattice (see Figure 5.2). In other words, the distances and angles between the points differ. In order to describe this lattice numerically, geographers use what is called a "spatial weighting function" (Odland, 1988) which assigns weights to pairs of values based on their relative location. In its simplest form, the analyst specifies a distance (or scale) at which she would like to measure spatial autocorrelation. The weighting function would then assign a weight of 0 to all pairs that are further apart in space than that distance and a value of 1 to all pairs which fell within that distance from each other. In this way, only those pairs of points within the chosen distance are included in the calculation of the statistic.

While there are a number of different spatial autocorrelation statistics, this analysis made use of Moran's I statistic, which is based on the covariance amongst values distributed in space (Getis, 1995). Like all spatial autocorrelation statistics, Moran's I is based on a matrix cross-product (Sawada, 1999):

$$I = \frac{n}{\sum_{i=1}^{i=n}\sum_{j=1}^{j=n} W_{ij}} * \frac{\sum_{i=1}^{i=n}\sum_{j=1}^{j=n} W_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{i=n}\left(x_i - \bar{x}\right)^2}$$

Here, i and j are two points and x is the variable of interest.

$\sum_{i=1}^{i=n}\sum_{j=1}^{j=n} W_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)$ is the sum of the cross-product of the spatial weighting function, W_{ij}, which is made up of 0s and 1s, and the matrix of the pair-wise products of the difference of the values of variable x at two points (i and j) from the mean value of all the points. The sum of the cross-products is then divided by the sum of the squared difference of the value of x at each point from the mean value. Then this entire term is weighted by the ratio of the number of points to the number of pairs which meet the distance criteria.

Moran's I ranges from 1, which indicates very strong positive spatial autocorrelation, to -1, which indicates very strong negative spatial autocorrelation. Positive spatial autocorrelation occurs when like values are grouped (see Figure 6.1 for a simplified diagram of different spatial autocorrelation cases). Negative spatial autocorrelation involves a pattern in which unlike values are grouped. A Moran's I approaching zero indicates a random pattern. Sawada in the University of Ottawa's department of geography has developed an Excel macro which calculates both the Moran's I and Z-values for data sets with spatial dimensions to them (Sawada, 1999). Analysts can specify any distance at which to calculate the Moran's I. This distance is

called the lag distance and is used to determine which pairs of points are to be included in the calculation of the Moran's I. In other words, the lag distance is the distance used in deriving the spatial weights for pairs of points (see above). It is the scale at which the spatial autocorrelation test is being conducted.

One weakness in Moran's I is that it is not able to discriminate between patterns in which high values dominate from patterns in which low values dominate. Further, it is poor at discriminating between a random pattern and one in which there is little variation in the values from the mean (Getis and Ord, 1992).

All ecologic covariates, including pollution values and relative risks, were tested for spatial autocorrelation. Tests for spatial autocorrelation were also conducted on the residuals of all the multivariate analyses conducted for this thesis.

6.2 Filtered Analyses

While spatial autocorrelation can lead to biased estimates and unreliable tests of significance, it is not a problem in and of itself. Rather, it is an indication that there is a problem in the analysis: an important ecologic variable has been omitted, the model structure is incorrect, or there is systematic measurement error present in one of the variables (Miron 1984). Ideally, therefore, spatial autocorrelation should be dealt with by addressing these three problems.

However, if the spatial autocorrelation is due to measurement error, it may not be possible to correct for it at the time the analysis has begun. It may also be undesirable to change the structure of the model because that may complicate things in such a way

that it is difficult to interpret the results meaningfully (Getis, 1995). In some cases spatial autocorrelation cannot be removed from a model through introducing an omitted variable simply because no data are available for that variable. There are many spatial processes at work at different scales which can influence mortality. The question in this thesis, from a population health perspective, is what is the contribution that air pollution makes to mortality above and beyond the contribution of all other variables.

Controlling spatial autocorrelation can help investigators examine how strong the association between air pollution and mortality is when all extraneous spatial processes are controlled for. In effect, removing spatial autocorrelation is an attempt to remove the broader spatial patterns that can drive a regression model to the point that what is merely a coincidence of shared space appears to be an epidemiologic relationship. Once these broad common trends are removed, the remaining finer-grained variability can better reveal whether two variables are indeed related. Of course, removing the broad trends means that relationships that operate at this scale will not be detectable. It should be noted, that while correcting for spatial autocorrelation removes spatial dependencies, filtering out the spatial autocorrelation cannot identify biases (Miron 1984).

Getis (1995) has developed a method of filtering out spatial autocorrelation when it cannot be corrected for. Getis proposes removing all spatial dependence from both the independent and dependent variables before regressing them. Once this is done, ordinary least squares regression can be conducted, since the error terms should no longer be spatially autocorrelated.

Getis' method of filtering is based on the $G_i(d)$ statistic, which measures the

concentration of the sum of values (not the variation as in the Moran's I) within the area being studied (Getis 1995). In other words, it gives the proportion of the entire summed values within a specified distance from each point. The Getis statistic for point *i* at a particular distance (*d*) for variable *x* is given by:

$$G_i(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{j=1}^{n} x_j}, j \neq i$$

Where, similar to the Moran's I, w_{ij} is a spatial weight function where ones are assigned to all points *j* within distance *d* of point *i* and zeros are assigned to all the others outside this distance (points are not compared to themselves so *j* cannot equal *i* in this equation). x_j is the value of the variable at point *j*. This statistic is then compared to what would be expected if the *x* values were distributed randomly in the same spatial configuration. The expected value of the Getis statistic at distance *d* for point *i* is given as follows:

$$E(G_i) = \frac{\sum_j w_{ij}(d)}{(n-1)}$$

Getis suggests using $G_i(d)$ to extract from each of the variables that which "embodies" spatial dependence. Getis suggests including in subsequent regression models both the pure spatial autocorrelation component of the variable and the

component from which the spatial autocorrelation has been removed. However, the spatially autocorrelated component of the original variable was discarded for the purposes of this thesis as the unmeasured spatial processes can be adequately controlled for through including only the filtered part of the variables.

Filtering is achieved by determining the degree of dissimilarity between the expected proportion of the sum of the value at distance *d* from point *i* to the actual proportion (excluding the value at point *i* itself). The filtered variable values x_i^* are thus given by,

$$x_i^* = \frac{x_i\left(E(G_i)\right)}{G_i(d)}$$

The challenge in filtering out spatial autocorrelation is to determine the best distance *d*. A distance should be chosen so that when regressions are run with the filtered variables, no spatial autocorrelation can be detected in the residuals (Getis 1995). The choice of a distance for filtering is more of an art than a science. Generally, the distance at which the spatial autocorrelation begins to level off in a semivariogram[2] (the sill) is used as the starting point for determining *d*. In Figure 6.3, which shows the semivariogram for all cause mortality in MAs, the semivariance increases sharply and

[2] A semivariogram is a graph which compares the semivariance of the values of points at different distances (or lags) from each other. Generally semivariance increases as points at a greater distance from each other are assessed until a certain point (the sill) when the semivariance levels off. This implies that points that are closer together are more similar than those which are farther apart and is an indication that spatial autocorrelation is present in the data. GS+ software was used to generate the semivariogram shown in Figure 6.3.

steadily from 0 km to approximately 600 km. After this point the semivariance does not show such a clear cut trend and, in fact, appears to level off. The point at which semivariance levels off is called the sill. The observed sill was used as the starting point for testing for spatial autocorrelation. In the case of all cause mortality, filtering at a *d* of 600 km removed the spatial autocorrelation, however, in cases where spatial autocorrelation persisted, a new distance of 50 km more or less was tried and so on until an appropriate *d* was determined.

Filtered variables were then entered into random effects models outlined in the previous chapter.

6.3 Results of Filtering

Spatial Autocorrelation in Individual Covariates

Before discussing the degree of spatial autocorrelation found in the individual covariates, some comment on the distribution of points in this study is warranted. The 151 metropolitan areas were spread over the entire United States, with the majority being located in the Eastern half of the country. Although the metropolitan areas are in fact polygons, they were converted into point data for the purposes of the spatial analysis. This was done using an ArcView macro which calculated the geographical centroids of the MA polygons. The nearest two points in this irregular lattice were slightly more than 9 km apart. On average, points were 111 km distant from their nearest neighbour, and were as far as 578 km from the next closest study city. As can be seen, the MAs are more densely clustered in the New England area, and are quite far apart in the West and Mid-

west. This is far from an ideal pattern, since spatial autocorrelation at distance *d* will be calculated over a greater number of data points in the East than in the West. Sulfates and mortality are also higher in East than in the West as well which may mean that tests for spatial autocorrelation of lower values of sulfates and lower mortality rates are less reliable than tests at the higher levels.

All of the relative risks produced for the random effects analysis and each of the ecologic covariates, including sulfates, were tested for spatial autocorrelation at a range of distances from 50km to 5000km at 50km intervals. The resulting Moran's I's show at what scale the proximity ceases to be a predictor of the variable in question. Lung cancer did not demonstrate any spatial autocorrelation and is thus not likely to be affected by any processes which operate at scales larger than a metropolitan area. Therefore, no further spatial analyses were done with lung cancer. Four covariates (number of physicians per 100,000 population, mean income, poverty rate and ozone) also showed no sign of spatial autocorrelation at distances of 50km or greater.

Filtered Variables Compared to Raw Variables

For the metropolitan-area analyses, descriptive statistics are provided for both raw and filtered data. Four variables were found to exhibit no spatial autocorrelation in their raw form (mean income, number of doctors per 100,000, poverty index and ozone) and were therefore not filtered. Of the spatially autocorrelated variables, it was not possible to eliminate spatial autocorrelation in mean maximum temperature, carbon monoxide and nitrogen dioxide through filtering, precluding their use in the filtered analyses. In most cases, filtering the data resulted in only minor changes to the range of

the ecologic covariate and to its correlation with sulfate air pollution. However, for two ecologic covariates, population change and mean relative humidity, filtering has an appreciable impact on their range and correlation with sulfates (see Table 6.1).

For the most part, the raw variables are highly correlated with their filtered counterpart, with coefficients of 80% or more. However, raw population change and raw mean relative humidity are weakly correlated with their filtered counterparts (7% and 6% respectively). This is because the two variables in raw form demonstrated a high degree of spatial autocorrelation which was then removed by filtering. The high level of spatial autocorrelation in population change is likely due to the regional patterns of economic growth in the US. The West is booming with high tech industry and the older industries of the East, particularly in the Ohio Valley are dwindling so people are moving from East to West.

Filtering also changed the correlation between these two variables and other variables. For example, the correlation between population change and raw sulfates is -40% before filtering and 2% after filtering. Similarly the correlation between mean relative humidity and sulfates is 59% before filtering and -12% after filtering. In general, the correlations between filtered variables are lower than those between unfiltered variables, since the spatial filtering process removes the correlation between ecologic-level variables attributable to broad spatial patterns.

6.4 Results of Residual Tests

Previously Reported MA Models

The random effects models outlined in the previous chapter used variables which are spatially autocorrelated. This is not a problem unless the residuals of these models are also spatially autocorrelated. It can be seen in Tables 5.1 and 5.2 that the vast majority of the independent MAs models do have spatially autocorrelated residuals at a distance of 600km. The notable exceptions are the all-cause and cardiopulmonary models including SO_2 and the all-cause model including water hardness. The lung cancer models were not tested for spatial autocorrelation because the independent variable was not spatially autocorrelated. Unlike the two covariate models, very few of the multiple ecologic covariate models had spatially autocorrelated residuals.

With a few exceptions, all of the spatial autocorrelation was eliminated in the sulfate models when broad regions were controlled for through the inclusion of indicator variables. A comparison of Figures 5.2 and 6.4 reveals that the EPA regions are of a similar size to the search circle of 600km radius used in calculating Moran's I. The residuals of the altitude and all-cause mortality and variation in maximum temperature and cardiopulmonary mortality models continued to show spatial autocorrelation at 0.10 level of confidence. This may be due to the small number of MAs included in these models, so that the number of points in each of the regions became too small to be stable.

Previously Reported County Models

Significant spatial autocorrelation was also detected in the Independent Counties models for both all-cause mortality and cardiopulmonary mortality. Spatial

autocorrelation that was significant at 0.05 level of significance was found in the residuals of the model containing only sulfates, as well as in the residuals of all the models with sulfates and an ecologic covariate. Unlike the Metropolitan-scale models, the inclusion of sulfur dioxide did not remove the spatial autocorrelation.

Because the counties with air pollution monitors and ACS participants are concentrated in the Eastern US, I was unable to derive a suitable matrix to enable the spatial filtering of the county-scale data. Therefore, at the county-scale, it is not possible to calculate estimates of the risk posed by sulfate air pollution on all-cause or cardiopulmonary mortality with spatial filtering techniques.

6.5 Results of Metropolitan Analysis With Filtered Variables

While the regional adjustment model adjusts for omitted variable(s) that operate within the boundaries of the EPA airshed regions, the filtered variable analysis allows for the control of processes which operate on the basis of proximity. Filtering further reduced the RR from sulfates (see Tables 6.3 and 6.4) to 1.09 (1.01, 1.19) for all-cause mortality and 1.13 (1.01, 1.27) for cardiopulmonary mortality (as compared to 1.25 and 1.29, respectively, in the independent cities model). However, it also rendered the RR for sulfates more robust to the inclusion of other covariates. Nevertheless, SO_2 still had the largest effect, reducing the sulfate RR to 1.05 (0.97 - 1.14) for all-cause and 1.10 (0.99 - 1.22) for cardiopulmonary mortality.

Table 6.1. Metropolitan-scale Place-specific Covariates, Descriptive Statistics Comparing Raw and Filtered Variables

	Raw				Filtered			
	N	min	max	mean	cor†	min	max	mean	cor
Sulfates	151	3.6	23.5	10.6	100	3.6	19.0	9.7	100
Demographic Data									
Population Change	139	-9.7	28.4	5.7	-40	-2126.1	90.9	-9.8	2
Race Data									
Whites %	151	57.3	99.2	87.4	-22	63.3	106.7	88.0	-19
African Americans %	151	0.1	46.6	10.1	28	0.1	28.4	8.7	27
Health Services Data									
Doctors per 100000	138	97	512	206	-13	Not necessary to filter[1]			-12
Hospital Beds per 10000	139	222	1424	609	-5	222	1233	592	-3
Socio-Economic Data									
High School Educated %	151	41.8	82.7	68.4	-44	42.6	83.6	69.7	-30
1979 Income mean	151	5700	10360	7363	-13	Not necessary to filter[1]			-6
Poverty Rate	151	5.69	20.02	11.12	8	Not necessary to filter[1]			11
Income Inequality	151	0.34	0.45	0.39	2	0.34	0.46	0.39	11
Unemployment Rate	151	2.7	14.5	6.9	16	2.7	15.2	6.8	13

Table 6.1. Metropolitan-scale Place-specific Covariates, Descriptive Statistics Comparing Raw and Filtered Variables

	Raw					Filtered			
	N	min	max	mean	cor†	min	max	mean	cor
Climate Data									
Maximum Temperature mean	135	49.3	87.0	65.4	-12	Not possible to filter[2]			
Variation in Max. Temp.	135	14.0	144.7	80.1	-6	14.0	144.7	79.3	-6
Relative Humidity mean	95	17.0	56.4	46.3	59	31.4	98.5	46.9	-12
Variation in Relative Humidity	95	76.3	312.6	213.0	23	31.4	98.5	46.9	29
Physical Environment Data									
Water Hardness	109	9.0	422.0	106.2	-9	14.0	422.0	101.2	-4
Elevation	110	2	1831	279.9	-49	3	1831	351.4	-43
Gaseous Co-pollutants									
Carbon Monoxide	107	194.2	3952.3	1561.3	-7	Not possible to filter[2]			
Nitrogen Dioxide	74	7.8	51.1	25.5	-1	Not possible to filter[2]			
Ozone	117	10.4	41.1	23.5	2	Not necessary to filter[1]			9
Sulfur Dioxide	113	0.02	29.32	9.27	48	0.02	22.59	7.93	24

† "Cor" stands for Pearson correlation coefficients, expressed here as percentages.

1. These variables did not exhibit spatial autocorrelation, and were thus not filtered.

2. These variables did exhibit spatial autocorrelation but a suitable distance could not be found at which the spatial autocorrelation could be removed.

Table 6.2. Pearson correlations (%) between metropolitan-scale place-specific covariates, raw and filtered.*

Raw Data / Filtered Data	Pop. Change	White	Black	Doctors	Hospital Beds	High School	Mean Income	Poverty rate	Gini	Unemployment	Temperature	Var. in Temp.	R. Humidity	Var. in R.H.	H2O Hardness	Altitude	Sulfates	CO	NO_2	O_3	SO_2
Pop. Change	7	-17	11	8	-26	25	14	19	28	-28	70	-36	-60	-15	-6	27	-40	10	10	27	-43
Whites	-7	90	-96	-35	-12	24	-12	-45	-51	3	-46	35	-13	-22	27	28	-22	-3	-4	-3	12
Blacks	10	-90	80	24	17	32	2	50	49	4	46	27	17	23	-24	-29	28	-7	-7	4	-6
Doctors†	-12	-41	29	100	32	26	41	-2	20	-36	51	-28	-42	-13	-12	-13	-13	12	21	-14	-7
Hospital Beds	-6	-12	10	32	96	-12	-19	10	5	-9	-7	26	7	26	-9	-14	-5	-12	-20	-8	7
High School	-9	7	-18	31	-13	92	55	-44	-27	-31	-19	17	-23	-35	18	38	-44	15	22	-0	-22
Income‡	-4	-25	22	41	-18	50	100	-58	-20	-22	-2	-10	-7	-26	5	1	-13	16	42	-20	-12
Poverty‡	8	-26	23	-2	9	-34	-58	100	76	25	50	-24	-1	22	-15	-6	8	-7	-22	6	-16
Gini	8	-38	35	20	4	-21	-18	74	99	9	60	36	12	27	-22	-8	2	10	-1	1	-22
Unemploy.	2	26	-15	-37	-6	-40	-28	18	5	89	4	4	15	-4	5	9	16	-27	-23	-13	3
Temperature†												-50	-48	-9	5	0	-12	-3	-6	9	-45
Var. in Temp.	-1	26	-19	-18	23	5	-18	-9	-22	19		88	-6	10	23	46	-6	-12	-16	-1	34
R. Humidity	-0	14	-21	2	-19	26	5	-10	-15	5		-13	6	32	-16	-73	59	-21	-8	-8	25
Var. in R.H.	-6	-16	13	2	16	-24	-22	17	17	3		22	-22	96	-27	-30	23	-7	8	28	12
H2O Hardness	-19	12	-7	-3	-8	6	0	-0	-6	1		5	7	-8	91	20	-9	-17	-16	3	4
Altitude	0	37	-36	-19	-7	30	-11	-4	-12	11		46	35	-21	13	94	-49	6	7	7	-8
Sulfates	2	-19	27	-12	-3	-30	-6	11	11	13		-6	-12	29	-4	-43	88	-7	-1	2	48
CO†																			66	5	12
NO_2†																				10	21
O_3‡	-9	-5	3	-14	-7	-5	-20	6	2	-11		9	3	29	7	9	9			100	-6
SO_2	2	2	2	-9	4	-11	-11	-12	-19	8		33	25	7	18	8	24			-4	94

*. Pearson correlation coefficients between the raw variables are in the upper right half, and between the filtered variables are in the lower left. Correlations between the raw and filtered versions of the same variable are shown on the diagonal.

‡. These variables did not exhibit spatial autocorrelation, and were thus not filtered.

†. These variables did exhibit spatial autocorrelation but a suitable distance could not be found at which the spatial autocorrelation could be removed.

Table 6.3a A comparison of using raw data versus using spatially filtered data in assessing the impact of metropolitan area-specific covariates on all-cause mortality rate ratio associated with sulfate exposure.

Ecologic Covariates	Independent MAs		Spatial Filtering	
	Sulfate	Covariate	Sulfate	Covariate
Sulfates Alone	1.25** (1.13, 1.37)		1.09 (1.01, 1.19)	
Demographic				
Population Change	1.16** (1.05, 1.29)	0.88 (0.80, 0.96)	1.1 (1.00, 1.20)	1.02 (0.86, 1.22)
Race				
Whites %	1.27* (1.15, 1.39)	1.06 (0.99, 1.14)	1.1 (1.01, 1.20)	1.03 (0.95, 1.11)
African Americans %	1.26** (1.15, 1.39)	0.96 (0.88, 1.04)	1.1 (1.01, 1.20)	0.99 (0.93, 1.05)
Health Services				
Doctors per 100000	1.23** (1.12, 1.36)	0.92 (0.84, 1.02)	1.09 (1.00, 1.19)	0.94§ (0.86, 1.03)
Hospital Beds per 10000	1.24** (1.13, 1.36)	1.13 (1.02, 1.25)	1.09 (1.00, 1.19)	1.08 (0.99, 1.17)
Socio-Economic				
High School Educated %	1.20** (1.08, 1.33)	0.92 (0.82, 1.02)	1.07 (0.98, 1.17)	0.92 (0.83, 1.02)
1979 Mean Income	1.23** (1.12, 1.35)	0.91 (0.83, 0.99)	1.09 (1.00, 1.18)	0.91§ (0.84, 0.98)
Poverty Rate	1.25** (1.14, 1.37)	0.97 (0.89, 1.05)	1.09 (1.00, 1.18)	1.01§ (0.94, 1.09)
Income Inequality	1.24* (1.13, 1.36)	0.92 (0.83, 1.01)	1.1 (1.01, 1.19)	0.96 (0.88, 1.06)
Unemployment Rate	1.22** (1.11, 1.34)	1.11 (1.01, 1.22)	1.08 (1.00, 1.18)	1.08 (0.98, 1.19)
Climate				
Mean Maximum Temperature	1.23** (1.11, 1.36)	0.94 (0.87, 1.01)	1.09 (1.00, 1.19)	0.95 (0.40, 2.26)
Variation in Max. Temp.	1.22** (1.11, 1.35)	1.13 (1.03, 1.25)	1.08 (0.99, 1.18)	1.07 (0.96, 1.19)
Mean Relative Humidity	1.17** (1.01, 1.35)	1.02 (0.90, 1.16)	1.12 (0.99, 1.26)	1.1 (1.03, 1.18)
Variation in Relative Humidity	1.21** (1.07, 1.37)	0.92 (0.83, 1.03)	1.09 (0.96, 1.24)	0.96 (0.86, 1.08)
Physical Environment				
Water Hardness	1.24 (1.11, 1.39)	1.07 (0.98, 1.17)	1.08 (0.98, 1.19)	1.08 (0.98, 1.20)
Elevation	1.24** (1.09, 1.40)	1.07 (0.95, 1.20)	1.08 (0.97, 1.19)	1.05 (0.95, 1.17)
Gaseous Co-pollutants				
Carbon Monoxide	1.25** (1.14, 1.37)	0.99 (0.90, 1.09)	Not possible to remove spatial autocorrelation in these covariates	
Nitrogen Dioxide	1.27** (1.13, 1.44)	0.93 (0.83, 1.04)		
Ozone	1.26** (1.15, 1.38)	0.93 (0.84, 1.02)	1.08 (1.00, 1.18)	0.96§ (0.87, 1.05)
Sulfur Dioxide	1.13 (1.02, 1.25)	1.27 (1.15, 1.40)	1.05 (0.97, 1.14)	1.19 (1.10, 1.29)

* The residuals are spatially autocorrelated at 0.10 confidence using a Moran's I test statistic.
** The residuals are spatially autocorrelated at 0.05 confidence using a Moran's I test statistic.

§ There was no evidence of spatial autocorrelation found in the original for these variables, therefore they were not filtered.

Table 6.3b A comparison of using raw data versus using spatially filtered data in assessing the impact of multiple metropolitan area-specific covariates on the all-cause mortality rate ratio associated with sulfate exposure.

Ecologic Covariate	Variables Included	Number of MAs	Relative Risk from Sulfates Alone	Adjusted Relative Risk From Exposure to	
				Sulfate	SO_2
Socio-economic and Demographic Covariates					
Independent MAs	Population change, Education, Unemployment,	139	1.24 (1.13, 1.37)	1.17 (1.05, 1.31)	
Spatial Filtering	Poverty, Income, Income inequality	139	1.11 (1.01, 1.21)	1.11* (1.01, 1.21)	
All Gases					
Independent MAs	SO_2, NO_2, CO, O_3	58	1.25 (1.10, 1.43)	1.05 (0.93, 1.18)	1.39 (1.24, 1.55)
Spatial Filtering	SO_2, O_3	102	1.09 (0.99, 1.19)	1.05 (0.96, 1.14)	1.19 (1.09, 1.29)
All Covariates Which Changed the RR from Sulfates by 25% on their Own					
Independent MAs	SO_2 and Population Change	103	1.28 (1.16, 1.42)	1.10 (0.99, 1.22)	1.10 (0.97, 1.24)
Spatial Filtering	SO_2, Altitude, Mean Relative Humidity, Education	50	1.05 (0.91, 1.22)	1.09 (0.94, 1.26)	1.07 (0.95, 1.22)

* The residuals are spatially autocorrelated at 0.10 confidence using a Moran's I test statistic.

** The residuals are spatially autocorrelated at 0.05 confidence using a Moran's I test statistic. (See Chapter 6)

Table 6.4a A comparison of using raw data versus using spatially filtered data in assessing the impact of metropolitan area-specific covariates on the cardiopulmonary mortality rate ratio associated with sulfates.

	Independent MAs		Spatial Filtering	
	Sulfate	Covariate	Sulfate	Covariate
Sulfates Alone	1.29 (1.15, 1.46)		1.13 (1.01, 1.27)	
Demographic				
Population Change	1.17** (1.03, 1.33)	0.84 (0.76, 0.94)	1.12 (1.00, 1.25)	1.08 (0.85, 1.37)
Race				
Whites %	1.32** (1.17, 1.50)	1.07 (0.98, 1.18)	1.14 (1.02, 1.28)	1.04 (0.94, 1.16)
African Americans %	1.31** (1.15, 1.49)	0.96 (0.86, 1.07)	1.14 (1.01, 1.28)	0.99 (0.91, 1.07)
Health Services				
Doctors per 100000	1.25** (1.11, 1.41)	0.85 (0.75, 0.96)	1.10 (0.98, 1.23)	0.85 § (0.76, 0.96)
Hospital Beds per 10000	1.27** (1.13, 1.44)	1.15 (1.01, 1.30)	1.11 (0.99, 1.25)	1.08 (0.96, 1.20)
Socio-Economic				
High School Educated %	1.22** (1.06, 1.39)	1.50 (1.14, 1.98)	1.08 (0.96, 1.22)	0.86 (0.74, 0.99)
1979 Mean Income	1.27** (1.13, 1.43)	0.87 (0.78, 0.97)	1.13 (1.01, 1.26)	1.00 § (1.00, 1.00)
Poverty Rate	1.29** (1.14, 1.46)	1.00 (0.91, 1.09)	1.12 (1.00, 1.26)	1.00 § (1.00, 1.00)
Income Inequality	1.29** (1.14, 1.46)	0.94 (0.83, 1.07)	1.16 (1.03, 1.30)	1.02 (0.90, 1.16)
Unemployment Rate	1.24** (1.10, 1.40)	1.21 (1.07, 1.36)	1.10 (0.99, 1.23)	1.19 (1.05, 1.36)
Climate				
Mean Maximum Temperature	1.25** (1.09, 1.42)	0.93 (0.84, 1.03)	1.12 (0.99, 1.26)	0.94 (0.83, 1.06)
Variation in Max. Temp.	1.23** (1.08, 1.40)	1.19 (1.05, 1.36)	1.10 (0.97, 1.24)	1.11 (0.96, 1.28)
Mean Relative Humidity	1.17** (0.96, 1.43)	1.05 (0.88, 1.26)	1.10 (0.91, 1.31)	1.09 (0.98, 1.20)
Variation in Relative Humidity	1.25** (1.06, 1.48)	0.89 (0.76, 1.03)	1.09 (0.90, 1.31)	0.93 (0.79, 1.09)
Physical Environment				
Water Hardness	1.26* (1.10, 1.44)	1.11 (1.00, 1.24)	1.11 (0.98, 1.26)	1.11 (0.97, 1.26)
Elevation	1.28** (1.09, 1.52)	1.04 (0.89, 1.22)	1.12 (0.96, 1.29)	1.02 (0.88, 1.19)
Gaseous Co-pollutants				
Carbon Monoxide	1.32** (1.18, 1.48)	0.94 (0.84, 1.05)	Not possible to remove spatial autocorrelation in these covariates	
Nitrogen Dioxide	1.35** (1.18, 1.55)	0.91 (0.81, 1.02)		
Ozone	1.32** (1.18, 1.48)	0.96 (0.85, 1.09)	1.13 (1.01, 1.26)	0.99 § (0.88, 1.12)
Sulfur Dioxide	1.18 (1.04, 1.34)	1.30 (1.16, 1.47)	1.10 (0.99, 1.22)	1.23 (1.11, 1.37)

* The residuals are spatially autocorrelated at 0.10 confidence using a Moran's I test statistic.
** The residuals are spatially autocorrelated at 0.05 confidence using a Moran's I test statistic.

§ There was no evidence of spatial autocorrelation found in the original for these variables, therefore they were not filtered.

Table 6.4b A comparison of using raw data versus using spatially filtered data in assessing the impact of multiple metropolitan area-specific covariates on the cardiopulmonary mortality rate ratio associated with sulfate exposure.

Ecologic Covariate	Variables Included	Number of MAs	Relative Risk from Sulfates Alone	Relative Risk From Exposure to	
				Sulfate	SO_2
Socio-economic and Demographic Covariates					
Independent MAs	Population change, Education, Unemployment, Poverty, Income, Income inequality	139	1.28 (1.13, 1.45)	1.18** (1.02, 1.37)	
Spatial Filtering		139	1.12 (1.00, 1.26)	1.12 (0.99, 1.27)	
All Gases					
Independent Cities	SO_2, NO_2, CO, O_3	58	1.31 (1.13, 1.52)	1.11 (0.97, 1.27)	1.42 (1.25, 1.61)
Spatial Filtering	SO_2, O_3	102	1.14 (1.02, 1.28)	1.10 (0.99, 1.23)	1.33 (1.17, 1.51)
All Covariates Which Changed the RR from Sulfates by 25% on their own					
Independent Cities	SO_2, Population Change and Education	103	1.34 (1.18, 1.51)	1.07 (0.93, 1.24)	1.29 (1.14, 1.46)
Spatial Filtering		72	1.10 (0.92, 1.31)	1.20 (1.01, 1.43)	1.19 (1.02, 1.38)

* The residuals are spatially autocorrelated at 0.10 confidence using a Moran's I test statistic.

** The residuals are spatially autocorrelated at 0.05 confidence using a Moran's I test statistic. (See chapter 6)

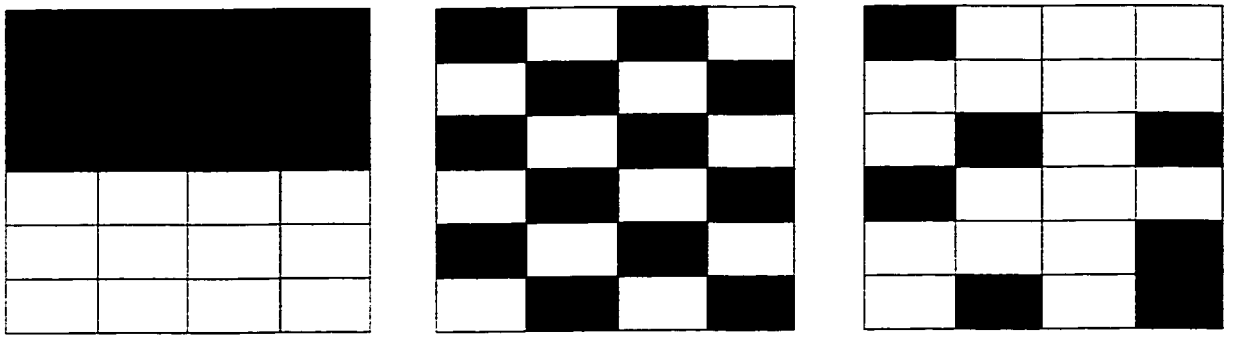


Figure 6.1 Examples of different types of spatial autocorrelation as measured by a Moran's I statistic.

Figure 6.2 Selected Counties and Centroids In New York and Adjacent States

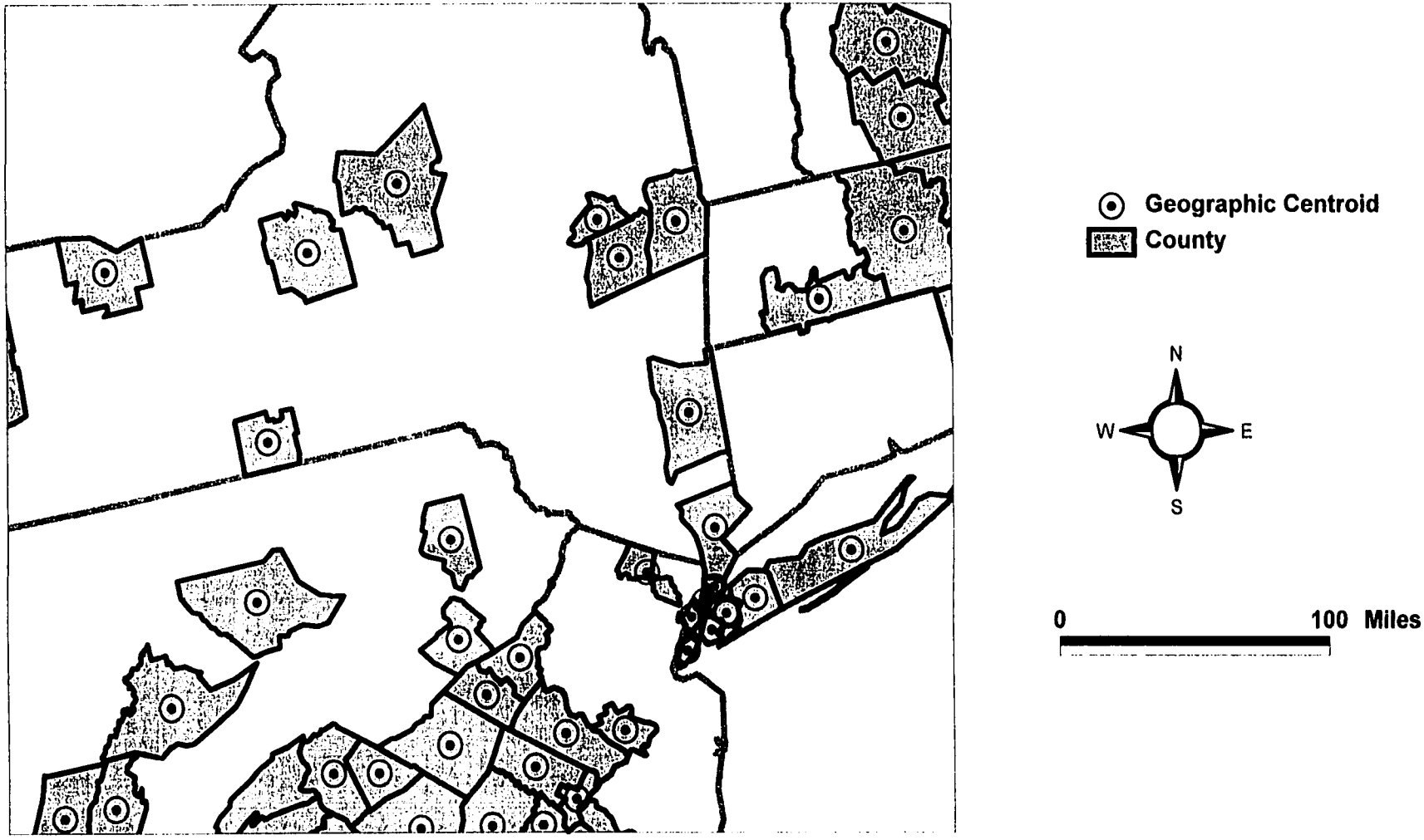


Figure 6.3. Semivariogram of All-cause Mortality for MAs

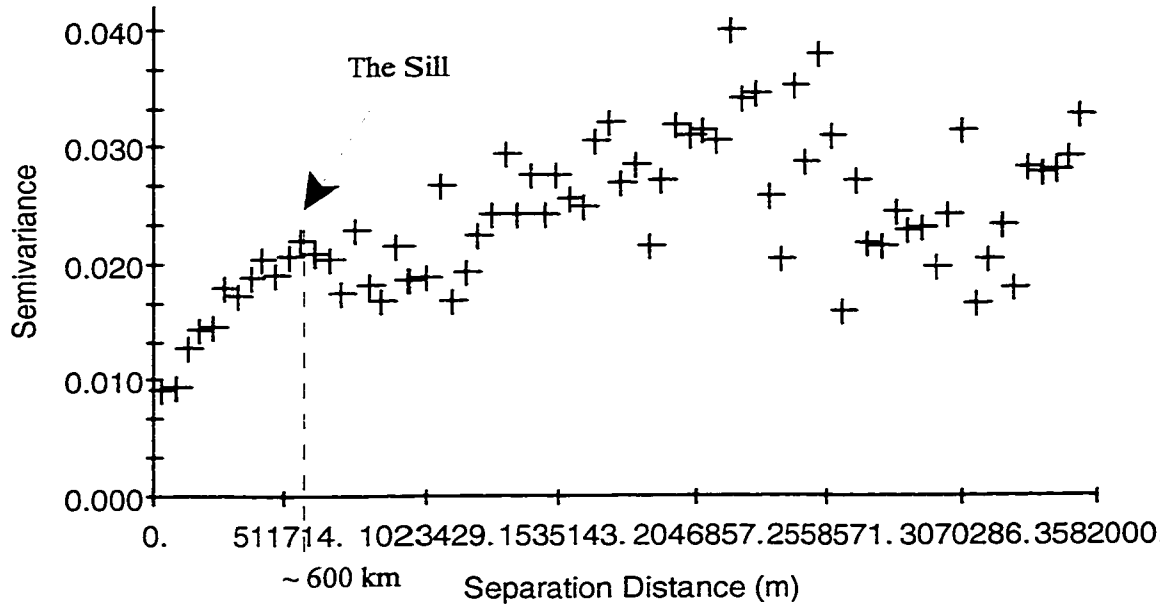
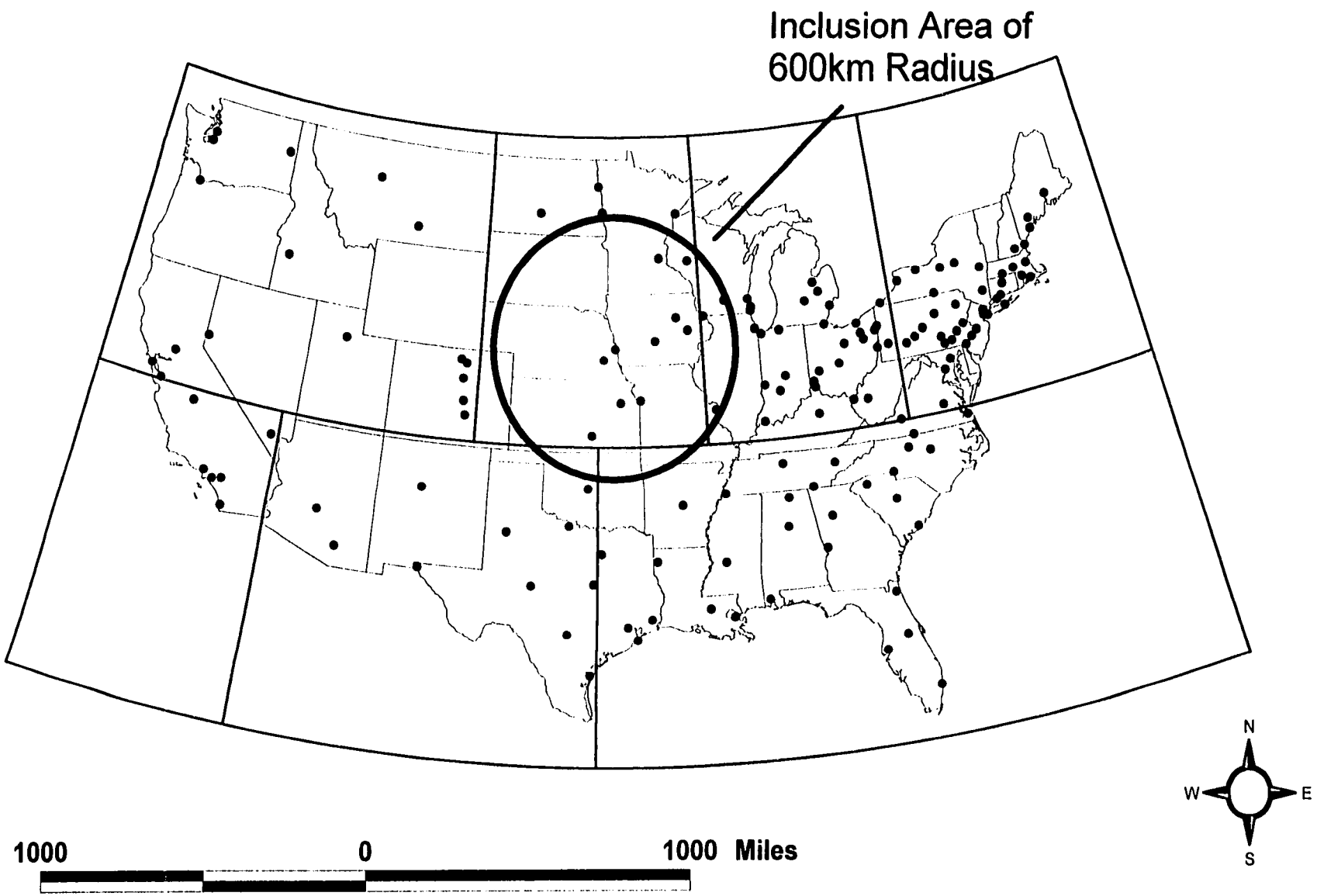


Figure 6.4
Distance Used in Calculating Spatial Statistics



7. Discussion

This thesis examines the truism that people are exposed to factors that impact their health in space, and that in a population health approach, epidemiology needs to pay more attention to space in order to understand disease. Medical geography – a discipline which, amongst other things, explores issues, of place, scale and patterns of influence in the context of health – is "concerned with the ways human behaviour, in its cultural and socioeconomic context, interacts with environmental conditions to produce or prevent disease among susceptible people" (Meade et al., 1988, p. 29).

In a sense, epidemiology's classic triangle of causation and confounding fits within the broader context outlined by Meade (see figure 7.1). The medical geography triangle is similar to the determinants of health approach to population health. Within each of the vertices of the second triangle, an investigator might find both relevant exposures and confounders. Every study of the risks presented by a particular exposure is done within the context of this broader framework of multiple causal factors for disease. By remembering that studies are located within this larger framework, it becomes easier to understand that space plays a part in most research, as two out of the three vertices involve exposures in space.

7.1 Place-specific Attributes

Within the discipline of geography, a great deal of qualitative research has been done into the impact places have on their inhabitants (cf. Willis, 1997; Pile, 1996; Duncan 1996, Keith and Pile 1993). Some of this work has come to the conclusion that places in and of themselves can negatively impact upon people's psycho-social well-being (cf. Sibley, 1995; Parr and Philo, 1995). Quantitative research in various fields has also delved into these issues (cf. Wilkinson, 1996). For the most part, however, epidemiologists have not embraced the concept of place in their quantitative research.

Many epidemiologists dismiss ecologic-level studies because of the so-called ecologic fallacy (cf. Morgenstern, 1998). However, if a population health approach is to be taken in epidemiology, investigators must accept that a legitimate area of research is how truly ecologic-level covariates (ones which are, in the words of Susser, integral or contextual) affect people's health. In this thesis, I have used the term place-specific ecologic-level covariates to differentiate contextual and integral variables that refer to attributes of geographic areas from ecologic-level variables in general.

This thesis has shown that place-specific social and environmental covariates, not to mention sulfates, can influence health outcomes at the individual level. At the metropolitan scale, a series of place-specific covariates were investigated as potential confounders. These variables were chosen from both the individuals' physical environment as well as from their socio-economic context (see Figure 7.1). When the covariates were individually entered into Cox proportional hazards regression models that include only sulfates, a few appeared to change the RR of mortality from sulfate exposure

(see Figure 7.2). For all-cause mortality and cardiopulmonary mortality, only sulfur dioxide and population change reduced the RR from sulfates to any great extent. Lung cancer mortality was a little different, with relative humidity and altitude causing the greatest decrease in risk. In the latter case, humidity, altitude and ozone were the only place-specific variables which had a significant RR on their own.

For all-cause and cardiopulmonary mortality, a large number of place-specific covariates had significant relative risks on their own. While most of the relative risks were in the expected direction, a few covariates performed contrary to expectation. Using a Cox regression model, income disparity, which has been widely found to be positively associated with mortality (see Wilkinson, 1996, for an overview of these studies), was found to be negatively associated with all-cause and cardiopulmonary mortality.

There are several potential explanations as to why, for the set of people in this cohort, income inequality appears to be protective. Since the other studies on income inequality examined in chapter 2 were done entirely at the ecological level, it is possible that the results are biased by a lack of control for the individual-level covariates that this study was able to take into account, such as smoking and alcohol consumption. It is also possible that the selection of metropolitan areas somehow affected the results. The RR of all-cause mortality from sulfates alone varied from 1.09 to 1.16 depending upon which sub-cohort of metropolitan areas was used. Another possible explanation stems from the make-up of the ACS cohort itself. As stated earlier, the members of the cohort were recruited by volunteers; they were not chosen randomly from a well-defined sampling

frame. The ACS cohort members in each metropolitan area are largely more educated than the general population (as measured in the 1980 census) and white people are over-represented (see Figures 7.3 and 7.4). The studies which have found income inequality to be positively associated with adverse health effects used the general population. Wilkinson (1996) postulates that income inequality causes health problems through psychosocial mechanisms, whereby the poorer segments of society suffer from stress and low self-esteem through being reminded of their lower status in comparison to the richer elements in their community. Perhaps the converse is true, that the well-off people have lowered stress and inflated self-esteem through being reminded of their higher status in comparison to the poorer people. It is conceivable that for well-educated white Americans, the income inequality of the place in which they live is actually protective.

The effect of the subset of cities on the observed effect of sulfates on mortality warrants some further discussion. The choice of places to study likely changes the observed relationship because of the distribution of sulfates over the US. The Re-analysis Team found some indication of a dose response relationship between sulfates and mortality (Health Effects Institute, 2000). If a subset of cities in which lower levels of sulfates are over represented is used, one would expect the observed risk to be lower. Added to this is the fact that in a smaller sample the power is diminished, making it even more difficult to detect these small effects. For these two reasons, subsets of the 151 cities for which sulfate data are available may demonstrate different relationships between sulfates and mortality.

Nitrogen dioxide and water hardness were also found to have an opposite

association than was expected. In chapter 2 I discussed my misgivings about including the gaseous co-pollutants as I felt they were too heterogeneous over metropolitan areas to be considered integral variables. In hindsight it was useful to include them because they have behaved as anticipated. Elsom (1992) review a large amount of literature that link nitrogen dioxide to health effects. Elsom also discusses the dispersion of nitrogen dioxide, namely that it varies over relatively small areas. What happens when a variable which varies over an area is treated as if it is an integral exposure to residents of that area? It behaves contrary to expectation. So even though there is a good deal of evidence that nitrogen dioxide is related to health effects, when it is measured at an inappropriate scale an inverse relationship is seen. The same process is likely at work with water hardness. Indeed, previous studies have demonstrated that the scale of analysis has a dramatic effect on the findings (Borgman, 1985; Sharrett, 1981; Comstock, 1979; Neri and Johansen, 1978).

Of course, the final potential explanation for these discrepancies is that there is a certain bias inherent to treating an ecologic-level variable as an individual-level variable. For one thing, while there are only 151 metropolitan areas, by assigning each of the 552,138 cohort members an individual value for the ecologic covariate, the analysis overstates the effective sample size, and correspondingly, the statistical significance of the results.

Indeed, when the place-specific covariates are more appropriately entered into the two-stage analysis as ecologic-level variables, fewer were found to be significantly associated with mortality. In fact, for lung cancer, ozone is the only place-specific

variable other than sulfates itself to have a significant rate ratio. Sulfur dioxide and population change continue to lower the effect of sulfates on both all-cause and cardiopulmonary mortality, but no longer reduce the sulfate RR to the point of insignificance. Education also decreases the RR of cardiopulmonary mortality from sulfates by more than 25%, although it does not affect all-cause mortality to the same extent.

Population change was conceived as a variable to measure the point on the economic boom and bust cycle that a particular area was in during the early 1980s. However, during this time, the economic health of a region was heavily tied into the type of industry that predominated. Those areas with high tech industry were experiencing growth, while those areas with heavy manufacturing were experiencing a decline. Unfortunately, these two types of technology also differ greatly in terms of their polluting emissions. Indeed population change at the MA scale is negatively correlated with both sulfur dioxide and sulfates at about 40%. Although 40% is not considered high enough to preclude using both variables in a regression analysis, to some extent they may be measuring the same thing. To exacerbate the problem, movers tend to be healthier that non-movers. Therefore, negative population growth may also indicate an increase in the relative proportion of individuals who are vulnerable to the effects of exposure to ambient air pollution.

The regional analyses give further clues as to why population change has such a powerful effect on mortality risk. Once an indicator variable is introduced that measures the broad region into which an MA fits, population change no longer poses a

significant mortality risk, nor does it change the RR from sulfates to any great extent. Region and population change apparently measure the same thing. It would appear that along with everything else population change measures, it also serves, in part, as an indicator of region itself. However, since the residuals from the independent MAs model with population change and sulfates were significantly spatially autocorrelated and the residuals from the model with sulfates alone and the regional indicator were not, the EPA airshed regions appear to provide a better means of controlling broad spatial patterns than population change. However, both reduced the sulfate effect by approximately the same amount.

That population change exhibits a strong spatial pattern is confirmed by filtering the data. In fact, so much of population change is spatially dependent, that its filtered and unfiltered values are only 7% correlated. The only other variable to exhibit such a change was relative humidity. Interestingly, when spatial autocorrelation is removed using the more sophisticated filtering methods (as opposed to the regional indicators), more place-specific variables reduce the rate ratio from sulfates by 25% or more (see Figure 7.5). Mean relative humidity, sulfur dioxide and education all change the RR of both all-cause and cardiopulmonary mortality by at least 25%, while altitude also affects the RR of all-cause mortality, and variation in relative humidity affects the RR of cardiopulmonary mortality.

Spatial filtering removes the broad regional trends in the data, leaving behind the finer-grained changes in a variable's values. The fact that more covariates change the sulfate effect by 25% or more after filtering (and even more reduce the sulfate RR to the

point of insignificance) indicates that at least part of the sulfate effect seen earlier (in the two-stage analysis for example), was due to the coincidence of broad spatial trends in data.

When the scale of measurement was changed from MAs to counties, the rate ratio of either all-cause or cardiopulmonary mortality from sulfate exposure became more robust to the inclusion of place-specific ecologic-level covariates. In fact, when sulfates were included in the two-stage regression model for all-cause mortality, all of the ecologic variables were either not significant or barely significant (with a lower 95% confidence limit near 0.99 or 1.00) and no ecologic covariate changed the RR from sulfates by 25% or more. In contrast to the metropolitan-scale analysis, sulfur dioxide is not significantly associated with all-cause mortality when sulfates are included in the model. However, it is still one of the two variables (the other being population change) that have the greatest impact on the RR from sulfates.

At the county scale there are two ecologic-level variables that decrease the RR of cardiopulmonary mortality associated with sulfates by 25% or more: population change and the unemployment rate. Because of the spatial arrangement of counties, it was not possible to do either a regional or filtered analysis at the county scale. However, assuming that population change exhibits similar characteristics at the county scale as at the MA scale, it is most likely operating as an indicator of region again here.

It appears that the unemployment rate, which is measuring the place-specific construct of employment insecurity, partially confounds the relationship between sulfate exposure and cardiopulmonary mortality. Even when unemployment is controlled for,

however, the RR of sulfates remains strongly significant.

7.2 Scale

As demonstrated in the preceding discussion of place-specific attributes, there are two issues of scale that can affect the results of an analysis involving ecologic-level covariates. The first is that the scale of measurement affects the observed relationships between the variables of interest. The second is that different results are obtained when variables are analysed at the same scale as they were measured as compared to when they are treated as referring to a construct on a different scale.

I will discuss the second statement first, as it is the least complicated. Conceptually, it is not correct to treat as an individual-level covariate, a covariate that actually measures a place-specific attribute like the average ambient sulfate levels in a metropolitan area. Statistically, assigning the same area attribute to each individual within an area contradicts the assumption that the individuals represent independent observations, (Kleinbaum et al, 1988), as all individuals within an MA will have the same value for that variable.

The two-stage method of analysis developed by the Re-analysis Team (Health Effects Institute, 2000) allows investigators to treat individual-level data and ecologic-level data appropriately within the framework of a cohort study. The difference in the results obtained when the two-stage method is used, as opposed to the Cox model used in the original study, can be seen in figure 7.7. The estimated relative risk of all-cause mortality from sulfates is higher in the independent MAs model than in the Cox model,

and the 95% confidence interval is larger, due to the incorporation of both the variance within and between MAs. Similar differences can be seen when comparing the Cox and the independent MA models for cardiopulmonary and lung cancer mortality although the difference in RR estimates for cardiopulmonary mortality is less than that for all-cause and lung cancer.

Although the importance of the scale of measurement of place-specific variables was touched upon in section 7.1, a more complete discussion of the issue is warranted. The root of this issue is that, in order to avoid the ecologic fallacy, ecologic-level covariates must be either integral or contextual; that is to say, they must represent exposures which are the same for all individuals within the group (Susser, 1994a). By no means are all groups in epidemiological studies defined on the basis of geography, but in the ACS Study, they were. For the sake of distinguishing contextual and integral ecologic-level variables which are associated with places from those which are associated with groups defined according to some other characteristic, I have called the geographical variables place-specific ecologic covariates. Therefore, in order for these place-specific covariates to have construct validity, they must actually be integral or contextual variables. In the ideal case, every individual within an area is exposed to an integral variable to the same degree. In practical terms, the actual variation in exposure to that variable amongst the individuals within the area should be smaller than that between individuals in different areas.

Thus, as discussed above, one of the reasons that variables such as nitrogen dioxide did not behave as expected in relation to mortality may be that it is not a truly

integral variable at the metropolitan scale. What is clear from the reviews of the literature on water hardness provided by Sharrett (1981) and Comstock (1979) is that the association between water hardness and health is sensitive to the scale of analysis used. The most appropriate scale at which to measure water hardness would be at the scale of drinking water provision zones or of geological substrata.

However, sulfates, not water hardness, were the focus of this study, and the scale at which it is appropriate to measure water hardness is probably not the best scale at which to measure sulfate exposure. Obviously it is not possible at this time to conduct multi-level analyses that allow for each variable to be measured and incorporated at the exact scale at which it operates. No such methods exist; even if appropriate methods were available, it is by no means apparent that it would be more desirable to use them than to approximate integral and contextual variables. Instead, I propose a guideline for a place-specific variables: exposure to a place-specific variable within an area should be more homogenous than exposure between areas.

In comparing the metropolitan-scale analysis to the county-scale analysis, there is reason to believe that the county-scale might be the more appropriate scale of measurement. As Figure 7.7 illustrates, the estimated RR for all-cause mortality associated with long-term sulfate exposure from the county-scale analysis is much higher than that from the metropolitan-scale analysis (1.50 as opposed to 1.25).

As explained in chapter 3, the US Census Bureau aggregates metropolitan areas from counties. The county scale, therefore, is smaller than the MA scale. Since points (monitors) were the source of the air pollution data and both the MA and county-

scale analyses used sulfate data aggregated from these points, there is reason to believe that exposure may be more accurately measured at the county scale, as it was averaged over fewer values. Moreover, the method by which cohort members were assigned to MAs meant that the area the cohort represented was actually larger than the MAs and, therefore, larger than the area for which exposure was calculated. By using the complete 5-digit ZIP codes, instead of only the first 3-digits, in assigning individuals to counties, the area covered by the cohort members matched the area over which exposure was assessed more accurately. Therefore, in all likelihood, the true variation in exposure of individuals within an area to sulfates is probably smaller in the county-scale analysis than in the metropolitan-scale analysis.

In essence, using an inappropriate scale of measurement for a place-specific covariate represents a misclassification error. As one would expect in a situation in which there is a misclassification error (Hennekens and Buring, 1988), the RR of mortality associated with sulfate exposure is less in the problematic Metropolitan-scale analysis than in the county-scale analysis. Because of the suspected misclassification at the MA scale, the rate ratios of mortality in the independent county models probably provide a more accurate estimate of the effect of long-term sulfate exposure than do the independent MA models. The county-scale RR estimate for all-cause mortality is 1.50 (1.30, 1.73), and the RR for cardiopulmonary mortality is 1.75 (1.48, 2.08). Unfortunately, I was unable to perform a county-scale analysis of the effect of sulfate exposure on lung cancer because there were too few (sometimes zero) cases of lung cancer deaths in some of the counties. The analytic techniques outlined in this thesis

break down in such a situation.

Of course, while misclassification may be less of a problem in the county analysis, residuals of all the county models did exhibit significant spatial autocorrelation. Consequently, the estimates of risk may be biased, and the tests of significance may be inaccurate.

7.3 Spatial Autocorrelation

Issues relating to spatial autocorrelation and methods to test for it have been recognized by medical geographers for some time. Meade et al. (1988) cite a 1976 study that used spatial autocorrelation to study disease diffusion. While many epidemiologists are now comfortable with time-series analyses, which require filtering for temporal autocorrelation, epidemiology has not given spatial autocorrelation the same attention. One reason for this may be that ecologic studies – the studies which most commonly use geographical data – have been considered to be less desirable than individual-level studies such as cohort and case-control studies. With the advent of new analytical methods which can incorporate both individual-level and ecologic-level data and the upsurge in interest in population health approaches, there may be renewed interest in place-specific ecologic-level covariates.

Because ecologic studies have been regarded with skepticism in the past, investigators who choose to examine the influence of place-specific ecologic covariates on health outcomes need to be extremely careful to avoid bias in their studies. This means that spatial autocorrelation will need to be tested for, and adequately controlled for

if it is detected, otherwise, the risk estimates might be biased and the standard tests of significance may not apply.

By conducting analyses at the metropolitan scale with an indicator variable for EPA airshed regions, a larger scale was incorporated into the research. However, the regional indicator variable also served as a rudimentary control for spatial autocorrelation. When it was included in the two-stage analysis, the spatial autocorrelation in the residuals of the models was eliminated in most cases. The exceptions were altitude in the all-cause mortality models, and variation in maximum temperature in the cardiopulmonary models. These two variables were only available for 110 and 135 MAs respectively. Therefore, as with the set of 132 counties, the spatial matrix of MA points may not have been appropriate for the use of an indicator variable for controlling spatial autocorrelation. Provided that appropriate regional boundaries can be derived for the exposure variable, it appears to be possible to eliminate spatial autocorrelation using an indicator variable for region.

Although the EPA airshed regions were defined prior to testing the variables for spatial autocorrelation, one of the reasons why their incorporation so successfully removed the autocorrelation was because the size of the regions happened to be roughly the same size as double the distance at which the sill was observed for the sulfate data. This distance was used as the radius for the areas used to derive the spatial matrix in calculating the spatial filters. In other words, both the regional analysis and the filtering operated at the same scale to remove spatial autocorrelation from the models.

Comparing the regional model to the filtered model for all-cause mortality (see

figure 7.8), it can be seen that despite the fact that both methods operate on more or less the same scale, and both successfully remove spatial autocorrelation from the models, the estimated RR in the filtered model (RR of 1.09 with a 95% CI of 1.01 to 1.19) is lower than that in the regional model (RR of 1.19 with a 95% CI of 1.06 to 1.34). The question is, why do these two methods produce different estimates.

The use of regions defined on the basis of previous knowledge regarding the spatial processes underlying the exposure of interest is conceptually appealing and is relatively easy to interpret. The regional analysis is based on an understanding that different regions of the US have different air pollution profiles. When these broad regional trends are controlled for, what remains are the more pollutant-specific smaller area fluctuations. It appears that the smaller scale changes in sulfate levels are associated with the smaller scale changes in mortality rates within the cohort.

The spatial filtering method based on the Getis statistic outlined in chapter 6 can be understood in a similar way, only the regions are relative rather than fixed. With this spatial filtering approach, the regional influence is defined as being an area with a radius of 600km around each MA. In other words, while the regional analysis assumes that each MA is influenced by the airshed in which it is located, the filtered analysis assumes that each MA is influenced by the area, 600 km or less, surrounding it. As both techniques remove spatial autocorrelation from the models, both have merit.

7.4 Conclusions

As epidemiologists, our role is both to use the best available analytic methods

and to provide useful information to population health risk management decision makers. So, while I hope that this thesis has demonstrated that spatial issues are central to many population health questions and may bias results if they are not taken into account, I would also like to offer some discussion of the sulfate effect.

Despite the fact that some of the analyses presented in this thesis reduced the relative risk associated with sulfates to the level of insignificance, the weight of evidence points in the direction of confirming an association between long-term exposure to ambient sulfates and all-cause, cardiopulmonary, and lung cancer mortality.

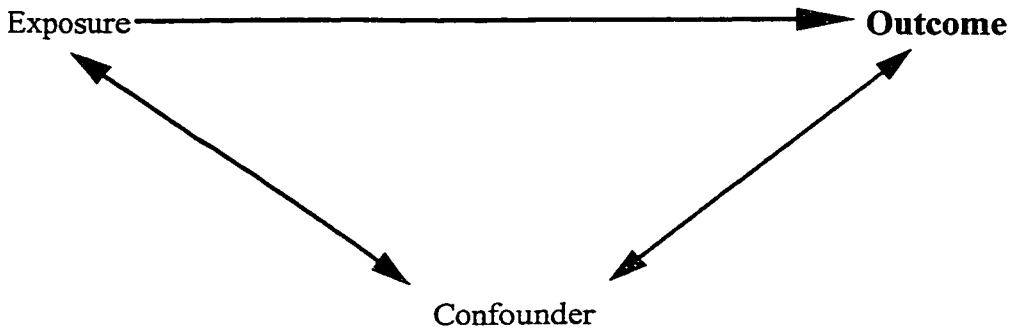
The most telling results are those obtained at the county scale, where the risk of misclassification error in assigning exposure is lowered. These results showed higher rate ratios associated with sulfates, which were more robust to the inclusion of place-specific covariates, including sulfur dioxide. However, the residuals of all the county-scale models exhibited highly significant spatial autocorrelation, indicating that these estimates are probably biased. Unfortunately, there were not enough counties in the western half of the US with sulfate data to enable either a regional or a filtered analysis, so the spatial autocorrelation could not be fully controlled for. It is also unfortunate that the methods outlined herein could not be applied to lung cancer at the county-scale because of the lack of lung cancer deaths in some areas.

Many analyses were done for this thesis leading to concerns that the findings are merely the result of multiple testing. However, this work was done in the broader context of research on the health effects of particulate air pollution. The ACS Study is one of three cohort studies that have been examined long-term exposure to particulates

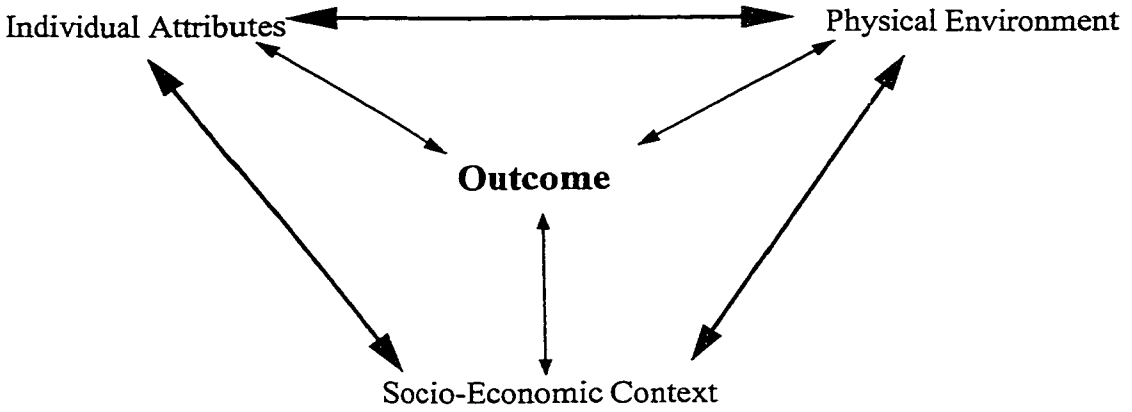
(of which sulfate is one component). All three cohort studies have shown significant positive associations between chronic exposure and mortality (Pope et al., 1995; Dockery et al. 1993; and Abbey et al., 1991). Like the ACS Study, the Harvard Six-cities Study found a significant association between sulfates and all-cause mortality (Docker et al. 1993). Burnett et al. (1998) provide a comprehensive overview of the literature on the health effects of particulates. The results of this thesis add to the growing weight of evidence linking particulates, and sulfates specifically, to adverse health outcomes.

The rate ratios found in the county analyses were larger than those found in the metropolitan-area analyses, but they are still relatively small risks (1.50 for all-cause mortality and 1.75 for cardiopulmonary mortality). However, because the vast majority of people are exposed to ambient sulfate air pollution, and the levels of sulfates are higher in the more populated Eastern half of the US, sulfates likely represent a significant risk factor from a population health perspective.

In summary, while it appears that long-term exposure to ambient sulfur air pollution is associated with all-cause mortality, cardiopulmonary mortality and lung-cancer mortality, this thesis was not able to calculate accurate unbiased estimates of the magnitude of this association. It is hoped that subsequent investigations will be able to obtain sulfate data for a greater geographical spread of counties so that the spatial autocorrelation at this scale of analysis can be controlled for. Also, if data on the next nine years of ACS follow-up can be obtained, it may be possible to examine the association between sulfate exposure and lung-cancer at the county scale, as more lung cancer deaths may have occurred during this time period.



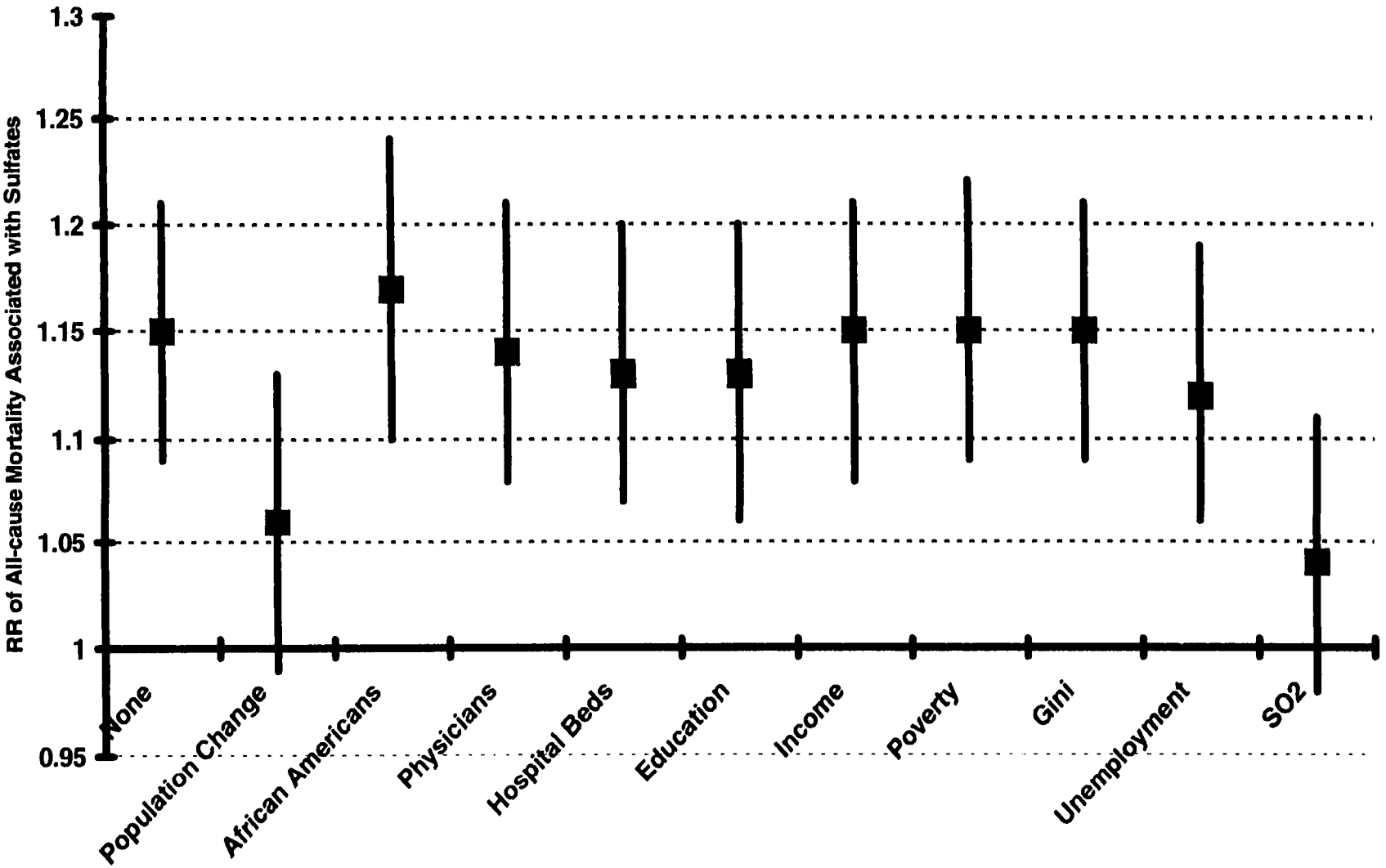
The Classic Epidemiological Triangle of Confounding



A Modification of the Medical Geography Triangle.

Figure 7.1. A Comparison of Etiological Models

Figure 7.2. Impact of Place-specific Ecologic Covariates on the Sulfate Effect
(Rate ratios shown with 95% confidence intervals)



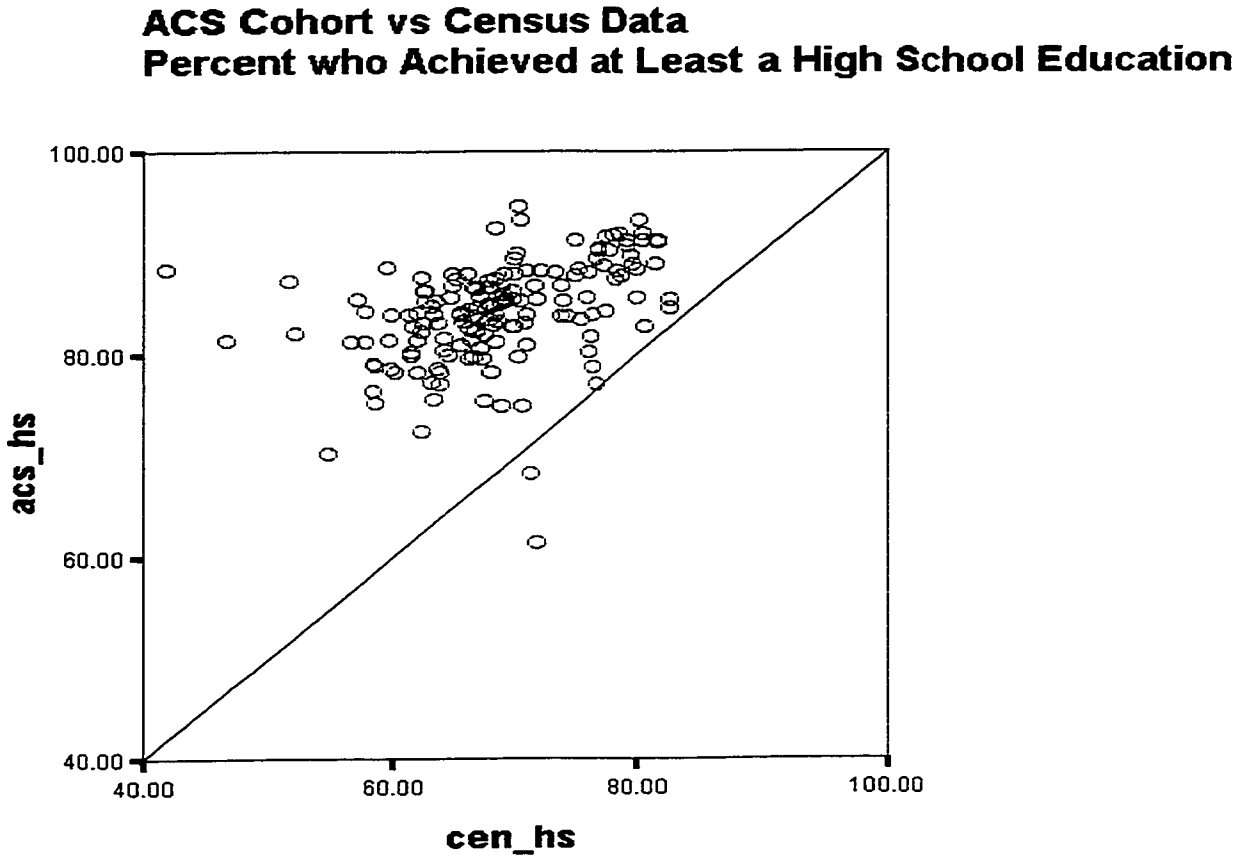


Figure 7.3. A comparison of the ACS cohort to Census data for metropolitan areas in the US.

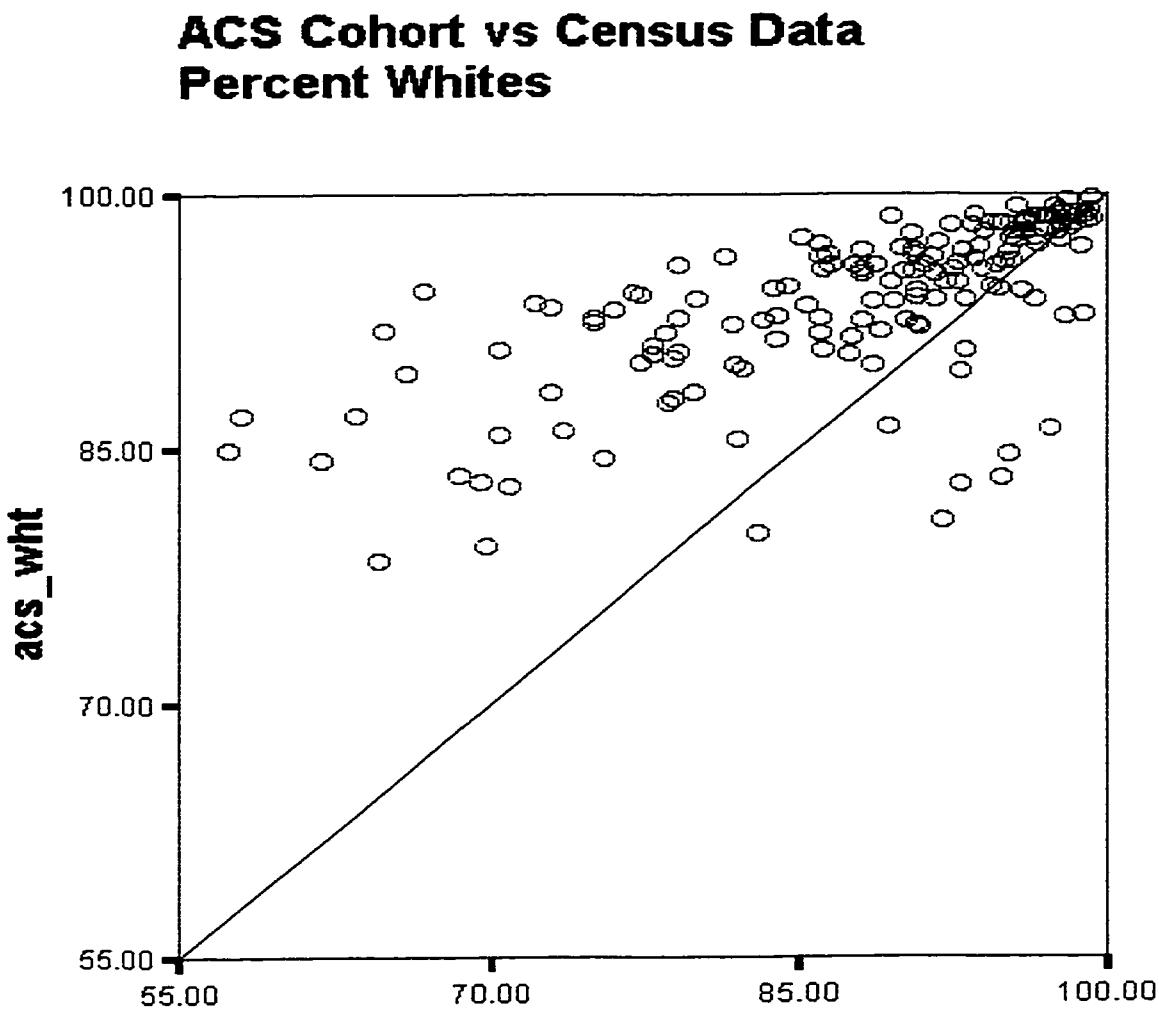


Figure 7.4. A comparison of the ACS cohort to Census data for metropolitan areas in the US.

Figure 7.5. Impact of Filtered Place-specific Covariates on Filtered Sulfate Effect
(Rate ratios shown with 95% confidence intervals)

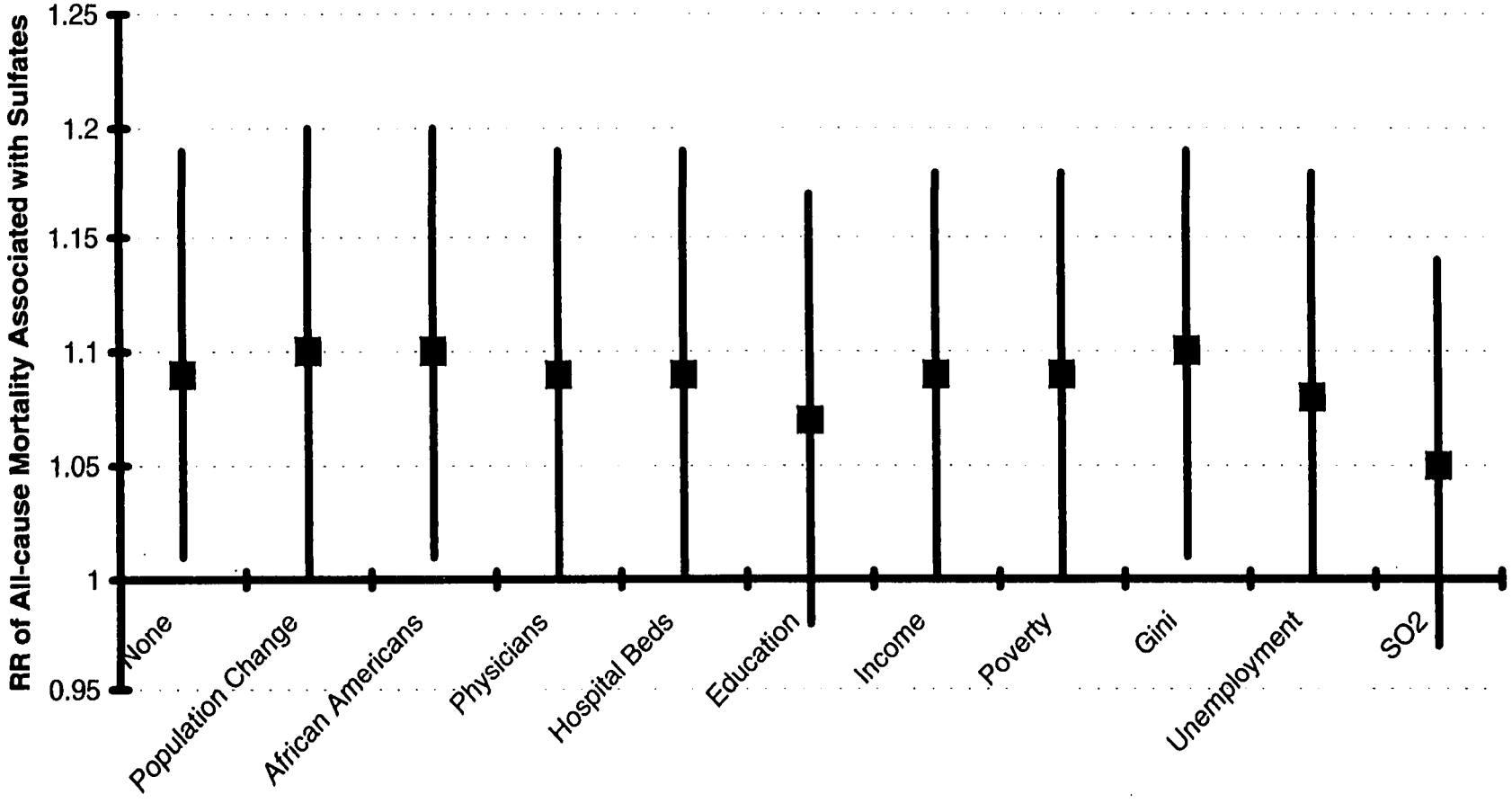


Figure 7.6. Impact of County-scale Place-specific Covariates on the Sulfate Effect
(Rate ratios shown with 95% confidence intervals)

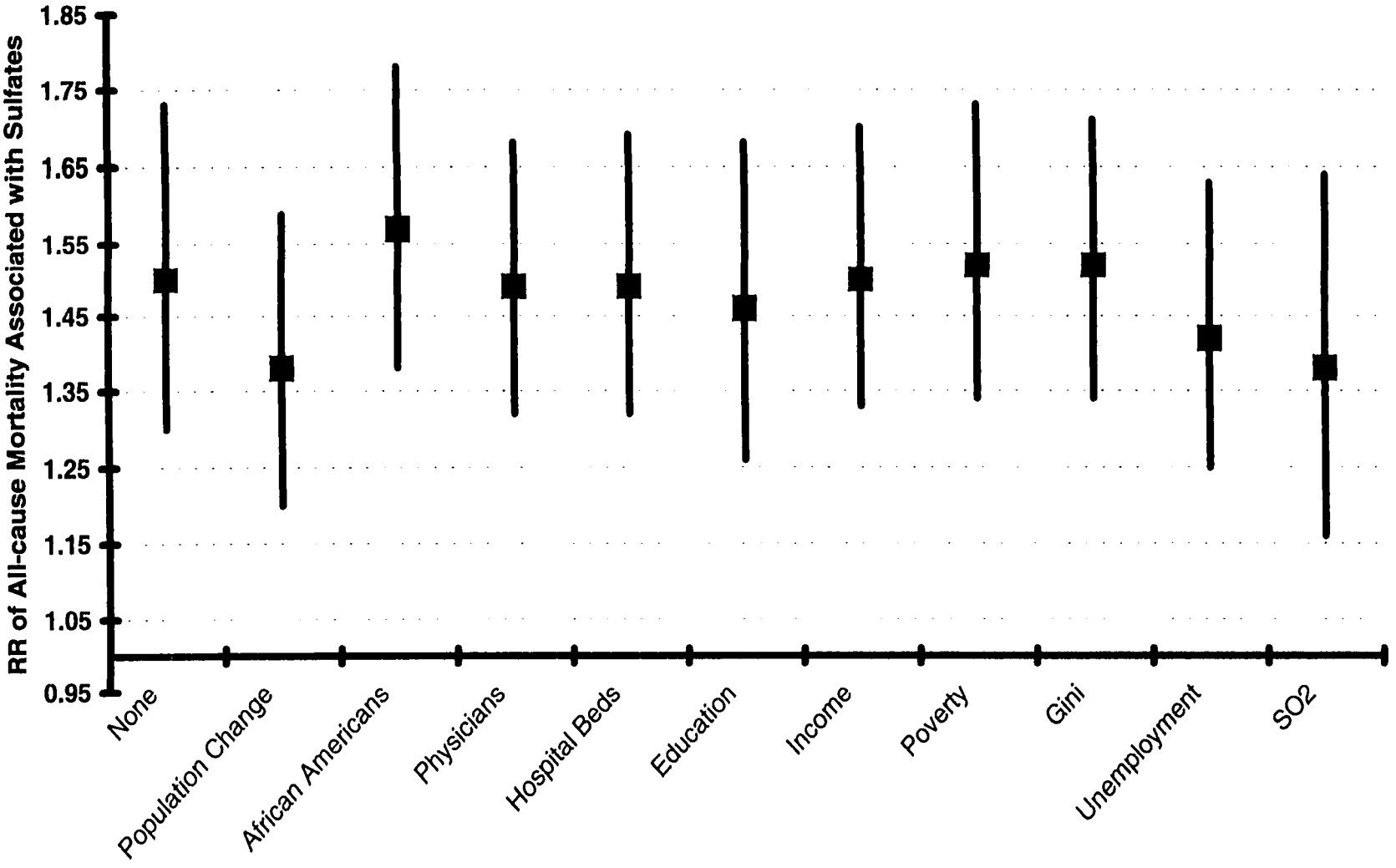


Figure 7.7. Impact of Scale of Analysis on Sulfate Effect
(Rate ratios shown with 95% confidence intervals)

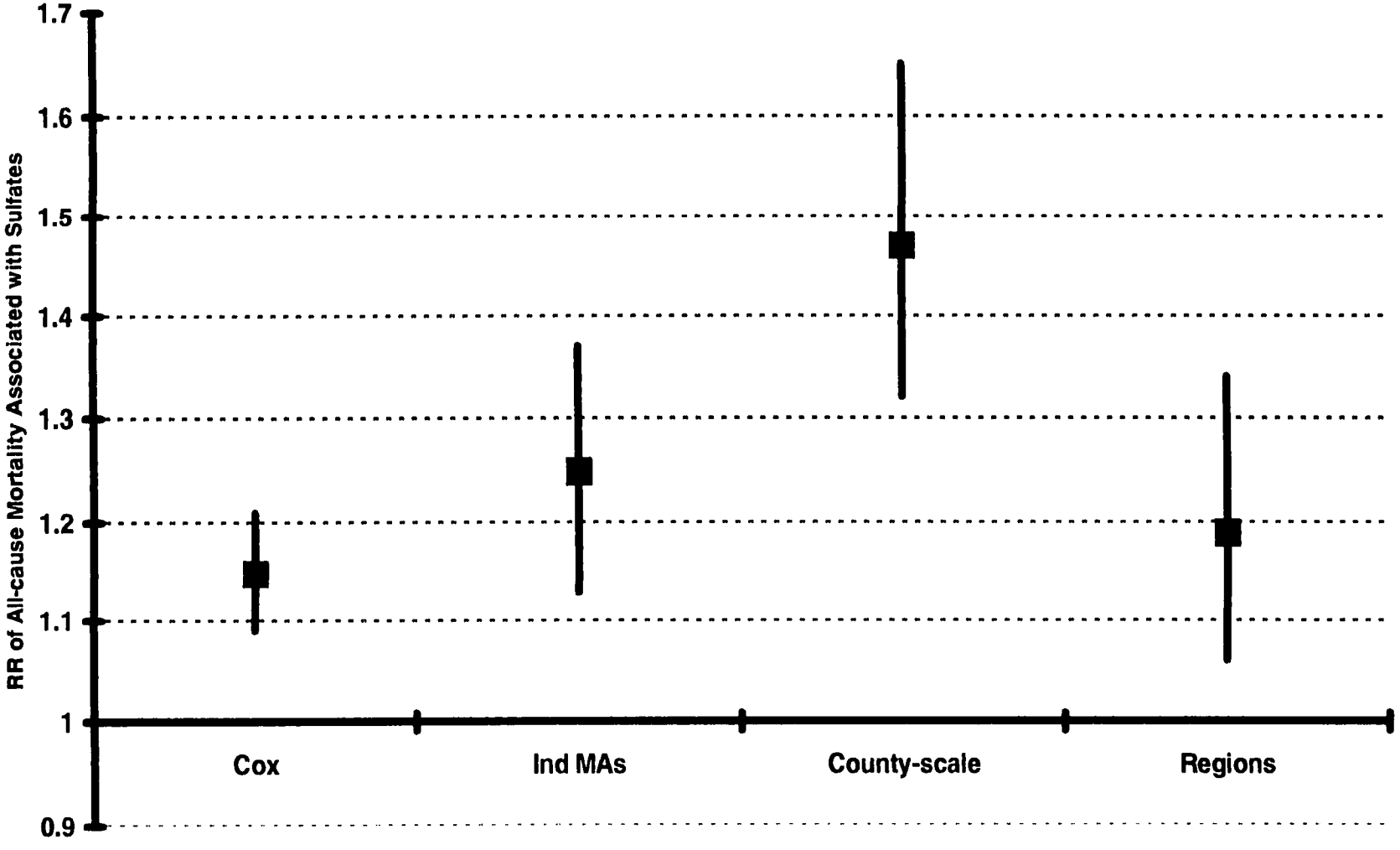
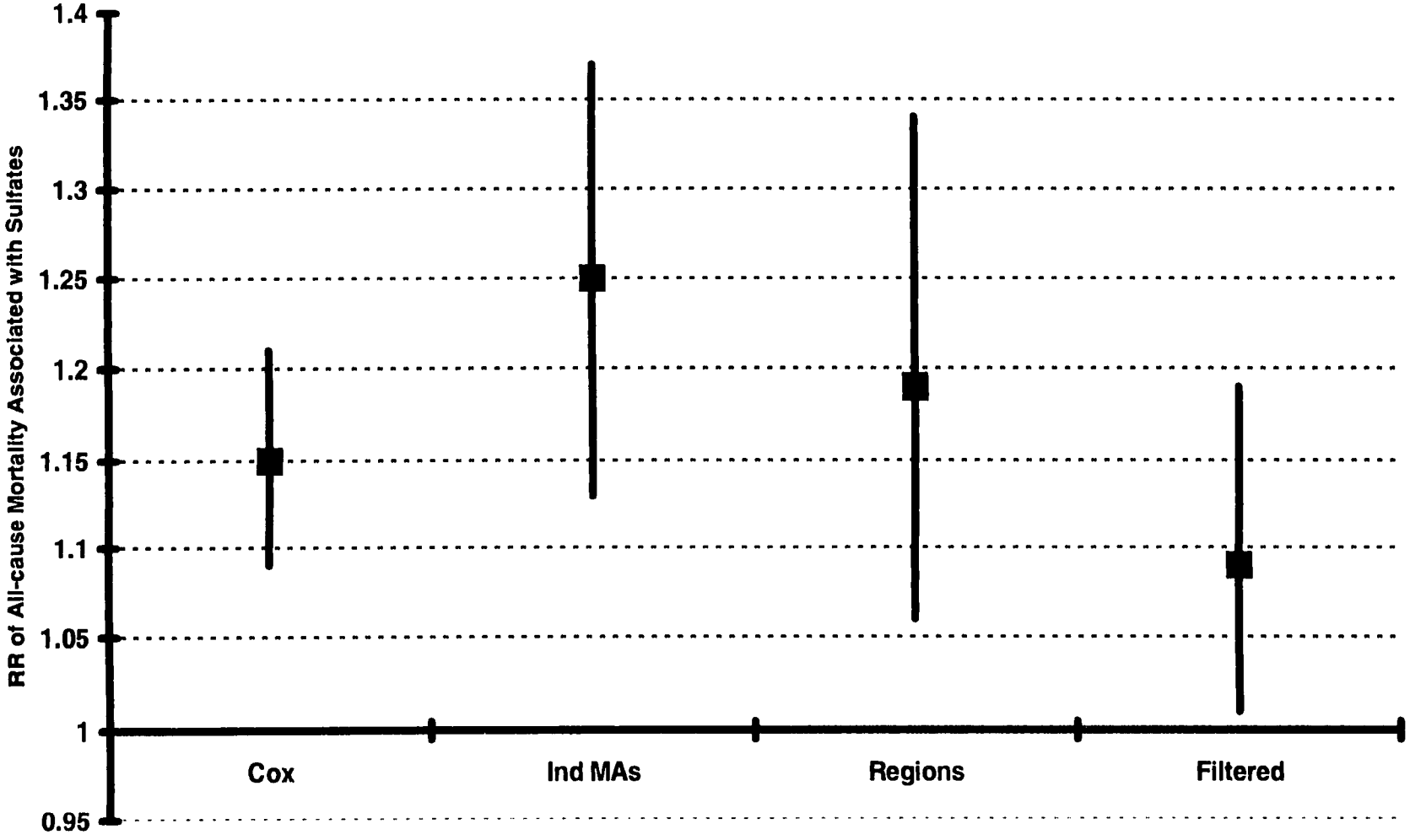


Figure 7.8. Impact of Spatial Autocorrelation on Sulfate Effect
(Rate ratios shown with 95% confidence intervals)



8. Further Research

Interestingly, both epidemiology and medical geography trace their roots back through John Snow's study of cholera in London, and claim Snow as one of their own (Hennekens and Buring, 1987; Meade et al., 1988). At one time, therefore, epidemiology and medical geography were effectively one and the same. Moreover, Snow's approach to discovering the source of the epidemic was inherently spatial. Since then, however, epidemiology has diverged from medical geography, and has lost interest in the role space plays in disease etiology. Now that certain sectors of epidemiology are redefining themselves as population health investigators, space may regain some of its lost prominence in epidemiology.

This thesis has demonstrated that the inclusion of ecologic-level variables that refer to place-specific constructs can change the estimated relative risk from long-term exposure to ambient sulfates. Further research is needed on other potential place-specific confounders and on the effect of including place-specific covariates on estimates of risk from exposure to other factors.

Changing the scale of measurement for both the exposure of interest (sulfates) and the ecologic-level confounders also changes the findings. Basic research into the scale at which different processes impact on health is needed. What are the physical processes in the distribution of sulfate air pollution? At what scale can sulfate exposure

really be called integral: county, city, region? At what scale does income inequality effect people: is it the inequality within a neighbourhood, within a city, within a country?

While the two-stage method employed in this thesis enabled the inclusion of both individual-level and ecologic-level data, it only allowed for one geographic scale for the ecologic data. For population health approaches to be most effective, methods need to be developed which can efficiently incorporate data at many different scales of measurement.

Spatial autocorrelation was found in the residuals of most of the independent area models. To the best of my knowledge, spatial autocorrelation has not been tested for before in epidemiological studies. Given the implications of its presence in residuals, the lack of homoscedasticity, a great many results in ecological epidemiology may be biased and their p-values erroneous because of spatial autocorrelation. Future epidemiological research in which subjects are grouped by geographic location, or in which areas themselves are the cases (ecologic-level studies), should therefore include tests for spatial autocorrelation, and employ appropriate spatial methods of analysis if it is present.

The spatial filtering methods used in this thesis to remove spatial autocorrelation are only effective on the scale chosen as the filtering distance. Any spatial autocorrelation on broader scales than the filtering distance may remain. Methods need to be developed which control for spatial autocorrelation at more than one scale.

While the primary goal of this thesis was to explore the impact of spatial issues on a cohort study which examined an ecologic-level exposure, it has also raised the question of whether that exposure (sulfates) is related to mortality. While this question is

dealt with more definitively in the Re-analysis report (Health Effects Institute, 2000), some conclusions may be drawn here. Although it is difficult to specify the actual quantitative risk posed by long-term exposure to sulfates given all of the problems discussed in this thesis, it does appear that there is a risk.

One of the difficulties in conducting epidemiologic research on the health effects of ambient air pollution is the high correlation between the different combustion-source pollutants. However, no one pollutant has exactly the same distribution as another. That is to say, no one ambient air pollutant operates in exactly the same way spatially. While the more micro-level epidemiologists, toxicologists and medical investigators are working hard to find the biological mechanisms by which different pollutants might act on a person's health, population health epidemiologists can contribute to this endeavour by taking space into account in their population based studies. All of the further research outlined in this chapter may be usefully applied to the goal of identifying which ambient air pollutant(s) causes which adverse health effect.

Overall, the research agenda outlined here shows the importance of trans-disciplinary research into population health issues. Epidemiologists alone will not be able to address these research questions. Instead, research teams involving epidemiologists, geographers, sociologists, biostatisticians, atmospheric scientists, and others are needed to explore how spatial processes cause disease or promote good health. Space matters.

References

Abbey DE, Mills PK, Petersen FF, and Beeson WL. 1991. Long-term ambient concentrations of total suspended particulates and oxidants as related to incidence of chronic disease in California Seventh-day Adventists. Environmental Health Perspectives; 94:43–50.

Adami J, Gridley G, Nyren O, Dosemeci M, Linet M, Glimelius B, Ekbom A, Zahm SH. 1999. Sunlight and non-Hodgkin's lymphoma: a population-based cohort study in Sweden. International Journal of Cancer; 80(5): 641-5.

Amick 3rd, BC, Levine, S, Tarlov AR, Walsh, DC, eds. 1995. Society and Health. Oxford University Press, Oxford, England.

Amsel J, Waterbor JW, Oler J, Rosenwaike I, Marshall K. 1982 Relationship of site-specific cancer mortality rates to altitude. Carcinogenesis; 3(5): 461-5.

Anonymous. 1998. State differences in reported healthy days among adults -- United States, 1993-1996. MMWR - Morbidity & Mortality Weekly Report; 47(12): 239-43.

Borgman, RF. 1985. Dietary factors in essential hypertension. Progress in Food & Nutrition Science; 9(1-2): 109-47.

Brenner MH, Savitz DA, Jockel KH, Greeland S. 1992. Effects of nondifferential exposure misclassification in ecologic studies. American Journal of Epidemiology 135: 85–95.

Brenner MH, Mooney A. 1983. Unemployment and health in the context of economic change. Social Science & Medicine; 17(16): 1125-38.

Brown ME, Bindman AB, Lurie N. 1998. Monitoring the consequences of uninsurance: a review of methodologies. Medical Care Research & Review; 55(2): 177-210.

Bull GM, Morton J. 1978. Environment, temperature and death rates. Age & Aging; 7(4): 210-24.

Bullard, RD. (editor) 1993. Confronting Environmental Racism: Voices from the Grassroots. Boston: South End Press.

Burnett, R.T., Jessiman, B., Steib, D. Krewski, D. 1998. Health effects of ambient particulate matter: A Canadian perspective. Health Effects of Particulate Matter in Ambient Air. Edited by J.J. Vostal. Pittsburgh: Air and Waste Management Association.

Burnett R, Krewski D. 1994. Air pollution effects on hospital admission rates: A random effects modeling approach. Canadian Journal of Statistics; 22: 441–458.

Carr W, Zeitel L, Weiss K. 1992. Variations in asthma hospitalizations and deaths in New York City. American Journal of Public Health; 82(1): 59-65.

Catalano R.1991. The health effects of economic insecurity. American Journal of Public Health; 9:1148-52.

Christenson BA, Johnson NE. 1995. Educational inequality in adult mortality: an assessment with death certificate data from Michigan. Demography; 32(2): 215-29.

Clay CM, Dyer AR, Liu K, Stamle, J, Shekelle RB, Berkson DM, Paul O, Schoenberger JA, Lepper M, Lindberg HA., et al. 1988. Education, smoking and non-cardiovascular mortality: findings in three Chicago epidemiological studies. International Journal of Epidemiology; 17(2): 341-7.

Cohen BL. 1995. Test of the linear-no threshold theory of radiation carcinogenesis for inhaled radon decay products. Health Physics; 2:157-74.

Colledge M. 1982. Economic cycles and health. Towards a sociological understanding of the impact of the recession on health and illness. Social Science & Medicine; 16(22): 1919-27.

Committee on Environmental Justice. 1999. Toward Environmental Justice: 1999. Research, Education, and Health Policy Needs. National Academy Press: Washington.

Comstock GW. 1979. Water hardness and cardiovascular diseases: American Journal of Epidemiology; 110(4): 375-400.

Cote TR, Stroup DF, Dwyer DM, Horan JM, Peterson DE. 1993. Chronic obstructive pulmonary disease. A role for altitude. Chest; 103(4): 1194-7.

Coultas DB, Samet JM, Wiggins CL. 1984. Altitude and mortality from chronic obstructive lung disease in New Mexico. Archives of Environmental Health; 39(5): 355-9.

Danet S, Richard F, Montaye M, Beauchant S, Lemaire B, Graux C, Cottel D, Marecaux N, Amouyel P. 1999. Unhealthy effects of atmospheric temperature and pressure on the occurence of myocardial infarction and coronary deaths. Circulation; 100(1): E1-7.

Decarli A, La Vecchia C. 1986. Environmental factors and cancer mortality in Italy: correlational exercise. Oncology; 43(2): 116-26.

Delfino RJ, Murphy-Moulton AM, Burnett RT, Brook JR, Becklake MR. 1997. Effects of air pollution on emergency room visits for respiratory illnesses in Montreal, Quebec. American Journal of Respiratory and Critical Care Medicin; 155: 568-576.

DerSimonian R, Laird N. 1986. Meta-analysis in Clinical Trials. Control Clinical Trials 7:177–188.

Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG, Speizer FE. 1993. An association between air pollution and mortality in six US cities. New England Journal of Medicine; 329:1753–1759.

Driessen G, Gunther N, Van Os J. 1998. Shared social environment and psychiatric

disorder: a multilevel analysis of individual. Social Psychiatry and Psychiatric Epidemiology; 33(12): 606-12.

Duncan N (ed). 1996. Body Space. Routledge: New York.

Easton DF, Peto J, Babiker AGA. 1991. Floating absolute risk: An alternative to relative risk in survival and case-control analysis avoiding an arbitrary reference group. Statistics and Medicine; 10: 1025–1035.

Ebdon D. 1990. Statistics in Geography, second edition. Basil Blackwell: Cambridge.

Elo IT, Preston SH. 1996. Educational differentials in mortality: United States, 1979-85. Social Science & Medicine; 42(1): 47-57.

Elsom DM. 1992. Atmospheric Pollution. A Global Problem. Blackwell: Cambrdige, UK.

English D, Geographical epidemiology and ecological studies. In P Elliott, J Cuzick, D English, R Stern (eds.) Geographical & Environmental Epidemiology Methods for Small-Area Studies (Oxford University Press) pp. 4-13.

Evans RG, and Stoddart GL. 1994. Producing Health, Consuming Health Care, in Evans RG, Barer ML, and Marmor TR, Why are Some People Healthy and Others Not? (Walter de Gruyter Inc.: Hawthorne, New York) pp. 27-64.

Evans RG, Stoddart GL. 1990. Producing health, consuming health care. Social Science and Medicine; 31:1347–1363.

Fabsitz R, Feinleib M. 1980. Geographic patterns in county mortality rates from cardiovascular diseases. American Journal of Epidemiology; 111(3): 315-28.

Feldman JJ, Makuc DM, Kleinman JC, Cornoni-Huntley J. 1989. National trends in educational differentials in mortality. American Journal of Epidemiology; 129(5): 919-33.

Ferrie, J.E. 1995. Health effects of anticipation of job change and non-employment: Longitudinal data from the Whitehall study. British Medical Journal; 311:1264-9.

Fiscella K and Franks P. 1997. Poverty of income inequality as predictor of mortality: longitudinal cohort study. British Medical Journal; 314(7096): 1724-7.

Garrison RJ, Gold RS, Wilson PW, Kannel WB. 1993. Educational attainment and coronary heart disease risk: the Framingham Offspring Study. Preventive Medicine; 22(1):54-64.

Geronimus AT. 1986. The effects of race, residence and prenatal care on the relationship of maternal age neonatal mortality. American Journal of Public Health. 76(12):1416-21.

Geronimus AT, Bound J, Waidmann TA. 1999. Poverty, time, and place: variation in excess mortality across selected US populations, 1980-1990. Journal of Epidemiology & Community Health; 53(6): 325-34.

Getis A, and Ord K. 1992. The analysis of spatial association by use of distance statistics.

Geographical Analysis 24: 189-206.

Getis A. 1995. Spatial filtering in a regression framework: examples using data on urban crime, regional inequality, and government expenditures. In New Directions in Spatial Econometrics Anselin, L. and R.J.G.M. Florax ed. (Berline: Springer-Verlag).

Getis A, Ord C. 1996. Local spatial statistics: An overview. In Longley P, Batty M, (eds) Spatial Analysis: Modeling in a GIS Environment (GeoInformation International, Cambridge MA).

Gorjanc ML, Flanders WD, VanDerslice J, Hersh J, Malilay J. 1999. Effects of temperature and snowfall on mortality in Pennsylvania. American Journal of Epidemiology; 149(12): 1152-60.

Greeland, S., and Robbins, J. 1994. Invited commentary: Ecologic Studies - Biases, Misconceptions, and Counterexamples. American Journal of Epidemiology; 139:747-760.

Greenberg JH, Bromberg J, Reed CM, Gustafson TL, Beauchamp RA. 1983. The epidemiology of heat-related deaths, Texas–1950, 1970-79, and 1980. American Journal of Public Health; 73(7): 805-7.

Guest AM, Almgren G, Hussey JM. 1998. The ecology of race and socioeconomic distress: infant and working-age mortality in Chicago. Demography; 35(1): 23-34.

Haan M, Kaplan GA, Camacho T. 1987. Prospective evidence from the Alameda County Study. American Journal of Epidemiology; 125(6): 989–98.

Hahn RA, Eaker ED, Barker ND, Teutsch SM, Sosniak WA, Krieger N. 1996. Poverty and death in the United States. International Journal of Health Services; 26(4): 673-90.

Hart KD, Kunitz SJ, Sell RR, Mukamel DB. 1998. Metropolitan governance, residential segregation, and mortality among African Americans. American Journal of Public Health; 88(3): 434-8.

He XZ, Chen W, Liu ZY, Chapman RS. 1991. An epidemiological study of lung cancer in Xuan Wei County, China. Environmental Health Perspectives; 94:9-13.

Health Effects Institute. 2000. Re-analysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality: A Special Report of the Institute's Particle Epidemiology Reanalysis Project. Health Effects Institute, Cambridge MA.

Hennekens CH, Buring SL. 1988. Epidemiology in Medicine. Little, Brown and Company: Boston.

Hoek G and Brunekreef B. 1993. Acute effects of winter air pollution episode on pulmonary function and respiratory symptoms of children. Archives of Environmental Health; 4:328-335.

Higgins IT. 1977. Epidemiology of lung cancer in the United States. IARC Scientific

Publications; 16:191-203.

Jarvelin MR, Elliott P, Kleinschmidt I, Martuzzi M, Grundy C, Hartikainen AL, Rantakallio P. 1997. Ecological and individual predictors of birthweight in a northern Finland birth cohort 1986. Paediatric & Perinatal Epidemiology; 11(3): 298-312.

Kahn HA and Sempos CT. 1989. Statistical Methods in Epidemiology. Oxford University Press.

Kaplan GA, Pamuk ER, Lynch JW, Cohen RD, Balfour JL. 1996. Inequality in income and mortality in the United States: analysis of mortality and potential pathways. British Medical Journal; 312(7037): 999-1003.

Kaplan GA. 1996. People and places: contrasting perspectives on the association between social class and health. International Journal of Health Services; 26(3): 507-19.

Katsouyanni K, Pantazopoulou A, Touloumi G, Tselepidaki I, Moustris K, Asimakopoulos D, Poulopoulou G, Trichopoulos D. 1993. Evidence for interaction between air pollution and high temperature in the causation of excess mortality. Archives of Environmental Health; 48(4): 235-42.

Kawachi I, and Kennedy BP. 1997. The Relationship of Income Inequality to mortality: Does the Choice of Indicator Matter? Social Science and Medicine; 45(7): 1121-1127.

Kawachi I, Kennedy BP, Wilkinson RG. 1999. Crime: social disorganization and relative deprivation. Social Science & Medicine; 48(6):719-31.

Kee F, Wilson RH, Harper C, Patterson CC, McCallion K, Houston RF, Moorehead RJ, Sloan JM, Rowlands BJ. 1999. Influence of hospital and clinician workload on survival from colorectal cancer: cohort study. British Medical Journal; 319(7195): 576-7.

Keith M, Pile S (eds). 1993. Place and the Politics of Identity. Routledge: New York.

Kennedy BP, Kawachi I, Lochner K, Jones C, Prothrow-Stith D. 1997. (Dis)respect and black mortality. Ethnicity & Disease; 7(3): 207-14.

Kennedy BP, Kawachi I, Prothrow-Stith D. 1996. Income distribution and mortality: cross sectional ecological study of the Robin Hood index in the United States. British Medical Journal; 312(7037): 1004-7.

Kinney PL, Ozkaynak H. 1991. Associations of daily mortality and air pollution in Los Angeles County. Environmental Research; 54(2): 99-120.

Kleinbaum D G, Kupper L L, Muller K E, Nizam A. 1998. Applied Regression Analyis and Multivariable Methods. Duxbury Press: Pacific Grove, CA.

Kreindel S, Rosetti R, Goldberg R, Savageau J, Yarzebski J, Gore J, Russo A, Bigelow C. Health insurance coverage and outcome following acute myocardial infarction. A community-wide perspective. Archives of Internal Medicine; 157(7): 758-62.

Larsen U. 1990. The effects of monthly temperature fluctuations on mortality in the

United States from 1921 to 1985. International Journal of Biometeorology; 34(3): 136-45.

Lawther PJ, Waller RE. 1976. Coal fires, indicuatrial emissions and motor vehicles as sources of environmental carcinogens. IARC Scientific Publications; 13:27-40.

Lee JA. 1997. Declining effect of latitude on melanoma mortality rates in the United States. A preliminary study. American Journal of Epidemiology; 146(5): 413-7.

Lee JA, Scotto J. 1993. Melanoma: linked temporal and latitude changes in the United States. Cancer Causes & Control; 4(5): 413-8.

Lipfert FW. 1993. Community air pollution mortality: analysis of 1980 data from US metropolitan areas I. Particulate air pollution, BNL 48446-R, prepared for US Department of Energy, Brookhaven National Laboratory, August.

Loomis, D., Castillejos, M., Gold, DR., McDonnell, W., Borja-Aburto, VH. 1999. Air Pollution and Infant Mortality in Mexico City. Epidemiology; 10(2): 118-23.

Mannino JA, Washburn RA. 1989. Environmental temperature and mortality from acute myocardial infarction. International Journal of Biometeorology; 33(1): 32-5.

Mansfield CJ, Wilson JL, Kobrinski EJ, Mitchell J. 1999. Premature mortality in the United States: the roles of geographic area, socioeconomic status, household type, and availability of medical care. American Journal of Public Health; 89(6): 893-8.

Marmor M. 1975. Heat wave mortality in New York City, 1949 to 1970. Archives of Environmental Health; 30(3): 130-6.

Marmot, M and Wilkinson, RG. 1999. Social Determinants of Health. Oxford University Press. Oxford, England.

Meade M, Florin J, Gesler W. 1988. Medical Geography. Guilford Press: Oxford.

Menotti A. Blackburn H. Seccareccia F. Kromhout D. Nissinen A. Karyonen M. Fidanza F. Giampaoli S. Buzina R. Mohacek I. Nedeljkovic S. Aravanis C. Dontas A. 1998. Relationship of some risk factors with typical and atypical manifestations of coronary heart disease. Cardiology. 89(1):59-67.

Menotti A, Blackburn H, Kromhout D, Nissinen A, Karkoven M, Aravanis C. 1997. The inverse relation of average population blood pressure and stroke mortality rates in the seven countries study: a paradox. European Journal of Epidemiology; 13(4): 379-386.

Miron J. 1984. Spatial autocorrelation in regression analysis: A beginner's guide. In Gaile GL, Willmott CJ, (eds) Spatial Statistics and Models (D Reidel Publishing, Boston) pp 201–222.

Moore LG, Rohr AL, Maisenbach JK, Reeves JT. 1982. Emphysema mortality is increased in Colorado residents at high altitude. American Review of Respiratory Disease; 126(2):225-8.

Morgenstern H. 1998. Ecologic Studies, in KJ Rothman and S Greenland (eds.) Modern Epidemiology, Second Edition, (Lippincott-Raven Publishers: Philadelphia).

Morris RD, Munasinghe RL. 1994. Geographic variability in hospital admission rates for respiratory disease among the elderly in the United States. Chest; 106(4): 1172-81.

Mortimer, EA Jr., Monson, RR., MacMahon, B. 1977. Reduction in mortality from coronary heart disease in men residing at high altitude. New England Journal of Medicine. 11:581-5.

Mortimer EA Jr, Monson RR, MacMahon B. 1977. Reduction in mortality from coronary heart disease in men residing at high altitude. New England Journal of Medicine; 296(11): 581-5.

Nakanishi Y, Chen S, Inutsuka S, Ma Y, Jiang X, Hara N, Sera N, Tokiwa H. 1997. Possible role of indoor environment and coal combustion emission in lung carcinogenesis in Fuyuan County, China. Neoplasma; 44(1):69-72.

National Research Council, United States. 1999. Health Effects of Exposure to Radon, BEIR VI, National Academy Press: Washington DC.

National Research Council. 1998. Research Priorities for Airborne Particulate Matter I: Immediate Priorities and a Long-Term Research Portfolio. Washington, D.C.: National Academy Press.

National Research Council. 1999. Research Priorities for Airborne Particulate Matter II: Evaluating Research Progress and Updating the Portfolio. Washington, D.C.: National Academy Press.

Navidi W. Thomas D, Stram D, Peters J. 1994. Design and analysis of multilevel analytic studies with applications to a study of air pollution. Environmental Health Perspectives; 102(sup 8): 25-32.

Neri LC, Johansen HL. 1978. Water hardness and cardiovascular mortality. Annals of the New York Academy of Sciences; 304: 203-21.

Odland J. 1988. Spatial Autocorrelation. Sage Publications: Beverly Hills CA.

Özkaynak H, Thurston G. 1987. Associations between 1980 US mortality rates and alternative measures of airborne particle concentration. Risk Analysis; 7:449–460.

Parr H, Philo C. 1995. Mapping 'Mad' Identities in Pile S, Thrift N (eds) Mapping the Subject. Routledge: New York.

Pearce N. 1996. Traditional Epidemiology, Modern Epidemiology and Public Health. American Journal of Public Health; 86(5): 678-683.

Pile S, Thrift N (eds). 1995. Mapping the Subject. Routledge: New York.

Peter, I., Otremski, I., Livshits, G. 1996. Geographic variation in vascular mortality in Eurasia: spatial autocorrelation analysis of mortality variables and risk factors. Annals of Human Biology;. 6:471-90.

Pile S. 1996. The Body and the City: Psychoanalysis, Space and Subjectivity. Routledge: New York.

Polednak AP. 1998. Mortality in Hartford, Connecticut: a comparison with the South Bronks, New York. Journal of Urban Health; 75(3): 550-7.

Polednak AP. 1991. Black-white differences in infant mortality in 38 standard metropolitan statistical areas. American Journal of Public Health; 81(11): 1480-2.

Pope CA 3rd, Hill RW, and Villegas GM. 1999. Particulate air pollution and daily mortality on Utah's Wasatch Front. Environmental Health Perspectives; 107(7): 567-73.

Pope CA 3rd, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, Heath CW. 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. Am J Respir Critical Care Medicine; 151:669–674.

Ramlow, JM., Kuller, LH. 1990. Effects of the summer heat wave of 1988 on daily mortality in Allegheny County, PA. Public Health Reports; 3:283-9.

Roemer W, Hoek G, and Brunekreef B. 1993. Effect of ambient winter air pollution on respiratory health of children with chronic respiratory symptoms. American Review of Respiratory Disease; 147: 118-124.

Rogot E, Sorlie PD, Backlund E. 1992. Air-conditioning and mortality in hot weather. American Journal of Epidemiology; 136(1): 106-16.

Rogot E, Padgett SJ. 1976. Associations of coronary and stroke mortality with temperature and snowfall in selected areas of the United States, 1962-1966. American Journal of Epidemiology; 103(6): 565-75.

Samet JM, Zeger SL, Dominici F, Curriero F, Coursac I, Dockery DW, Schwarz J, Zanobetti A. 2000. The National Morbidity and Mortality Air Pollution Study, Part II: Morbidity and Mortality from Air Pollution in United States. Research Report 94. Health Effects Institute: Cambridge MA.

Sartor F, Snacken R, Demuth C, Walckiers D. 1995. Temperature, ambient ozone levels, and mortality during summer 1994, in Belgium. Environmental Research; 70(2): 105-13.

Sawada M. 1999. Global spatial autocorrelation indices– Moran's I, Geary's C and the general cross-product statistic. University of Ottawa website www.uottawa.ca/academic/arts/geographie/lpcweb/sections1/reports/moransi/moran.htm.

Schwartz J. 1994a. Air pollution and daily mortality: a review and meta analysis. Environmental Research; 64: 36-52.

Schwartz J. 1991. Particulate air pollution and daily mortality: a synthesis. Public Health Review; 19:36-52.

Schwartz S. 1994b. The Fallacy of the Ecological Fallacy: The Potential Misuse of a Concept and the Consequences. American Journal of Public Health; 84(5): 819-824.

Seretakis D, Lagiou P, Lipworth L, Signorello LB, Rothman KJ, Trichopoulos D. 1997. Changing seasonality of mortality from coronary heart disease. Journal of the American Medical Association; 278(12): 1012-4.

Sharrett AR. 1981. Water hardness and cardiovascular disease. Circulation; 63(1):247A-250A.

Shryock HS. 1976. The Methods and Materials of Demography. New York: Academic Press.

Sibley D. 1995. Geographies of Exclusion. Routledge: New York.

Sly RM, O'Donnell R. 1989. Lack of effect of geographic elevation on mortality from asthma. Annals of Allergy; 63(6 pt 1): 495-7.

Smoyer KE. 1998. A comparative analysis of heat waves and associated mortality in St. Louis, Missouri - 1980 and 1995. International Journal of Biometeorology; 42(1): 44-50.

Sorlie PD, Rogot E. 1990. Mortality by employment status in the National Longitudinal Mortality Study. 132(5): 983-92.

Stengard JH. Kardia SL. Tervahauta M. Ehnholm C. Nissinen A. Sing CF. 1999. Utility of the predictors of coronary heart disease mortality in a longitudinal study of elderly Finnish men aged 65 to 84 years is dependent on context defined by Apo E genotype and area of residence. Clinical Genetics; 56(5):367-77.

Susser M and E. 1996a. Choosing a Future for Epidemiology: I. Eras and Paradigms. American Journal of Public Health; 86(5): 668-673.

Susser M and E. 1996b. Choosing a Future for Epidemiology: II. From Black Boxes to Chinese Boxes and Eco-Epidemiology., American Journal of Public Health; 86(5): 674-677.

Susser M. 1994a. The Logic in Ecological: I. The Logic of Analysis. American Journal of Public Health; 84(5): 825-829.

Susser M. 1994b. The Logic in Ecological: II. The Logic of Design. American Journal of Public Health; 84(5): 830-835.

Tselepidaki IG, Asimakopoulos DN, Katsouyanni K, Moustris C, Touloumi G, Pantazopoulou A. 1995. The use of a complex thermohygrometric index in predicting adverse health effects in Athens. International Journal of Biometeorology; 38(4): 194-8.

US Bureau of the Census. 1988. County and City Data Book. US Government Printing Office: Washington DC.

Waddell CE. 1983. Population crowding and death rates due to heart disease. Ecology of Disease;. 4:271-5.

Waitzman, NJ.1998. Smith KR. Separate but lethal: the effects of economic segregation on mortality in metropolitan America. Milbank Quarterly; 3:341-73. 304.

Wilkinson, RG. 1996. Unhealthy Societies: The Afflictions of Inequality. Routledge; London.

Willis A. 1997. The Role of Space and Place in Enabling Resistance at a Global Scale: The Case of Pearson College. M.A. Thesis. Carleton University: Ottawa.

Wissow LS, Gittelsohn AM, Szklo M, Starfield B, Mussman M. 1988. Poverty, race, and hospitalization for childhood asthma. American Journal of Public Health; 78(7): 777-82.

Woodruff, TJ., Grillo, J., Schoendorf, KC. 1997. The relationship between selected causes of postneonatal infant mortality and particulate air pollution in the United States. Environmental Health Perspectives; 105(6): 608-12.

World Health Organization. 1975. International Classification of Diseases. 9th Version. Vol. 1. Geneva: WHO.

Xu Z, Brown LM, Pan GW, Liu TF, Gao GS, Stone BJ, Cao RM, Guan DX, Sheng JH, Yan ZS, Dosemeci M, Fraumeni JF Jr, Blot WJ. 1996. Cancer risks among iron and steel workers in Anshan, China, Part II. American Journal of Industrial Medicine; 30(1):7-15.

Yen IH, Kaplan GA, 1999. Poverty area residence and changes in depression and perceived health status: evidence from the Alameda County Study. International Journal of Epidemiology; 28(1): 90-4.

Yen IH, Kaplan GA. 1998. Poverty area residence and changes in physical activity level: evidence from the Alameda County Study. American Journal of Public Health; 88(11): 1709-12.

Zahm SH. 1999. Sunlight and non-Hodgkin's lymphoma: a population-based cohort study in Sweden. International Journal of Cancer; 80(5):641-5.

Appendix A - Data Used in the Thesis

Table A.1 Metropolitan Area-specific Demographic and Race Data

MA Code	Metropolitan Area	State	Total Population (1980)	Percent Population Change ('80-'86)	Percent White	Percent African American
1000	Birmingham	AL	883993	3.1	72.06	27.17
3440	Huntsville	AL	196966	18.6	79.20	19.8
5160	Mobile	AL	443536	6.0	70.34	28.64
4400	Little Rock-North Little Rock	AR	474464	6.6	79.17	19.12
6200	Phoenix	AZ	1509227	25.9	93.84	3.17
8520	Tucson	AZ	602400	13.4	92.67	2.7
360	Anaheim	CA	2166800	12.1	90.10	1.27
2840	Fresno	CA	587600	14.2	89.24	5.00
4480	Los Angeles-Long Beach	CA	8295900	10.9	78.88	12.61
6780	Riverside-San Bernardino	CA	1558215	28.4	90.60	5.06
6920	Sacramento	CA	1099814	17.4	87.38	5.60
7320	San Diego	CA	2201300	18.2	86.31	5.61
7360	San Francisco	CA	1488895	6.7	73.45	8.56
7400	San Jose	CA	1401600	8.2	83.94	3.31
1720	Colorado Springs	CO	380400	22.9	90.41	6.18
2080	Denver	CO	1428836	14.3	91.96	5.30
2670	Fort Collins-Loveland	CO	174600	17.0	97.68	0.39
3060	Greeley	CO	123438	9.4	98.00	0.49
6560	Pueblo	CO	125972	0.9	97.18	1.93
1160	Bridgeport-Milford	CT	383836		75.36	12.65
3280	Hartford	CT	527006		85.13	12.09
5480	New Haven-Meriden	CT	421782		85.92	13.35
8840	Washington	DC	2655167	5.3	63.27	26.78
9160	Wilmington	DE	523221	5.3	84.03	13.99
2680	Fort Lauderdale-Hollywood-ompano Beach	FL	2643766	10.1	81.96	14.93
3600	Jacksonville	FL	722252	18.1	77.25	21.57
5960	Orlando	FL	699906	28.4	85.98	12.92
8280	Tampa-St. Petersburg-learwater	FL	1964296	19.1	90.55	8.75
520	Atlanta	GA	2138143	19.7	74.91	24.58
1800	Columbus	GA	239196	4.8	61.69	34.99
1360	Cedar Rapids	IA	251492	1.0	97.21	1.65

MA Code	Metropolitan Area	State	Total Population (1980)	Percent Population Change ('80-'86)	Percent White	Percent African American
2120	Des Moines	IA	367561	3.8	93.93	3.81
2200	Dubuque	IA	93745	-2.9	99.20	0.33
8920	Waterloo-Cedar Falls	IA	162781	-6.9	93.41	5.38
1080	Boise City	ID	173125	12.0	97.56	0.33
1600	Chicago	IL	7171437	2.9	76.91	19.90
1020	Bloomington	IN	98785	2.9	95.76	2.67
2440	Evansville-Henderson	IN	276252	1.8	93.67	5.63
2960	Gary-Hammond	IN	642733	-4.3	78.69	19.66
3480	Indianapolis	IN	1305911	3.0	85.93	12.79
7800	South Bend-Mishawaka	IN	378947	2.3	91.53	7.26
8320	Terre Haute	IN	137247	-2.2	94.72	4.50
8440	Topeka	KS	154916	3.8	89.92	7.66
9040	Wichita	KS	411870	6.6	88.47	7.78
4280	Lexington-Fayette	KY	317548	4.3	77.31	10.91
760	Baton Rouge	LA	494151	10.4	70.83	27.85
5560	New Orleans	LA	1256668	6.2	64.65	34.51
7680	Shreveport	LA	333158	9.4	65.83	33.15
1120	Boston	MA	2428005		90.53	6.55
5400	New Bedford	MA	108007		90.61	2.54
8000	Springfield	MA	400451		88.59	6.93
9240	Worcester	MA	261069		95.97	2.06
720	Baltimore	MD	1543882	4.3	66.67	25.44
730	Bangor	ME	82561		98.74	0.24
4240	Lewiston-Auburn	ME	78075		99.23	0.36
6400	Portland	ME	94436		99.13	0.27
2162	Detroit-Ann Arbor	MI	4752764	-3.2	78.61	19.35
2640	Flint	MI	450499	-3.5	80.14	17.47
4040	Lansing-East Lansing	MI	419750	1.2	92.23	5.55
6960	Saginaw-Bay City-Midland	MI	421518	-4.3	89.29	8.85
2240	Duluth	MN	266650	-8.7	97.65	0.34
5120	Minneapolis-St Paul	MN	2137133	7.4	94.68	2.31
3760	Kansas City	MO	1433464	5.9	85.98	12.52
7040	St. Louis	MO	2376971	2.6	81.80	17.13

MA Code	Metropolitan Area	State	Total Population (1980)	Percent Population Change ('80-'86)	Percent White	Percent African American
3560	Jackson	MS	362038	8.3	57.33	46.59
880	Billings	MT	108035	11.2	96.75	0.19
3040	Great Falls	MT	80696	-1.6	94.27	1.19
1520	Charlotte-Gastonia-Rock Hill	NC	885552	9.9	77.87	20.00
3120	Greensboro-Winston-Salem-gh Point	NC	950763	5.4	79.96	19.01
6640	Raleigh-Durham	NC	560775	16.0	72.84	26.15
1010	Bismarck	ND	79988	7.4	97.36	0.08
2520	Fargo-Moorhead	ND	137574	5.6	98.45	0.16
2985	Grand Forks	ND	66100	5.0	95.62	1.62
4360	Lincoln	NE	192884	6.8	96.03	1.70
5920	Omaha	NE	585122	5.0	90.43	7.51
4760	Manchester	NH	100706	7.0	98.85	1.18
6450	Portsmouth-Dover-Rochester	NH	159749	.	98.16	0.94
3640	Jersey City	NJ	556972	-0.7	82.31	12.52
5640	Newark	NJ	1879147	0.5	74.99	21.75
875	Paterson	NJ	1292970	0.4	89.27	7.13
8480	Trenton	NJ	307863	4.2	78.86	17.98
200	Albuquerque	NM	420262	12.9	92.74	2.14
4120	Las Vegas	NV	463087	23.0	86.08	9.99
6720	Reno	NV	193623	16.0	93.33	1.92
160	Albany-Schenectady-Troy	NY	835880	0.9	95.29	3.65
1280	Buffalo	NY	1015472	-5.0	88.04	10.20
2335	Elmira	NY	97656	-7.3	95.23	4.05
5380	Nassau-Suffolk	NY	2605813	1.1	91.56	6.22
5600	New York	NY	8274961	2.4	70.30	23.09
6460	Poughkeepsie	NY	245055	4.8	90.85	6.94
6840	Rochester	NY	971230	0.9	90.30	8.02
8160	Syracuse	NY	642971	1.0	93.49	4.82
8680	Utica-Rome	NY	320180	-1.5	96.88	2.39
80	Akron	OH	660328	-2.4	89.79	9.14
1320	Canton	OH	404421	-1.0	92.98	6.03
1640	Cincinnati	OH	1401471	1.2	86.50	2.03
1680	Cleveland	OH	2173734	-2.4	81.43	16.80

MA Code	Metropolitan Area	State	Total Population (1980)	Percent Population Change ('80-'86)	Percent White	Percent African American
1840	Columbus	OH	1243827	4.5	87.58	11.36
2000	Dayton-Springfield	OH	942083	-0.9	86.03	12.59
3200	Hamilton-Middletown	OH	258787	4.9	94.34	4.76
4800	Mansfield	OH	131205	-1.8	92.26	7.10
8080	Steubenville-Weirton	OH	163734	-5.5	95.51	3.89
8400	Toledo	OH	616864	-0.9	88.00	10.62
9320	Youngstown-Warren	OH	531350	-4.0	88.35	10.49
5880	Oklahoma City	OK	860969	14.2	86.11	9.13
6440	Portland	OR	1105750	4.3	92.64	2.91
240	Allentown-Bethlehem	PA	635481	3.4	97.55	1.41
280	Altoona	PA	136621	-3.0	98.77	0.76
2360	Erie	PA	279780	-0.2	94.64	4.34
3240	Harrisburg-Lebanon-Carlisle	PA	556242	3.8	92.55	6.10
3680	Johnstown	PA	264506	-3.9	98.44	1.35
4000	Lancaster	PA	362346	8.6	96.59	1.92
6160	Philadelphia	PA	4716559	2.3	79.16	18.73
6280	Pittsburgh	PA	2423311	-4.4	91.51	7.49
6680	Reading	PA	312509	2.7	96.46	2.46
42069	Scranton (Lackawanna)	PA	227908	-2.2	98.87	0.47
7610	Sharon	PA	128299	-3.7	95.25	4.30
8050	State College	PA	112760	1.6	97.07	1.51
9140	Williamsport	PA	118416	-1.8	98.07	1.44
9280	York	PA	381255	4.3	96.50	2.64
6480	Providence	RI	487781	.	92.84	4.83
1440	Charleston	SC	430346	12.8	68.36	31.05
1760	Columbia	SC	409955	8.5	69.73	28.68
3160	Greenville-Spartanburg	SC	747100	6.2	82.11	17.06
1560	Chatanooga	TN	426540	-0.2	84.42	14.01
3840	Knoxville	TN	565970	4.5	92.94	6.04
4920	Memphis	TN	913472	5.0	57.95	39.84
5360	Nashville	TN	933847	9.4	83.25	16.24
840	Beaumont-Port Arthur	TX	373210	0.7	75.89	21.77
1880	Corpus Christi	TX	326228	11.4	95.17	3.94

MA Code	Metropolitan Area	State	Total Population (1980)	Percent Population Change ('80-'86)	Percent White	Percent African American
1920	Dallas	TX	2877238	23.4	83.83	14.48
2320	El Paso	TX	479899	17.0	94.82	3.83
2920	Galveston-Texas City	TX	365325	10.4	85.25	13.56
3360	Houston	TX	2734595	18.1	77.89	18.78
4600	Lubbock	TX	211651	6.2	91.34	7.25
7200	San Angelo	TX	84784	15.8	95.00	4.02
7240	San Antonio	TX	1072125	19.1	91.83	6.75
8800	Waco	TX	170755	9.9	83.13	15.95
9080	Wichita Falls	TX	121082	5.0	87.99	8.94
7160	Salt Lake City-Ogden	UT	910222	14.4	96.07	0.98
1950	Danville	VA	283677	0.3	72.80	26.26
5720	Norfolk-Virginia Beach-ewport News	VA	1160347	12.8	69.36	28.11
6760	Richmond-Petersburg	VA	787044	6.3	64.45	29.09
6800	Roanoke	VA	220393	2.0	87.30	11.73
7600	Seattle	WA	1607618	8.9	89.46	3.60
7840	Spokane	WA	341835	4.4	95.63	1.29
8200	Tacoma	WA	632819	11.0	88.81	5.19
2290	Eau Claire	WI	130932	4.4	98.54	0.14
3800	Kenosha	WI	123137	-2.5	96.36	2.54
4720	Madison	WI	462965	4.3	95.87	2.24
5080	Milwaukee	WI	1497955	-1.0	87.86	10.09
6600	Racine	WI	173132	-0.5	90.65	7.99
1480	Charleston	WV	311341	-1.0	94.93	4.44
3400	Huntington-Ashland	WV	388888	-1.6	97.67	1.86

Table A.2 Metropolitan Area-specific Health Services and Socio-economic Data

MA Code	Metropolitan Area	State	Percent Finishing High School	Mean Income (1979) ($)	Poverty Rate Individuals (%)	Unemploy-ment Rate (%)	Gini Coeffcient
1000	Birmingham	AL	61.7	6824	15.3	8.2	0.426
3440	Huntsville	AL	70.1	7068	13.8	7.2	0.406
5160	Mobile	AL	61.4	6066	18.3	10.7	0.423
4400	Little Rock-North Little Rock	AR	68.2	6844	12.8	6.9	0.404
6200	Phoenix	AZ	75.0	7744	10.5	5.6	0.395
8520	Tucson	AZ	74.6	7181	13.0	5.7	0.409
360	Anaheim	CA	80.4	9612	7.3	4.0	0.380
2840	Fresno	CA	63.7	7010	14.5	12.3	0.420
4480	Los Angeles-Long Beach	CA	69.8	8365	13.4	6.7	0.430
6780	Riverside-San Bernardino	CA	70.1	7316	11.2	6.4	0.403
6920	Sacramento	CA	77.7	7958	11.2	6.2	0.401
7320	San Diego	CA	78.0	7992	11.3	5.0	0.410
7360	San Francisco	CA	79.2	10360	9.7	4.6	0.421
7400	San Jose	CA	79.5	9576	7.1	5.8	0.369
1720	Colorado Springs	CO	82.7	7061	10.3	7.3	0.391
2080	Denver	CO	80.5	8891	8.2	7.2	0.388
2670	Fort Collins-Loveland	CO	82.7	7489	11.0	6.6	0.389
3060	Greeley	CO	68.8	6542	14.1	8.4	0.400
6560	Pueblo	CO	66.5	6744	13.7	12.1	0.411
1160	Bridgeport-Milford	CT	62.3	7055	10.1	3.7	0.395
3280	Hartford	CT	70.6	8417	9.2	4.2	0.396
5480	New Haven-Meriden	CT	68.7	7055	10.3	6.1	0.395
8840	Washington	DC	79.2	10131	9.2	3.8	0.383
9160	Wilmington	DE	69.6	7785	10.9	5.0	0.392
2680	Fort Lauderdale-Hollywood-ompano Beach	FL	66.6	8100	12.7	5.8	0.435
3600	Jacksonville	FL	67.0	6828	15.1	5.4	0.415
5960	Orlando	FL	71.3	7126	15.0	4.7	0.425
8280	Tampa-St. Petersburg-learwater	FL	66.5	7230	11.4	5.0	0.414
520	Atlanta	GA	66.7	7746	12.5	4.6	0.408
1800	Columbus	GA	58.3	5734	18.8	7.6	0.430

MA Code	Metropolitan Area	State	Percent Finishing High School	Mean Income (1979) ($)	Poverty Rate Individuals (%)	Unemploy-ment Rate (%)	Gini Coeffcient
1360	Cedar Rapids	IA	80.0	7977	9.4	4.8	0.382
2120	Des Moines	IA	78.5	8175	8.1	5.7	0.380
2200	Dubuque	IA	68.6	7155	8.3	7.7	0.384
8920	Waterloo-Cedar Falls	IA	73.5	7684	9.0	12.2	0.377
1080	Boise City	ID	81.7	7805	8.5	5.8	0.389
1600	Chicago	IL	67.6	8614	11.2	7.2	0.401
1020	Bloomington	IN	74.7	6322	15.0	3.9	0.430
2440	Evansville-Henderson	IN	64.2	7426	9.7	6.8	0.398
2960	Gary-Hammond	IN	65.4	7897	9.9	11.7	0.370
3480	Indianapolis	IN	68.8	7765	9.4	5.2	0.387
7800	South Bend-Mishawaka	IN	66.9	7331	8.7	5.5	0.379
8320	Terre Haute	IN	67.7	6668	10.7	7.0	0.393
8440	Topeka	KS	78.3	7970	7.8	5.1	0.378
9040	Wichita	KS	76.2	8135	8.9	6.0	0.384
4280	Lexington-Fayette	KY	66.3	7000	14.1	5.3	0.415
760	Baton Rouge	LA	68.2	7160	14.9	11.0	0.419
5560	New Orleans	LA	62.7	6999	17.5	10.9	0.439
7680	Shreveport	LA	63.4	7071	15.8	12.0	0.432
1120	Boston	MA	77.3	7950	9.8	4.7	0.411
5400	New Bedford	MA	41.8	5752	15.1	8.9	0.415
8000	Springfield	MA	66.1	6701	12.5	6.1	0.407
9240	Worcester	MA	64.8	6648	11.1	5.2	0.399
720	Baltimore	MD	62.0	7747	5.7	5.6	0.396
730	Bangor	ME	76.0	6042	11.3	7.3	0.405
4240	Lewiston-Auburn	ME	56.5	5700	12.6	6.5	0.391
6400	Portland	ME	75.6	6937	7.7	5.3	0.392
2162	Detroit-Ann Arbor	MI	67.7	8513	10.1	7.9	0.394
2640	Flint	MI	67.8	7983	10.6	10.6	0.377
4040	Lansing-East Lansing	MI	77.3	7644	10.7	7.0	0.376
6960	Saginaw-Bay City-Midland	MI	67.3	7423	10.4	10.0	0.385
2240	Duluth	MN	71.0	6966	9.3	9.5	0.385
5120	Minneapolis-St Paul	MN	79.7	8679	6.8	4.2	0.380
3760	Kansas City	MO	73.6	8116	9.0	4.7	0.390

MA Code	Metropolitan Area	State	Percent Finishing High School	Mean Income (1979) ($)	Poverty Rate Individuals (%)	Unemploy-ment Rate (%)	Gini Coeffcient
7040	St. Louis	MO	64.1	7704	10.4	7.0	0.394
3560	Jackson	MS	67.9	6536	18.8	8.0	0.440
880	Billings	MT	76.6	7639	9.3	7.8	0.396
3040	Great Falls	MT	75.2	7006	10.3	2.7	0.391
1520	Charlotte-Gastonia-Rock Hill	NC	57.6	7005	10.6	4.6	0.408
3120	Greensboro-Winston-Salem-igh Point	NC	57.0	6980	11.0	4.6	0.397
6640	Raleigh-Durham	NC	69.0	7278	12.3	3.3	0.401
1010	Bismarck	ND	71.5	7344	8.8	7.1	0.340
2520	Fargo-Moorhead	ND	76.5	7302	9.8	7.8	0.337
2985	Grand Forks	ND	76.2	6561	11.0	3.8	0.357
4360	Lincoln	NE	81.5	7739	8.6	3.1	0.381
5920	Omaha	NE	76.6	7570	9.1	5.3	0.385
4760	Manchester	NH	61.9	6822	9.7	5.2	0.340
6450	Portsmouth-Dover-Rochester	NH	73.5	6626	8.8	5.0	0.369
3640	Jersey City	NJ	51.5	6504	16.9	8.0	0.419
5640	Newark	NJ	68.9	8514	11.2	5.3	0.414
875	Paterson	NJ	68.6	9200	7.1	4.6	0.397
8480	Trenton	NJ	67.8	8178	9.4	4.3	0.398
200	Albuquerque	NM	76.5	7154	13.2	6.5	0.408
4120	Las Vegas	NV	74.0	8284	9.1	6.2	0.393
6720	Reno	NV	80.1	9524	7.0	5.1	0.387
160	Albany-Schenectady-Troy	NY	69.7	7094	9.8	5.2	0.386
1280	Buffalo	NY	65.4	7147	10.6	7.2	0.394
2335	Elmira	NY	69.5	6393	11.0	6.8	0.382
5380	Nassau-Suffolk	NY	75.8	8866	5.7	4.4	0.381
5600	New York	NY	62.4	7739	18.1	6.7	0.455
6460	Poughkeepsie	NY	70.6	7834	7.3	3.8	0.372
6840	Rochester	NY	70.4	7911	8.8	5.7	0.380
8160	Syracuse	NY	70.7	6942	10.4	7.4	0.385
8680	Utica-Rome	NY	64.5	6140	11.4	7.1	0.384
80	Akron	OH	69.5	7549	9.4	8.0	0.384
1320	Canton	OH	67.1	7273	8.6	10.2	0.370

MA Code	Metropolitan Area	State	Percent Finishing High School	Mean Income (1979) ($)	Poverty Rate Individuals (%)	Unemploy-ment Rate (%)	Gini Coeffcient
1640	Cincinnati	OH	63.2	7542	10.3	6.5	0.398
1680	Cleveland	OH	68.2	8062	9.8	7.8	0.395
1840	Columbus	OH	71.8	7497	11.1	6.1	0.389
2000	Dayton-Springfield	OH	69.4	7492	10.3	6.6	0.386
3200	Hamilton-Middletown	OH	64.7	7384	9.8	7.7	0.373
4800	Mansfield	OH	66.1	6907	9.4	9.2	0.369
8080	Steubenville-Weirton	OH	62.3	7374	9.3	8.7	0.380
8400	Toledo	OH	68.2	7533	11.0	8.4	0.418
9320	Youngstown-Warren	OH	66.7	7259	9.7	10.6	0.382
5880	Oklahoma City	OK	73.0	7700	10.6	6.8	0.406
6440	Portland	OR	78.7	8454	9.0	7.2	0.396
240	Allentown-Bethlehem	PA	63.2	7565	7.4	7.4	0.372
280	Altoona	PA	65.4	6105	10.9	9.2	0.379
2360	Erie	PA	69.9	6692	10.0	8.5	0.376
3240	Harrisburg-Lebanon-Carlisle	PA	68.3	7478	8.0	4.7	0.365
3680	Johnstown	PA	60.1	6292	10.3	10.7	0.387
4000	Lancaster	PA	59.6	7151	8.5	3.4	0.364
6160	Philadelphia	PA	66.0	7522	11.9	5.2	0.406
6280	Pittsburgh	PA	67.3	7624	9.2	8.3	0.395
6680	Reading	PA	58.5	7404	8.2	6.4	0.374
42069	Scranton (Lackawanna)	PA	63.8	6162	9.8	7.6	0.391
7610	Sharon	PA	66.8	6811	9.1	8.4	0.374
8050	State College	PA	75.8	5967	16.7	6.0	0.408
9140	Williamsport	PA	65.6	6217	10.8	7.1	0.376
9280	York	PA	61.3	7189	7.2	5.3	0.363
6480	Providence	RI	61.4	6985	11.1	7.3	0.384
1440	Charleston	SC	63.3	6144	16.5	4.6	0.405
1760	Columbia	SC	66.3	6729	13.3	3.8	0.394
3160	Greenville-Spartanburg	SC	51.9	6461	12.4	5.7	0.399
1560	Chatanooga	TN	58.3	6481	13.7	7.1	0.410
3840	Knoxville	TN	59.8	6584	15.0	7.8	0.420
4920	Memphis	TN	63.8	6540	20.0	6.8	0.434
5360	Nashville	TN	63.7	7158	11.7	5.2	0.404

MA Code	Metropolitan Area	State	Percent Finishing High School	Mean Income (1979) ($)	Poverty Rate Individuals (%)	Unemploy-ment Rate (%)	Gini Coeffcient
840	Beaumont-Port Arthur	TX	62.5	7460	12.4	14.5	0.396
1880	Corpus Christi	TX	57.7	6548	17.0	12.2	0.417
1920	Dallas	TX	69.6	8304	10.0	6.5	0.401
2320	El Paso	TX	59.5	5322	21.7	11.5	0.421
2920	Galveston-Texas City	TX	65.2	8099	9.5	11.2	0.373
3360	Houston	TX	69.9	8987	10.3	10.3	0.394
4600	Lubbock	TX	66.4	6970	14.3	6.8	0.416
7200	San Angelo	TX	59.8	6886	12.5	7	0.425
7240	San Antonio	TX	62.7	6200	18.1	7.2	0.417
8800	Waco	TX	58.5	6298	17.2	7.8	0.431
9080	Wichita Falls	TX	65.4	7148	12.3	8.3	0.415
7160	Salt Lake City-Ogden	UT	80.7	6853	8.4	5.3	0.373
1950	Danville	VA	46.7	6181	13.4	7.8	0.400
5720	Norfolk-Virginia Beach-ewport News	VA	66.2	6802	14.0	4.9	0.396
6760	Richmond-Petersburg	VA	62.3	7640	11.3	4.5	0.391
6800	Roanoke	VA	62.0	7380	11.0	4.8	0.401
7600	Seattle	WA	81.7	9368	7.7	6.5	0.385
7840	Spokane	WA	78.2	7244	11.5	8.2	0.401
8200	Tacoma	WA	77.1	7565	10.0	7.8	0.388
2290	Eau Claire	WI	70.3	6264	11.8	6.6	0.389
3800	Kenosha	WI	64.8	7818	7.0	12.4	0.361
4720	Madison	WI	79.9	7934	8.9	5.6	0.381
5080	Milwaukee	WI	71.4	8279	7.9	6.1	0.378
6600	Racine	WI	67.7	8063	7.0	8.5	0.363
1480	Charleston	WV	63.3	7208	11.2	10.2	0.398
3400	Huntington-Ashland	WV	54.6	6128	16.8	11.8	0.42

Table A.3 Metropolitan Area-specific Climate Data

MA Code	Metropolitan Area	State	Mean Maximum Temperature (F)	Variation in Max. Temperature	Mean Relative Humidity (%)	Variation in Relative Humidity
1000	Birmingham	AL	73.483	67.127	46.02	237.971
3440	Huntsville	AL	71.065	77.473	48.543	223.498
5160	Mobile	AL	77.052	49.448	50.281	254.294
4400	Little Rock-North Little Rock	AR	71.181	84.096		
6200	Phoenix	AZ	86.992	45.758	19.766	83.22
8520	Tucson	AZ	83.256	49.208	19.269	90.692
360	Anaheim	CA				
2840	Fresno	CA	76.901	54.597		
4480	Los Angeles-Long Beach	CA	76.304	47.884		
6780	Riverside-San Bernardino	CA	78.156	64.513		
6920	Sacramento	CA	73.455	51.751		
7320	San Diego	CA	71.494	23.846		
7360	San Francisco	CA	66.081	32.754		
7400	San Jose	CA	71.4	39.008		
1720	Colorado Springs	CO	61.971	124.504	29.698	262.457
2080	Denver	CO	64.172	125.832	30.424	247.582
2670	Fort Collins-Loveland	CO	63.044	104.388		
3060	Greeley	CO	64.388	116.908		
6560	Pueblo	CO	68.892	132.146	29.873	265.449
1160	Bridgeport-Milford	CT	59.597	56.018	51.392	301.323
3280	Hartford	CT	60.31	79.545	45.254	258.808
5480	New Haven-Meriden	CT				
8840	Washington	DC	65.809	83.727	47.206	238.952
9160	Wilmington	DE	63.919	71.426	49.176	218.97
2680	Fort Lauderdale-Hollywood-ompano Beach	FL	83.6	14.002		
3600	Jacksonville	FL	79.308	47.381		
5960	Orlando	FL	82.737	32.827		
8280	Tampa-St. Petersburg-learwater	FL	81.664	27.609		
520	Atlanta	GA	72.094	66.377	45.104	253.867
1800	Columbus	GA	76.196	58.475	43.475	250.563
1360	Cedar Rapids	IA	58.802	98.163		

MA Code	Metropolitan Area	State	Mean Maximum Temperature (F)	Variation in Max. Temperature	Mean Relative Humidity (%)	Variation in Relative Humidity
2120	Des Moines	IA	60.72	106.249		
2200	Dubuque	IA				
8920	Waterloo-Cedar Falls	IA	58.179	98.537		
1080	Boise City	ID				
1600	Chicago	IL				
1020	Bloomington	IN	63.451	93.315		
2440	Evansville-Henderson	IN	66.596	90.466	49.832	193.01
2960	Gary-Hammond	IN				
3480	Indianapolis	IN	62.363	87.774	52.142	188.18
7800	South Bend-Mishawaka	IN	59.079	87.935	51.898	180.824
8320	Terre Haute	IN				
8440	Topeka	KS	65.694	106.189	48.159	203.09
9040	Wichita	KS	67.677	106.284	45.517	222.171
4280	Lexington-Fayette	KY	64.885	91.81	50.428	204.753
760	Baton Rouge	LA	77.556	56	50.735	212.94
5560	New Orleans	LA	77.679	50.525	54.455	187.249
7680	Shreveport	LA	76.204	75.25	47.53	224.358
1120	Boston	MA	58.955	80.378	49.99	312.554
5400	New Bedford	MA	58.86	60.35		
8000	Springfield	MA				
9240	Worcester	MA	56.282	79.6	50.363	267.508
720	Baltimore	MD	65.025	78.552	46.635	217.851
730	Bangor	ME	53.971	75.735	50.327	294.875
4240	Lewiston-Auburn	ME	56.062	74.03		
6400	Portland	ME	55.468	69.52	49.602	293.792
2162	Detroit-Ann Arbor	MI	58.236	81.009	51.276	167.709
2640	Flint	MI	56.843	82.282	53.43	180.633
4040	Lansing-East Lansing	MI	56.986	84.858	54.082	183.235
6960	Saginaw-Bay City-Midland	MI	55.339	80.263		
2240	Duluth	MN	49.343	87.054	53.069	233.956
5120	Minneapolis-St Paul	MN	55.624	95.108	49.641	166.658
3760	Kansas City	MO	66.347	102.239		
7040	St. Louis	MO	65.749	102.366	47.598	214.043

MA Code	Metropolitan Area	State	Mean Maximum Temperature (F)	Variation in Max. Temperature	Mean Relative Humidity (%)	Variation in Relative Humidity
3560	Jackson	MS	76.273	72.74	48.247	241.288
880	Billings	MT	60.242	144.673	34.795	207.9
3040	Great Falls	MT	58.208	137.738	37.511	217.616
1520	Charlotte-Gastonia-Rock Hill	NC	70.583	69.045	44.062	237.446
3120	Greensboro-Winston-Salem-igh Point	NC	68.306	76.258		
6640	Raleigh-Durham	NC	70.189	77.629	46.202	242.704
1010	Bismarck	ND	55.897	120.976	46.434	187.583
2520	Fargo-Moorhead	ND	53.816	106.375	51.526	174.358
2985	Grand Forks	ND	53.252	106.479		
4360	Lincoln	NE	64.25	118.318		
5920	Omaha	NE	61.199	113.366	47.655	197.081
4760	Manchester	NH				
6450	Portsmouth-Dover-Rochester	NH				
3640	Jersey City	NJ	58.783	68.45		
5640	Newark	NJ	63.308	74.438	45.266	234.839
875	Paterson	NJ				
8480	Trenton	NJ				
200	Albuquerque	NM	70.483	58.571	25.002	123.587
4120	Las Vegas	NV	79.883	53.78	17.008	76.295
6720	Reno	NV	66.896	75.51	25.769	103.634
160	Albany-Schenectady-Troy	NY	57.754	81.985	50.566	216.129
1280	Buffalo	NY	56.363	89.134	52.5	198.676
2335	Elmira	NY	58.591	86.239		
5380	Nassau-Suffolk	NY	61.657	67.6	48.6	215.178
5600	New York	NY	61.222	56.896	50.301	274.772
6460	Poughkeepsie	NY	59.55	76.607		
6840	Rochester	NY	57.024	90.52	52.63	188.31
8160	Syracuse	NY	57.36	89.319	53.146	198.225
8680	Utica-Rome	NY	55.204	89.55		
80	Akron	OH	59.26	91.836	53.277	195.646
1320	Canton	OH				
1640	Cincinnati	OH	64.694	89.749		
1680	Cleveland	OH	58.937	97.021	52.297	171.211

MA Code	Metropolitan Area	State	Mean Maximum Temperature (F)	Variation in Max. Temperature	Mean Relative Humidity (%)	Variation in Relative Humidity
1840	Columbus	OH	61.774	91.119	49.059	176.035
2000	Dayton-Springfield	OH	61.29	93.068	50.735	176.128
3200	Hamilton-Middletown	OH				
4800	Mansfield	OH	58.65	90.931	56.376	230.089
8080	Steubenville-Weirton	OH	61.634	92.367		
8400	Toledo	OH	58.8193	89.532	53.14	201.258
9320	Youngstown-Warren	OH	58.462	95.962	53.329	207.117
5880	Oklahoma City	OK	70.894	102.006	44.364	228.225
6440	Portland	OR	62.958	52.558	52.228	191.552
240	Allentown-Bethlehem	PA	61.168	73.526	47.959	221.242
280	Altoona	PA	57.564	89.178		
2360	Erie	PA	57.116	89.345	54.118	178.944
3240	Harrisburg-Lebanon-Carlisle	PA	61.506	75.293	47.688	225.686
3680	Johnstown	PA	62.913	93.445		
4000	Lancaster	PA	63.369	70.26		
6160	Philadelphia	PA	63.963	72.352	47.773	207.258
6280	Pittsburgh	PA	60.263	95.953	48.84	186.769
6680	Reading	PA	62.56	77.216		
42069	Scranton (Lackawanna)	PA	58.386	80.435	49.675	208.78
7610	Sharon	PA				
8050	State College	PA	58.861	87.629		
9140	Williamsport	PA	59.912	74.326	47.772	216.125
9280	York	PA				
6480	Providence	RI	60.145	72.594	48.25	302.517
1440	Charleston	SC	75.579	57.336	49.417	232.415
1760	Columbia	SC	75.036	67.391	42.807	242.471
3160	Greenville-Spartanburg	SC	70.653	64.357	46.034	280.725
1560	Chatanooga	TN	70.74	69.21	46.849	228.6
3840	Knoxville	TN	69.53	73.763	48.51	212.604
4920	Memphis	TN	72.094	80.693	48.842	217.506
5360	Nashville	TN	70.104	86.491	46.794	225.746
840	Beaumont-Port Arthur	TX				
1880	Corpus Christi	TX	80.843	56.036	52.82	193.358

MA Code	Metropolitan Area	State	Mean Maximum Temperature (F)	Variation in Max. Temperature	Mean Relative Humidity (%)	Variation in Relative Humidity
1920	Dallas	TX	76.136	89.332	42.965	215.114
2320	El Paso	TX				
2920	Galveston-Texas City	TX	74.607	35.248		
3360	Houston	TX	79.135	60.976	48.95	212.918
4600	Lubbock	TX	73.547	112.154	34.961	302.436
7200	San Angelo	TX	77.379	91.667	36.826	230.562
7240	San Antonio	TX	79.86	65.191	42.169	227.163
8800	Waco	TX	78.117	84.392	44.866	221.071
9080	Wichita Falls	TX	75.349	108.539	40.283	204.141
7160	Salt Lake City-Ogden	UT	63.636	73.673	36.106	119.894
1950	Danville	VA	70.04	84.886		
5720	Norfolk-Virginia Beach-Newport News	VA	68.494	81.635	49.332	214.962
6760	Richmond-Petersburg	VA	68.862	85.025	46.052	246.058
6800	Roanoke	VA	66.731	84.061	45.192	247.912
7600	Seattle	WA	59.591	41.19	54.953	192.37
7840	Spokane	WA	57.418	70.13	47.427	163.122
8200	Tacoma	WA				
2290	Eau Claire	WI	55.378	96.021	49.005	174.49
3800	Kenosha	WI	54.608	86.345		
4720	Madison	WI	56.537	89.688	52.39	193.574
5080	Milwaukee	WI	55.355	90.259	54.514	196.334
6600	Racine	WI	54.375	82.694		
1480	Charleston	WV	66.121	100.641	46.77	222.191
3400	Huntington-Ashland	WV	65.501	98.222	47.839	223.487

Table A.4. Metropolitan Area-specific Health Services and Physical Environment Data

MA Code	Metropolitan Area	State	Doctors per 100,000	Hospital Beds per 100,000	Water Hardness (ppm)	Altitude (m)
1000	Birmingham	AL	248	663	59	183
3440	Huntsville	AL	153	452		195
5160	Mobile	AL	182	628	20	2
4400	Little Rock-North Little Rock	AR	270	904	18	
6200	Phoenix	AZ	194	425	244	332
8520	Tucson	AZ	269	433		727
360	Anaheim	CA	239	379		49
2840	Fresno	CA	182	368	119	90
4480	Los Angeles-Long Beach	CA	261	479	125	101
6780	Riverside-San Bernardino	CA	157	367		262
6920	Sacramento	CA	213	259	21	8
7320	San Diego	CA	229	350	207	13
7360	San Francisco	CA	442	555	58	19
7400	San Jose	CA	257	354	194	27
1720	Colorado Springs	CO	127	435		1831
2080	Denver	CO	250	447	116	1609
2670	Fort Collins-Loveland	CO	148	222		1525
3060	Greeley	CO	123	254		1422
6560	Pueblo	CO	189	990		1421
1160	Bridgeport-Milford	CT			29	
3280	Hartford	CT			12	
5480	New Haven-Meriden	CT			43	
8840	Washington	DC	342	594	102	15
9160	Wilmington	DE	183	693	48	37
2680	Fort Lauderdale-Hollywood-ompano Beach	FL	256	591		3
3600	Jacksonville	FL	175	454	274	
5960	Orlando	FL	156	428	123	32
8280	Tampa-St. Petersburg-learwater	FL	177	568	81	17
520	Atlanta	GA	192	453	20	320
1800	Columbus	GA	132	740	25	76
1360	Cedar Rapids	IA	512	965	99	223
2120	Des Moines	IA	153	647	78	245

MA Code	Metropolitan Area	State	Doctors per 100,000	Hospital Beds per 100,000	Water Hardness (ppm)	Altitude (m)
2200	Dubuque	IA	149	657	282	198
8920	Waterloo-Cedar Falls	IA	137	558	241	264
1080	Boise City	ID	164	463	85	831
1600	Chicago	IL	243	564	133	182
1020	Bloomington	IN	146	253	111	
2440	Evansville-Henderson	IN	176	813	100	120
2960	Gary-Hammond	IN	129	525	136	185
3480	Indianapolis	IN	234	544	241	219
7800	South Bend-Mishawaka	IN	133	459	203	
8320	Terre Haute	IN	134	561	208	155
8440	Topeka	KS	229	1424	100	290
9040	Wichita	KS	209	529	86	398
4280	Lexington-Fayette	KY	354	1056		300
760	Baton Rouge	LA	143	369	38	6
5560	New Orleans	LA	286	659	74	
7680	Shreveport	LA	261	821	68	
1120	Boston	MA			12	6
5400	New Bedford	MA			15	15
8000	Springfield	MA			12	21
9240	Worcester	MA			9	146
720	Baltimore	MD	388	790	50	10
730	Bangor	ME				48
4240	Lewiston-Auburn	ME				37
6400	Portland	ME			14	21
2162	Detroit-Ann Arbor	MI	221	532	98	
2640	Flint	MI	156	408	86	
4040	Lansing-East Lansing	MI	165	398	86	
6960	Saginaw-Bay City-Midland	MI	130	572		
2240	Duluth	MN	156	655	44	189
5120	Minneapolis-St Paul	MN	225	448	68	256
3760	Kansas City	MO	209	649	276	244
7040	St. Louis	MO	224	649	182	143
3560	Jackson	MS	283	1231	47	90
880	Billings	MT	189	440	148	952

MA Code	Metropolitan Area	State	Doctors per 100,000	Hospital Beds per 100,000	Water Hardness (ppm)	Altitude (m)
3040	Great Falls	MT	189	763	151	1016
1520	Charlotte-Gastonia-Rock Hill	NC	133	450	22	213
3120	Greensboro-Winston-Salem-igh Point	NC	192	404	40	
6640	Raleigh-Durham	NC	490	736	30	
1010	Bismarck	ND	210	801		512
2520	Fargo-Moorhead	ND	233	760		274
2985	Grand Forks	ND	253	870		254
4360	Lincoln	NE	169	710	206	358
5920	Omaha	NE	252	833	236	313
4760	Manchester	NH	.	912		69
6450	Portsmouth-Dover-Rochester	NH	..			6
3640	Jersey City	NJ	175	630	36	3
5640	Newark	NJ	265	677	26	45
875	Paterson	NJ	272	425		30
8480	Trenton	NJ	255	739		15
200	Albuquerque	NM	269	549	112	1511
4120	Las Vegas	NV	133	378		616
6720	Reno	NV	249	689	40	1371
160	Albany-Schenectady-Troy	NY	235	563	54	46
1280	Buffalo	NY	259	806	124	183
2335	Elmira	NY	182	814		266
5380	Nassau-Suffolk	NY	299	643		18
5600	New York	NY	355	641	61	9
6460	Poughkeepsie	NY	180	1156		64
6840	Rochester	NY	257	627	82	157
8160	Syracuse	NY	237	443	109	124
8680	Utica-Rome	NY	141	1065	16	129
80	Akron	OH	187	551	106	
1320	Canton	OH	145	608	413	
1640	Cincinnati	OH	241	575	115	208
1680	Cleveland	OH	269	620	128	
1840	Columbus	OH	202	467	90	
2000	Dayton-Springfield	OH	170	720	358	

MA Code	Metropolitan Area	State	Doctors per 100,000	Hospital Beds per 100,000	Water Hardness (ppm)	Altitude (m)
3200	Hamilton-Middletown	OH	100	390		
4800	Mansfield	OH	120	516	229	
8080	Steubenville-Weirton	OH	97	369	111	
8400	Toledo	OH	227	692	170	
9320	Youngstown-Warren	OH	156	551	86	262
5880	Oklahoma City	OK	213	554	90	379
6440	Portland	OR	257	449	9	
240	Allentown-Bethlehem	PA	164	517	157	111
280	Altoona	PA	149	679		357
2360	Erie	PA	154	636	129	227
3240	Harrisburg-Lebanon-Carlisle	PA	204	709	13	110
3680	Johnstown	PA	139	722		360
4000	Lancaster	PA	125	348		116
6160	Philadelphia	PA	276	615	75	14
6280	Pittsburgh	PA	242	721		232
6680	Reading	PA	173	604	88	79
42069	Scranton (Lackawanna)	PA	179	951	26	230
7610	Sharon	PA	102	570		304
8050	State College	PA	129	363		352
9140	Williamsport	PA	174	648		161
9280	York	PA	115	278		122
6480	Providence	RI	..		24	7
1440	Charleston	SC	239	514		36
1760	Columbia	SC	221	1066	29	
3160	Greenville-Spartanburg	SC	158	410		
1560	Chatanooga	TN	213	582	59	209
3840	Knoxville	TN	181	665	116	271
4920	Memphis	TN	237	742	38	80
5360	Nashville	TN	237	766	90	134
840	Beaumont-Port Arthur	TX	126	568	66	
1880	Corpus Christi	TX	153	495		
1920	Dallas	TX	172	450	66	141
2320	El Paso	TX	132	433	146	1147
2920	Galveston-Texas City	TX	256	541	115	

MA Code	Metropolitan Area	State	Doctors per 100,000	Hospital Beds per 100,000	Water Hardness (ppm)	Altitude (m)
3360	Houston	TX	219	527	73	
4600	Lubbock	TX	260	696	422	
7200	San Angelo	TX	157	814		563
7240	San Antonio	TX	195	654	223	
8800	Waco	TX	149	885	156	
9080	Wichita Falls	TX	180	1044	67	291
7160	Salt Lake City-Ogden	UT	214	346	186	1298
1950	Danville	VA	134	379		
5720	Norfolk-Virginia Beach-ewport News	VA	177 570		74	
6760	Richmond-Petersburg	VA	272	757		
6800	Roanoke	VA	248	1161		
7600	Seattle	WA	274	347	18	
7840	Spokane	WA	219	621	156	
8200	Tacoma	WA	135	535	21	
2290	Eau Claire	WI	162	697	54	243
3800	Kenosha	WI	108	352	132	186
4720	Madison	WI	329	684	345	263
5080	Milwaukee	WI	227	646		193
6600	Racine	WI	119	328	130	191
1480	Charleston	WV	190	616	32	189
3400	Huntington-Ashland	WV	151	544		173

Table A.5 Metropolitan Area-specific Gaseous Co-pollutant Data

MA Code	Metropolitan Area	State	Fine Part ($\mu g/m^3$)	Sulfates ($\mu g/m^3$)	CO (ppb)	NO_2 (ppb)	O_3 (ppb)	SO_2 (ppb)
1000	Birmingham	AL	24.51	13.1	1497.74	.	24.2799	.
3440	Huntsville	AL		12	833.53	23.3892	27.4492	.
5160	Mobile	AL	20.89	12.6	.	.	.	.
4400	Little Rock-North Little Rock	AR	17.84	5.9	.	.	28.2315	.
6200	Phoenix	AZ	15.24	4.3	1950.43	7.7505	21.8269	5.595
8520	Tucson	AZ		4.378	2260.24	30.3248	27.9032	0.8273
360	Anaheim	CA		11.5	3178.48	43.49	21.4205	7.461
2840	Fresno	CA	10.32	5.8	1240.98	32.2721	29.8996	2.7319
4480	Los Angeles-Long Beach	CA	21.81	14	3050.16	51.0586	27.957	7.3806
6780	Riverside-San Bernardino	CA		14.6	2008.63	37.136	41.14	4.0829
6920	Sacramento	CA		5.8	1116.45	22.3883	25.4675	1.6195
7320	San Diego	CA		11.2	1165.04	27.5719	34.4784	5.7941
7360	San Francisco	CA	12.16	6.6	1514.5	26.0611	10.9253	0.6046
7400	San Jose	CA	12.44	6.2	1735.14	35.0364	19.1518	.
1720	Colorado Springs	CO		6.1	1828.34	27.3435	26.4474	.
2080	Denver	CO	16.09	6.2	2371.84	43.635	23.8645	12.8136
2670	Fort Collins-Loveland	CO		5.2	1819.44	.	26.7062	.
3060	Greeley	CO		4.7	1316.97	.	23.4081	.
6560	Pueblo	CO		6.7	1106.83	.	.	.
1160	Bridgeport-Milford	CT		9.9	3017.08	27.7215	35.6284	11.6687
3280	Hartford	CT	14.77	9.4	2134.18	29.0827	26.2289	10.9828
5480	New Haven-Meriden	CT		8.5	939.98	.	28.4976	8.3722
8840	Washington	DC	22.51	14.9	1010.07	24.5431	21.962	11.841
9160	Wilmington	DE		19.4	1109.56	31.3804	23.0042	6.5264
2680	Fort Lauderdale-Hollywood-ompano Beach	FL		6.9	1827.1	.	20.8301	2.1638
3600	Jacksonville	FL		11.2	1214.68	14.5925	31.5988	6.6877
5960	Orlando	FL		7.9	1247.29	.	22.5354	3.2398
8280	Tampa-St. Petersburg-learwate	FL	11.42	10.3	996.82	15.8848	21.2725	5.545
520	Atlanta	GA	20.29	12	1620.18	32.0999	21.8366	9.1701
1800	Columbus	GA		9.4	.	.	18.6757	.
1360	Cedar Rapids	IA		10.6	928.47	.	19.0759	10.4142
2120	Des Moines	IA		10.6	2893.02	.	36.6805	7.1919

MA Code	Metropolitan Area	State	Fine Part ($\mu g/m^3$)	Sulfates ($\mu g/m^3$)	CO (ppb)	NO_2 (ppb)	O_3 (ppb)	SO_2 (ppb)
2200	Dubuque	IA		8.9	2292.66	.	.	3.5766
8920	Waterloo-Cedar Falls	IA		9.1	.	.	.	3.245
1080	Boise City	ID	12.12		2915.17	.	.	3.011
1600	Chicago	IL	21.04		1847.58	34.7521	18.4675	12.3266
1020	Bloomington	IN		13.7	.	.	.	.
2440	Evansville-Henderson	IN		14.2	548.86	19.5714	16.7954	10.9113
2960	Gary-Hammond	IN	25.24	19.1	1253.95	.	18.1991	14.9361
3480	Indianapolis	IN	21.09	12.6	2288.52	36.0572	22.3678	14.7531
7800	South Bend-Mishawaka	IN		11.7	.	.	17.2343	11.6914
8320	Terre Haute	IN		13.2	.	.	.	15.1772
8440	Topeka	KS	10.33	6.8	.	.	.	.
9040	Wichita	KS	13.59	4.9	1236.13	.	24.0435	.
4280	Lexington-Fayette	KY		14.3	1280.73	18.091	23.8779	7.219
760	Baton Rouge	LA		11.2	.	.	20.3814	11.3686
5560	New Orleans	LA		14.6	.	20.9763	14.0168	2.7248
7680	Shreveport	LA		10.1	.	.	.	.
1120	Boston	MA		11	1992.48	32.3804	27.9584	10.445
5400	New Bedford	MA		11.8	.	.	.	11.5469
8000	Springfield	MA		12.8	3952.27	.	28.0353	13.1876
9240	Worcester	MA		10.7	.	.	31.2428	12.6826
720	Baltimore	MD		13	1831.54	26.0584	21.3704	10.7959
730	Bangor	ME		10.2	1137.05	.	.	8.6328
4240	Lewiston-Auburn	ME		10.7	1368.48	.	.	18.2383
6400	Portland	ME		11.25	.	.	38.7877	9.2342
2162	Detroit-Ann Arbor	MI		14.7	1171.14	25.5844	18.3015	7.5792
2640	Flint	MI		9.5	.	.	17.8606	5.1382
4040	Lansing-East Lansing	MI		13.1	.	30.9568	25.9848	12.5026
6960	Saginaw-Bay City-Midland	MI		11.3	1069.98	25.2117	.	5.7974
2240	Duluth	MN		7	2489.58	.	22.8789	5.9161
5120	Minneapolis-St Paul	MN	13.68	8.4	1862.99	23.1245	23.9132	9.6014
3760	Kansas City	MO		10.15	2273.33	.	28.0396	.181
7040	St. Louis	MO		13.5	1225.57	24.8372	23.2838	.788
3560	Jackson	MS	15.71	8.8	.	14.4303	27.4267	.
880	Billings	MT		7.1	1166.99	13.7613	16.7379	13.0969

MA Code	Metropolitan Area	State	Fine Part ($\mu g/m^3$)	Sulfates ($\mu g/m^3$)	CO (ppb)	NO_2 (ppb)	O_3 (ppb)	SO_2 (ppb)
3040	Great Falls	MT		3.6	840.74	9.5454	.	1.6874
4360	Lincoln	NB		6.6	1519.71	.	24.8052	.
5920	Omaha	NB	13.07	8.7	955.85	.	18.5345	6.5079
1520	Charlotte-Gastonia-Rock Hill	NC	22.6	11.5	1783.98	.	21.1752	8.789
3120	Greensboro-Winston-Salem-igh Point	NC		12.9	1314.84	.	26.0647	.
6640	Raleigh-Durham	NC	16.79	11.9	2489.21	.	.	.
1010	Bismarck	ND		5.267	.	.	.	.
2520	Fargo-Moorhead	ND		6.3	.	.	.	.
2985	Grand Forks	ND		4.8				
4760	Manchester	NH		11.05	1973.59	.	25.4629	8.8532
6450	Portsmouth-Dover-Rochester	NH		8.7	.	.	.	.
3640	Jersey City	NJ	17.34	13.8	2060.34	30.3302	22.0809	14.1362
5640	Newark	NJ		11.4	2711.83	33.4288	21.8421	11.9722
875	Paterson	NJ		12.8	.	30.9951	25.1851	10.317
8480	Trenton	NJ		12.1	1349.85	.	26.5732	9.0038
200	Albuquerque	NM	8.95	4.533	1608.82	19.6085	25.692	1.8612
4120	Las Vegas	NV		4.2	2587.61	35.1434	21.8161	.
6720	Reno	NV	11.82	4.1	2481.17	48.5684	23.2992	.
160	Albany-Schenectady-Troy	NY		10.5	627.75	.	19.7	15.1519
1280	Buffalo	NY	23.53	11.7	919.92	23.2878	25.3045	13.423
2335	Elmira	NY		6.6	573.72	.	22.8194	7.8
5380	Nassau-Suffolk	NY		8.4	1307.35	27.4797	23.9209	7.8225
5600	New York	NY		10.7	2194.97	34.7118	21.058	18.7846
6460	Poughkeepsie	NY		8.1	1036.04	.	14.1345	11.9676
6840	Rochester	NY		9.9	1096.44	.	21.0899	19.627
8160	Syracuse	NY		10.2	862.77	21.0792	20.3576	10.1229
8680	Utica-Rome	NY		8.6	.	.	.	9.3356
80	Akron	OH	24.6	14.1	877.59	.	27.9995	.
1320	Canton	OH		13.6	1094.5	28.8635	25.086	26.0653
1640	Cincinnati	OH	23.12	14.3	1438.41	24.9634	20.284	11.9489
1680	Cleveland	OH	24.58	13.7	2269.14	.	20.6126	13.7568
1840	Columbus	OH		11.8	1448.75	19.3025	17.0739	9.0792
2000	Dayton-Springfield	OH	18.8	13.5	1258.96	21.9899	21.1564	.

MA Code	Metropolitan Area	State	Fine Part ($\mu g/m^3$)	Sulfates ($\mu g/m^3$)	CO (ppb)	NO_2 (ppb)	O_3 (ppb)	SO_2 (ppb)
3200	Hamilton-Middletown	OH		13.2	835.82	15.277	22.3161	12.3369
4800	Mansfield	OH		13.1	.	.	.	7.9128
8080	Steubenville-Weirton	OH	23.09	23.5	2311.68	22.6784	19.1758	29.3167
8400	Toledo	OH		10.8	864.06	.	25.1098	8.5724
9320	Youngstown-Warren	OH	20.23	15.7	.	40.712	19.0045	10.6884
5880	Oklahoma City	OK	15.86	6.3	535.01	19.796	22.6629	0.0246
6440	Portland	OR	14.66	7.7	2174.62	27.268	10.4054	6.0041
240	Allentown-Bethlehem	PA	17.88	14.5	1798.81	24.9068	23.6763	11.0875
280	Altoona	PA		14.2	.	.	.	.
2360	Erie	PA		12.1	.	17.1442	29.4256	13.2245
3240	Harrisburg-Lebanon-Carlisle	PA		11.9	1924.6	22.0135	22.5786	9.7296
3680	Johnstown	PA		17.2	751.59	23.9155	22.4635	21.6266
4000	Lancaster	PA		12.5	.	.	.	.
6160	Philadelphia	PA	21.4	11.469	1403.37	33.7065	24.824	15.1254
6280	Pittsburgh	PA		15.8	2325.69	25.4752	22.9163	20.6491
6680	Reading	PA		12.1	.	25.4654	20.2168	14.2785
42069	Scranton (Lackawanna)	PA		11.7	.	.	.	.
7610	Sharon	PA		12.4	.	.	.	.
8050	State College	PA		15.3	.	.	.	.
9140	Williamsport	PA		12.4	.	.	21.4821	7.1145
9280	York	PA		13.217	1382.53	25.8918	24.6722	12.3812
6480	Providence	RI	12.89	8.7	1718.62	35.3876	22.6853	13.839
1440	Charleston	SC		13.8	.	.	.	5.609
1760	Columbia	SC		16	1083.87	.	25.8152	3.3531
3160	Greenville-Spartanburg	SC		11	1857.83	.	.	2.1164
1560	Chatanooga	TN	16.64	13.9	.	.	.	.
3840	Knoxville	TN		13.9	3168.89	.	19.9358	7.5971
4920	Memphis	TN		12.8	1820.59	15.6636	26.8985	10.426
5360	Nashville	TN	20.45	8.7	2166.5	.	17.4195	9.321
840	Beaumont-Port Arthur	TX		14.4	347.83	13.3138	27.7125	.2966
1880	Corpus Christi	TX		8.8	537.37	.	25.0577	4527
1920	Dallas	TX	16.48	10	746.09	18.142	24.9043	1.5769
2320	El Paso	TX	15.74		1635.98	24.805	21.6872	13.6428
2920	Galveston-Texas City	TX		9	194.2	16.9444	25.0872	4.8061

MA Code	Metropolitan Area	State	Fine Part ($\mu g/m^3$)	Sulfates ($\mu g/m^3$)	CO (ppb)	NO_2 (ppb)	O_3 (ppb)	SO_2 (ppb)
3360	Houston	TX	13.37	10.5	771.74	25.7925	22.2223	4.3977
4600	Lubbock	TX		4.5	.	.	.	.
7200	San Angelo	TX		4.4	.	.	.	.
7240	San Antonio	TX		6.6	671.4	15.5472	21.8569	0.847
8800	Waco	TX		10	.	.	.	.
9080	Wichita Falls	TX		6.9	.	.	.	.
7160	Salt Lake City-Ogden	UT	15.39	4.8	1770.17	27.2913	25.2682	16.537
1950	Danville	VA		13.6	.	.	.	.
5720	Norfolk-Virginia Beach-ewport News	VA	16.91	14.8	1159.63	.	35.1657	11.5687
6760	Richmond-Petersburg	VA		11	1409.57	30.5673	32.3228	12.1028
6800	Roanoke	VA		12.4	.	.	21.6118	6.413
7600	Seattle	WA	11.87	7.5	2360	.	14.5649	5.0855
7840	Spokane	WA	9.38	5.6	2579.79	.	15.0641	9.984
8200	Tacoma	WA		6.8	2718.32	.	15.0644	5.1042
2290	Eau Claire	WI		8.3	.	.	.	.
3800	Kenosha	WI		10.8	.	12.7029	29.0526	5.4446
4720	Madison	WI		9.8	.	.	27.1173	6.6926
5080	Milwaukee	WI		11.6	969.92	18.4156	22.6244	6.9378
6600	Racine	WI		11.1	616.33	.	23.349	6.6131
1480	Charleston	WV	20.1	17.8	971.63	18.8835	15.6787	15.0005
3400	Huntington-Ashland	WV	33.35	15.3	1049.67	11.7755	20.8387	9.3366

Table A.6. County-specific Data

County	State	County Code	Percent Population Change ('80-'86)	Percent White	Percent African American	Doctors per 100,000	Hospital Beds per 100,000	Percent Finishing Highschool	Median Income (1979)	Poverty Rate Individuals (%)	Gini Coefficient	Unemployment Rate (%)	Sulfates ($\mu g/m^3$)	SO_2 (ppb)
Broward	Florida	12011	12.2	86.48	11.16	188	586	70.4	16580	9.1	0.40	4.4	6.932	
Duval	Florida	12031	13.2	73.21	24.62	203	529	66.8	14938	15.8	0.41	5.3	11.229	7.339
Hillsborough	Florida	12057	19.9	85.61	13.36	215	567	65.8	14868	14.0	0.41	5.7	9.738	4.853
Orange	Florida	12095	22.2	83.17	14.77	188	547	70.4	15298	13.2	0.40	4.7	7.859	3.246
Pinellas	Florida	12103	11.9	91.44	7.58	174	710	68.3	13404	9.8	0.41	4.2	12.176	
Fulton	Georgia	13121	5.6	46.88	51.45	458	912	66.0	13988	21.2	0.47	6.3	12.018	10.152
Muscogee	Georgia	13215	6.4	62.32	34.03	168	752	60.9	13343	18.0	0.43	6.9	9.421	
Lake	Indiana	18089	-6.0	73.62	24.10	129	567	63.3	21307	11.0	0.37	12.2	19.110	17.100
Monroe	Indiana	18105	2.9	95.76	2.60	146	253	74.7	13721	15.0	0.43	3.9	13.671	
St. Joseph	Indiana	18141	-0.1	89.43	8.94	153	525	67.6	17570	9.2	0.38	5.9	11.650	11.764
Vanderburgh	Indiana	18163	0.1	91.87	7.12	233	1208	64.2	16070	10.3	0.40	5.9	14.217	11.330
Vigo	Indiana	18167	-2.6	93.64	5.49	151	611	68.3	15224	10.7	0.39	7.0	13.218	11.859
Black Hawk	Iowa	19013	-7.5	92.28	6.23	147	584	73.9	19494	9.3	0.37	12.7	9.119	3.274
Dubuque	Iowa	19061	-2.9	99.20	0.25	149	657	68.6	19396	8.3	0.38	7.7	8.950	3.607
Linn	Iowa	19113	-0.6	97.59	1.61	140	572	77.6	20084	6.8	0.36	6.2	10.626	10.285
Polk	Iowa	19153	4.2	92.84	4.52	184	791	78.8	18849	8.4	0.38	5.8	10.615	7.009
Sedgwick	Kansas	20173	6.6	87.26	8.67	229	535	76.5	18223	9.1	0.38	6.0	11.717	
Shawnee	Kansas	20177	3.8	89.92	7.64	229	1424	78.3	17713	7.8	0.37	5.1	6.761	
Fayette	Kentucky	21067	4.3	84.91	13.28	513	1510	71.6	15915	13.5	0.41	4.3	14.288	7.297
Kenton	Kentucky	21117	0.4	96.55	2.75	167	576	59.0	17195	10.1	0.38	6.1	15.464	
East Baton Rouge	Louisiana	22033	7.2	66.06	31.33	188	477	72.3	18041	14.9	0.42	9.4	11.242	11.383
Orleans	Louisiana	22071	-0.6	39.33	55.27	447	1037	59.2	11814	26.4	0.48	11.2	14.564	
Androscoggin	Maine	23001	1.6	99.32	0.24	174	545	58.5	13524	12.6	0.38	6.9	10.717	18.245
Cumberland	Maine	23005	5.7	98.71	0.46	270	687	75.0	15359	10.5	0.39	2.7	10.731	9.781
Penobscot	Maine	23019	0.9	98.54	0.20	183	742	71.9	14181	13.0	0.38	6.4	10.222	8.547
Prince Georges	Maryland	24033	2.4	53.57	37.27	156	276	77.4	22395	6.7	0.34	3.3	11.252	
Bristol	Massachusetts	25005	2.2	97.36	1.01	105	417	52.6	15473	10.1	0.39	5.8	11.721	8.889
Hampden	Massachusetts	25013	0.4	98.87	6.27	193	642	66.5	16166	11.7	0.39	4.3	12.844	12.966
Middlesex	Massachusetts	25017	.	96.13	1.85	321	735	77.4	20433	7.0	0.38	2.9	10.521	12.161
Plymouth	Massachusetts	25023	4.7	96.36	2.01	120	627	77.1	18749	8.0	0.37	4.5	8.462	
Worcester	Massachusetts	25027	2.3	97.61	1.35	259	679	66.7	17182	9.1	0.38	4.0	10.147	12.358
Genesee	Michigan	26049	-3.5	80.14	17.49	156	408	67.8	20996	10.6	0.37	10.6	9.540	5.121
Ingham	Michigan	26065	0.8	89.68	7.72	240	559	78.0	18089	13.2	0.39	7.3	14.100	11.994
Wayne	Michigan	26163	-7.4	61.09	35.50	175	654	61.4	18629	14.3	0.41	8.9	15.684	7.974

Table A.6. County-specific Data

County	State	County Code	Percent Population Change ('80-'86)	Percent White	Percent African American	Doctors per 100,000	Hospital Beds per 100,000	Percent Finishing Highschool	Median Income (1979)	Poverty Rate Individuals (%)	Gini Coeffic ient	Unemplo yment Rate (%)	Sulfates ($\mu g/m^3$)	SO_2 (ppb)
Hennepin	Minnesota	27053	4.9	92.89	3.50	340	683	81.7	20077	7.5	0.39	3.9	7.824	10.702
Ramsey	Minnesota	27123	3.1	92.30	3.20	273	380	77.2	18939	8.1	0.38	4.2	8.618	8.241
Hinds	Mississippi	28049	3.5	52.11	45.14	391	1132	70.0	14976	18.9	0.43	8.1	8.753	
Cascade	Montana	30013	-1.6	94.27	1.22	189	763	75.2	16050	10.3	0.39	2.7	3.780	1.687
Yellowstone	Montana	30111	11.2	96.75	0.27	189	440	76.6	17460	9.3	0.39	7.8	6.143	12.821
Douglas	Nebraska	31055	4.5	88.05	10.03	341	1050	76.8	17720	9.9	0.39	5.5	8.673	5.350
Lancaster	Nebraska	31109	6.8	96.03	1.83	169	710	81.5	17428	8.6	0.38	3.1	6.564	
Clark	Nevada	32003	23.0	86.08	9.99	133	378	74.0	18102	9.1	0.39	6.2	4.245	
Washoe	Nevada	32031	16.0	93.33	1.83	249	689	80.1	19560	7.0	0.38	5.1	4.124	
Hillsborough	New Hampshire	33011	13.6	98.62	0.50	147	441	71.1	18689	7.2	0.37	2.8	11.044	15.414
Merrimack	New Hampshire	33013	11.6	99.37	0.22	165	852	74.0	16717	8.0	0.37	2.0	10.489	12.354
Rockingham	New Hampshire	33015	16.5	98.42	0.01	112	264	77.3	18993	6.6	0.35	3.7	8.658	
Hudson	New Jersey	34017	-0.7	82.31	12.58	175	630	51.5	14384	16.9	0.42	8.0	14.071	14.267
Mercer	New Jersey	34021	4.2	78.86	18.04	255	739	67.8	19659	9.4	0.39	4.3	12.081	8.834
Passaic	New Jersey	34031	3.0	84.21	13.22	169	425	58.3	17907	12.8	0.40	6.0	12.823	
Bernalillo	New Mexico	35001	12.9	92.74	2.33	269	549	76.5	16239	13.2	0.40	6.5	4.798	1.871
Albany	New York	36001	-0.9	91.79	6.59	409	854	71.7	17006	10.2	0.39	4.2	10.773	22.052
Bronx	New York	36005	2.1	61.48	31.82	225	631	50.8	10947	27.6	0.45	8.4	12.070	22.182
Chemung	New York	36015	-7.3	95.23	4.03	182	814	69.5	15213	11.0	0.38	6.8	6.579	7.871
Dutchess	New York	36027	4.8	90.85	6.99	180	1156	70.6	20267	7.3	0.37	3.8	8.095	11.952
Erie	New York	36029	-5.0	88.04	10.14	259	806	65.4	17119	10.6	0.39	7.2	11.040	13.441
Kings	New York	36047	2.8	60.89	32.40	201	448	54.9	11919	24.0	0.44	8.6	12.932	18.655
Monroe	New York	36055	0.1	87.69	10.12	320	586	71.7	20194	8.8	0.38	5.0	9.922	18.785
Nassau	New York	36059	0.1	90.48	6.87	402	356	77.7	26090	4.8	0.36	4.0	8.326	8.028
New York	New York	36061	3.5	71.05	21.69	855	1258	68.0	13904	21.8	0.49	7.0	12.404	27.223
Oneida	New York	36065	-2.0	96.15	3.05	163	1247	65.0	15261	11.0	0.38	6.3	7.971	9.240
Onondaga	New York	36067	-0.2	91.35	6.49	297	496	72.7	17574	9.6	0.38	6.4	8.697	10.113
Queens	New York	36081	1.7	72.49	18.72	222	418	63.1	17028	11.4	0.39	6.3	9.973	16.283
Rensselaer	New York	36083	-0.1	96.48	2.71	127	420	66.0	15970	11.2	0.38	5.1	8.115	
Richmond	New York	36085	6.4	89.49	7.27	280	484	68.5	21204	8.2	0.37	5.3	11.889	15.740
Schenectady	New York	36093	0.2	95.28	3.12	257	554	71.5	16928	8.7	0.39	4.8	8.754	11.757
Suffolk	New York	36103	2.2	92.64	5.59	193	929	73.7	22359	6.6	0.35	4.8	9.234	7.859

Table A.6. County-specific Data

County	State	County Code	Percent Population Change ('80-'86)	Percent White	Percent African American	Doctors per 100,000	Hospital Beds per 100,000	Percent Finishing Highschool	Median Income (1979)	Poverty Rate Individuals (%)	Gini Coefficient	Unemployment Rate (%)	Sulfates ($\mu g/m^3$)	SO_2 (ppb)
Westchester	New York	36119	-0.4	84.29	12.09	456	623	75.2	22725	7.1	0.41	3.8	9.131	10.106
Durham	North Carolina	37063	9.4	60.62	36.28	931	1189	65.1	15395	14.0	0.40	3.6	11.938	
Forsyth	North Carolina	37067	6.7	75.09	24.38	409	631	62.7	16600	11.6	0.40	4.6	12.879	
Mecklenburg	North Carolina	37119	11.5	71.43	26.47	201	481	69.3	17837	10.9	0.39	3.9	14.043	
Butler	Ohio	39017	4.9	94.34	4.66	100	390	64.7	19584	9.8	0.37	7.7	12.598	12.423
Clark	Ohio	39023	-1.9	90.99	8.77	119	442	65.3	16842	11.6	0.38	7.0	12.127	
Cuyahoga	Ohio	39035	-3.5	74.87	22.76	349	765	66.7	18009	11.5	0.41	7.4	13.696	13.819
Franklin	Ohio	39049	4.4	83.19	15.07	258	580	73.0	17081	12.3	0.39	5.5	11.792	9.128
Hamilton	Ohio	39061	-0.9	79.15	19.01	335	783	65.0	17446	11.3	0.41	6.1	13.724	11.025
Jefferson	Ohio	39081	-6.4	93.82	5.51	103	658	62.3	18173	10.2	0.38	9.4	23.521	26.785
Montgomery	Ohio	39113	-0.9	81.59	16.54	213	953	69.4	17632	11.0	0.39	6.5	13.489	
Richland	Ohio	39139	-1.8	92.26	7.14	120	516	66.1	17162	9.4	0.37	9.2	13.089	8.106
Stark	Ohio	39151	-1.4	92.54	6.36	153	650	63.8	18620	8.5	0.37	10.1	13.223	25.821
Summit	Ohio	39153	-3.2	87.98	10.84	219	634	69.2	18381	9.4	0.38	7.9	12.598	
Trumbull	Ohio	39155	-3.5	92.85	6.04	100	424	68.2	19536	8.1	0.37	11.2	12.520	7.782
Warren	Ohio	39165	5.3	97.79	1.72	50	0	62.7	20769	7.3	0.34	7.7	11.914	7.159
Wood	Ohio	39173	2.7	98.02	1.22	92	115	75.0	18855	10.0	0.37	6.6	12.595	
Oklahoma	Oklahoma	40109	10.8	82.64	12.35	285	661	73.7	16456	10.7	0.40	6.8	6.253	0.025
Multnomah	Oregon	41051	0.8	88.74	5.30	403	718	76.4	16078	11.4	0.40	8.5	7.671	5.924
Allegheny	Pennsylvania	42003	-5.3	88.24	10.36	333	917	69.0	17944	9.2	0.40	7.1	15.907	19.816
Berks	Pennsylvania	42011	2.7	96.46	2.47	173	604	58.5	17530	8.2	0.37	6.4	11.891	14.257
Blair	Pennsylvania	42013	-3.0	98.77	0.70	149	679	65.4	14768	10.9	0.38	9.2	14.259	
Bucks	Pennsylvania	42017	9.0	96.04	2.49	138	351	74.7	22016	5.8	0.35	5.0	10.897	13.550
Cambria	Pennsylvania	42021	-5.5	97.80	1.85	162	686	61.1	15772	9.7	0.38	10.4	16.589	21.663
Carbon	Pennsylvania	42025	2.1	99.59	0.06	74	510	57.4	15369	7.6	0.37	10.4	24.121	
Centre	Pennsylvania	42027	1.6	97.07	1.30	129	363	75.8	14862	16.7	0.40	6.0	15.270	
Chester	Pennsylvania	42029	7.1	91.71	7.29	142	963	76.4	22206	6.4	0.37	4.0	11.524	
Cumberland	Pennsylvania	42041	5.7	97.43	1.38	167	407	72.9	19270	5.8	0.35	3.6	13.084	
Dauphin	Pennsylvania	42043	1.9	84.47	13.46	304	866	69.9	17139	9.9	0.37	5.0	11.945	
Delaware	Pennsylvania	42045	1.0	88.87	9.01	221	361	72.3	19950	7.4	0.38	4.4	11.470	11.152
Erie	Pennsylvania	42049	-0.2	94.64	4.44	154	636	69.9	16760	10.0	0.37	8.5	11.725	13.448
Fayette	Pennsylvania	42051	-2.3	95.89	3.79	73	283	57.8	14282	15.5	0.40	11.6	13.402	

Table A.6. County-specific Data

County	State	County Code	Percent Population Change ('80-'86)	Percent White	Percent African American	Doctors per 100,000	Hospital Beds per 100,000	Percent Finishing Highschool	Median Income (1979)	Poverty Rate Individuals (%)	Gini Coefficient	Unemployment Rate (%)	Sulfates ($\mu g/m^3$)	SO_2 (ppb)
Lackawanna	Pennsylvania	42069	-2.2	98.87	0.49	179	951	63.8	14267	9.8	0.39	7.6	12.757	11.731
Lancaster	Pennsylvania	42071	8.6	96.59	1.92	125	348	59.6	17933	8.5	0.36	3.4	12.554	
Lebanon	Pennsylvania	42075	3.1	98.89	0.43	123	1157	59.6	17669	7.4	0.35	5.6	12.089	
Lehigh	Pennsylvania	42077	3.4	97.22	1.52	200	651	64.6	18790	7.2	0.37	7.4	12.096	12.182
Lycoming	Pennsylvania	42081	-1.8	98.07	1.48	174	648	65.6	15025	10.8	0.37	7.1	12.408	7.051
Mercer	Pennsylvania	42085	-3.7	95.25	4.25	102	570	66.8	17160	9.1	0.37	8.4	12.420	
Montgomery	Pennsylvania	42091	4.5	92.60	4.80	421	657	75.8	22508	4.7	0.37	4.3	12.100	11.749
Northampton	Pennsylvania	42095	3.9	97.37	1.76	168	413	61.8	18470	7.8	0.37	7.7	12.804	6.261
Philadelphia	Pennsylvania	42101	-2.7	58.79	37.84	385	852	54.3	13169	20.6	0.43	6.9	19.352	17.072
Washington	Pennsylvania	42125	-2.1	95.99	3.47	102	334	62.4	17664	8.9	0.38	8.6	15.333	15.585
York	Pennsylvania	42133	4.3	96.17	2.90	125	310	61.8	18389	7.0	0.36	5.3	13.054	12.411
Kent	Rhode Island	44003	3.4	98.70	0.39	142	226	66.8	18455	7.1	0.37	3.7	7.327	
Newport	Rhode Island	44005	4.2	94.58	3.70	124	411	72.0	16847	9.9	0.38	3.6	9.352	
Providence	Rhode Island	44007	1.8	93.27	4.04	276	769	56.1	14834	11.9	0.41	4.4	10.082	13.825
Charleston	South Carolina	45019	3.5	64.15	34.39	384	862	63.6	14826	17.9	0.41	4.8	13.311	5.778
Greenville	South Carolina	45045	6.4	80.91	17.66	194	513	56.1	15998	11.8	0.40	4.8	9.972	2.331
Pickens	South Carolina	45077	10.4	92.81	7.38	61	193	50.5	15586	10.7	0.38	5.8	10.484	
Richland	South Carolina	45079	1.8	58.01	38.57	316	1567	67.4	15691	15.3	0.40	4.2	10.477	3.312
York	South Carolina	45091	13.2	76.93	22.26	86	317	52.6	17075	11.0	0.37	5.2	10.315	4.758
Davidson	Tennessee	47037	4.2	76.51	22.26	402	1159	66.1	16408	12.3	0.40	4.3	9.311	11.677
Hamilton	Tennessee	47065	-1.2	79.28	19.41	238	723	63.7	15694	13.5	0.41	6.8	13.926	
Shelby	Tennessee	47157	4.2	55.60	41.78	276	863	65.9	15289	19.6	0.43	6.6	12.777	11.145
Bexar	Texas	48029	18.3	91.55	7.00	211	719	63.1	15085	18.5	0.41	7.4	6.066	0.801
Dallas	Texas	48113	17.8	79.22	18.47	234	516	71.2	18571	10.6	0.40	5.7	9.133	1.465
Harris	Texas	48201	16.1	76.95	19.66	242	578	70.5	20805	10.4	0.39	10.3	10.673	3.724
Roanoke	Virginia	51161	2.1	97.39	2.31	272	326	70.0	20458	5.8	0.35	3.5	12.433	6.431
Hampton	Virginia	51650	2.8	61.29	34.31	107	852	67.5	16971	11.7	0.38	5.6	12.518	
Norfolk	Virginia	51710	2.9	61.59	35.20	259	739	61.7	12509	20.7	0.43	4.9	14.265	
King	Washington	53033	7.3	87.79	4.41	326	400	82.6	20717	7.7	0.38	6.2	5.943	4.998
Spokane	Washington	53063	4.4	95.63	1.26	219	621	78.2	15930	11.5	0.40	8.2	5.617	9.984

Table A.6. County-specific Data

County	State	County Code	Percent Population Change ('80-'86)	Percent White	Percent African American	Doctors per 100,000	Hospital Beds per 100,000	Percent Finishing Highschool	Median Income (1979)	Poverty Rate Individuals (%)	Gini Coefficient	Unemployment Rate (%)	Sulfates ($\mu g/m^3$)	SO_2 (ppb)
Dane	Wisconsin	55025	6.6	96.10	1.76	422	676	83.7	18309	9.7	0.38	4.5	9.786	9.656
Eau Claire	Wisconsin	55035	5.5	98.04	0.18	214	676	75.0	15300	12.9	0.38	5.9	8.258	
Milwaukee	Wisconsin	55079	-3.4	81.41	15.49	274	823	68.5	18122	10.2	0.38	6.5	11.608	8.776
Racine	Wisconsin	55101	-0.5	90.65	8.03	119	328	67.7	20944	7.0	0.36	8.5	11.097	6.654

Appendix B - Assignment of ACS Cohort Members to Metropolitan Areas and Selection of Counties Used to Aggregate Metropolitan Area-Specific Covariates

Table B.1. Assignment of ACS Cohort Members to Metropolitan Areas and Selection of Counties Used to Aggregate Metropolitan Area-specific Covariates

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
1000	BIRM	AL	350, 351, 352	1000	MSA		01009 Blount County, 01073 Jefferson County 01117 Shelby County 01115 St. Clair County 01127 Walker County
3440	HUNT	AL	357, 358	3440	MSA		01089 Madison County
5160	MOBI	AL	365, 366	5160	MSA		01003 Baldwin County 01097 Mobile County
4400	LROC	AR	720, 721, 722	4400	MSA		05045 Faulkner County 05085 Lonoke County 05119 Pulaski County 05125 Saline County
6200	PHOE	AZ	850, 852, 853	6200	MSA		04013 Maricopa County
8520	TUCS	AZ	856, 857	8520	MSA		04019 Pima County,
360	ANAH	CA	926, 927, 928	360	PMSA		06059 Orange County
2840	FRES	CA	936, 937, 938	2840	MSA		06019 Fresno County
4480	LANG	CA	900, 901, 902, 903, 904, 905, 906, 907, 908, 910, 911, 912, 913, 914, 915, 916, 917 918	4480	PMSA		06037 Los Angeles County
6780	RIVE	CA	923, 924, 925	6780	PMSA		06065 Riverside County 06071 San Bernardino County

Original Analysis				Re-analysis				
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis	
6920	SACR	CA	942, 956, 957, 958	6920	MSA		06017 06061 06067 06113	El Dorado County Placer County Sacramento County Yolo County
7320	SDIE	CA	920, 921	7320	MSA		06073	San Diego County
7360	SFRA	CA	940, 941, 943, 944	7360	PMSA		06041 06075 06081	Marin County San Francisco County San Mateo County
7400	SJOS	CA	950, 951	7400	PMSA		06085	Santa Clara County
1720	COLO	CO	808, 809	1720	MSA		08041	El Paso County
2080	DENV	CO	800, 801, 802	2080	PMSA		08001 08005 08031 08035 08059	Adams County Arapahoe County Denver County Douglas County Jefferson County
2670	FCOL	CO	805	2670	MSA		08069	Larimer County
3060	GREE	CO	806	3060	MSA		08123	Weld County
6560	PUEB	CO	810	6560	MSA		08101	Pueblo County
1163	BRID	CT	66	1160	PMSA	The MA number "1163" does not exist in the 1983 MSA definitions. The ZIP code appears to refer to PMSA 1160.	09001 09009 09001 09009	Fairfield County (pt.) New Haven County (pt.) Fairfield County (pt.), New Haven County (pt.)

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
3283	HART	CT	060, 061	3282 (excluding 5020)	CMSA, (Excludg a PMSA)	Includes the PMSAs of Hartford (3280) Bristol (1170), and New Britain (5440). Middletown (5020) which is also in this CMSA has a ZIP code of 064 which is in NHAV	*Bristol, CT PMSA* 09003 Hartford County (pt.) 09005 Litchfield County (pt.) *Hartford, CT PMSA* 09003 Hartford County (pt.) 09005 Litchfield County (pt.) 09007 Middlesex County (pt.) 09011 New London County (pt.) 09013 Tolland County (pt.) *New Britain, CT PMSA* 09003 Hartford County (pt.)
5483	NHAV	CT	064, 065	5480	MSA	Only a small portion of this MSA has a ZIP of 064	*New Haven, CT MSA* 09007 Middlesex County (pt.) 09009 New Haven County (pt.) *Middletown, CT PMSA* 09007 Middlesex County (pt.)
8840	WASH	DC	200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 220, 221, 222, 223	8840	MSA	The county of Frederick MD is in the MSA but has a ZIP code of 217.	11001 District of Columbia 24009 Calvert County, MD 24017 Charles County, MD 24021 Frederick County, MD 24031 Montgomery County, MD 24033 Prince George's County, MD 51510 Alexandria city, VA 51013 Arlington County, VA 51059 Fairfax County, VA 51600 Fairfax city, VA 51610 Falls Church city, VA 51107 Loudoun County, VA 51683 Manassas city, VA 51685 Manassas Park city, VA 51153 Prince William County, VA 51179 Stafford County, VA

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
9160	WILM	DE	197, 198, 219	9160	PMSA	The New Jersey portion has a ZIP code of 080 which is not included.	1003 New Castle County, DE 24015 Cecil County, MD 34033 Salem County, NJ
2680	FLAU	FL	330, 331, 332, 333	4992	CMSA	These ZIP codes apply to the PMSAs: Fort Lauderdale-Hollywood-Pompano Beach (2680) and Miami-Hialeah (5000)	*Fort Lauderdale-Hollywood-Pompano Beach, FL PMSA* 12011 Broward County *Miami-Hialeah, FL PMSA* 12025 Dade County
3600	JACK	FL	320, 322	3600	MSA		12019 Clay County 12031 Duval County 12089 Nassau County 12109 St. Johns County
5960	ORLA	FL	327, 328	5960	MSA	Osceola County is in this MSA but has a ZIP code of 347.	12095 Orange County 12117 Seminole County 12097 Osceola County
8280	TAMP	FL	335, 336, 337, 342, 346	8280, 1140 & 7510	MSA	Bradenton (1140) and Sarasota (7510) have ZIPs of 342 and are their own MSAs.	*Tampa-St. Petersburg-Clearwater, FL MSA* 12053 Hernando County 12057 Hillsborough County 12101 Pasco County 12103 Pinellas County *Bradenton, FL MSA* 12081 Manatee County *Sarasota, FL MSA* 12115 Sarasota County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
520	ATLA	GA	300, 301, 302, 303, 311, 399	520	MSA		13031 Barrow County 13035 Butts County 13057 Cherokee County 13063 Clayton County 13067 Cobb County 13077 Coweta County 13089 De Kalb County 13097 Douglas County 13113 Fayette County 13117 Forsyth County 13121 Fulton County 13135 Gwinnett County 13151 Henry County 13217 Newton County 13223 Paulding County 13247 Rockdale County 13255 Spalding County 13297 Walton County
1800	COLU	GA	318, 319, 368	1800	MSA		01113 Russell County, AL 13053 Chattahoochee County, GA 13215 Muscogee County, GA
1080	BCIT	ID	837	1080	MSA		16001 Ada County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
1600	CHIC	IL	600, 601, 602, 603, 604, 605, 606.	1602 (excl u-ding 2920 and 3800)	CMSA (exclu-ding PMSAs)	The CMSA includes the PMSAs: Aurora-Elgin (0620), Chicago (1600) which also includes the ZIP codes 607 and 608 (which may have replace 602 and 603), Gary-Hammond (2960) which is a separate city in the ACS study, Joliet (3690), Kenosha (3800), Lake County (3965). However, Gary-mmond and Kenosha are included as separate places in the ACS study (see below).	*Chicago, IL PMSA* 17031 Cook County 17043 Du Page County 17111 McHenry County *Aurora-Elgin, IL PMSA* 17089 Kane County 17093 Kendall County *Joliet, IL PMSA* 17063 Grundy County 17197 Will County *Lake County, IL PMSA* 17097 Lake County
1020	BLOO	IN	474	1020	MSA		18105 Monroe County
2440	EVAN	IN	424, 476, 477	2440	MSA		18129 Posey County, IN 18163 Vanderburgh County, IN 18173 Warrick County, IN 21101 Henderson County, KY
2960	GARY	IN	463, 464	2960	PMSA	Part of the CMSA of Chicago	18089 Lake County 18127 Porter County
3480	INDI	IN	460, 461, 462	3480 & 0400	MSAs	The adjacent MSA of Anderson (0400) includes the ZIP 460	*Indianapolis, IN MSA* 18011 Boone County 18057 Hamilton County 18059 Hancock County 18063 Hendricks County 18081 Johnson County 18097 Marion County 18109 Morgan County 18145 Shelby County *Anderson, IN MSA* 18095 Madison County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
7800	SBEN	IN	465, 466	7800 & 2330	MSAs	The adjacent MSA of Elkhart-Goshen (2330) also has a ZIP of 465	*South Bend, IN MSA* 18141 St. Joseph County *Elkhart-Goshen, IN MSA* 18039 Elkhart County
8320	THAU	IN	478	8320	MSA		18021 Clay County 18167 Vigo County
1360	CRAP	IO	522, 523, 524	1360 & 3500	MSAs	The adjacent MSA of Iowa City (3500) also has the ZIP codes 522 and 523. The state abbreviation should be IA	*Cedar Rapids, IO MSA* 19113 Linn COunty *Iowa City, IA MSA* 19103 Johnson County
2120	DMOI	IO	503	2120	MSA	ZIP code 509 is also in Des Moines (Polk County). The state abbreviation should be IA	19049 Dallas County 19153 Polk County 19181 Warren County
2200	DUBU	IO	520	2200	MSA	The state abbreviation should be IA	19061 Dubuque County
8920	WATE	IO	507	8920	MSA	The state abbreviation should be IA	19013 Black Hawk County 19017 Bremer County
8440	TOPE	KS	664, 665, 666	8440	MSA		20177 Shawnee County
9040	WICH	KS	670, 671, 672	9040	MSA		20015 Butler County 20173 Sedgewick County
4280	LEXI	KY	403, 404, 405	4280	MSA		21017 Bourbon County 21049 Clark County 21067 Fayette County 21113 Jessamine County 21209 Scott County 21239 Woodford County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
760	BATO	LA	707, 708	760	MSA		22005 Ascension Parish 22033 East Baton Rouge Parish 22063 Livingston Parish 22121 West Baton Rouge Parish
5560	NORL	LA	700, 701	5560	MSA	St. Tammany county is in this MSA but has a ZIP code of 704	22051 Jefferson Parish 22071 Orleans Parish 22087 St Bernard Parish 22089 St. Charles Parish 22095 St. John the Baptist Parish
7680	SHRE	LA	711	7680	MSA	The majority of the county of Caddo which is in this MSA has a ZIP code of 710	22015 Bossier Parish 22017 Caddo Parish
1123	BOST	MA	017, 020, 021, 022	1120	PMSA	The SMSA number "1123" does not exist in the 1983 definitions, however the ZIP codes appear to refer to PMSA 1120.	25005 Bristol County (pt.) 25009 Essex County (pt.) 25017 Middlesex County (pt.) 25021 Norfolk County (pt.) 25023 Plymouth County (pt.) 25025 Suffolk County (pt.) 25027 Worcester County (pt.)
5403	NBED	MA	25	5400	MSA	Now has the ZIP code 027. The ZIP code 025 appears to apply to part of Plymouth cty, Barnstable Cty and Dukes Cty.	*New Bedford, MA MSA* 25005 Briston County (pt.) 25023 Plymouth County (pt.) *Rural Counties in MA* 25 001 Barnstable County 25 007 Dukes County
8003	SPRI	MA	010, 011	8000	MSA	The SMSA number "8003" does not exist in the 1983 definitions, however the ZIP codes appear to refer to PMSA 8000.	25013 Hampden County (pt.) 25015 Hampshire County (pt.)

Original Analysis				Re-analysis				
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis	
9243	WORC	MA	015, 016	9240	MSA	The SMSA number "9243" does not exist in the 1983 definitions, however the ZIP codes appear to refer to PMSA 9240.	25027	Worcester County (pt.)
720	BALT	MD	210, 211, 212	720	MSA		24003	Anne Arundel County
							24510	Baltimore city
							24005	Baltimore County
							24013	Carroll County
							24025	Harford County
							24027	Howard County
							24035	Queen Anne's County
733	BANG	ME	44	730	MSA	The SMSA number "0733" does not exist in the 1983 definitions, however the ZIP codes appear to refer to PMSA 0730.	23019	Penobscot County (pt.)
							23027	Waldo County
4243	LEWI	ME	42	4240	MSA	The SMSA number "1123" does not exist in the 1983 definitions, however the ZIP codes appear to refer to PMSA 1120.	23001	Androscoggin County (pt.)
6403	PORT	ME	041, 040	6400	MSA	The SMSA number "6403" does not exist in the 1983 definitions, however the ZIP codes appear to refer to PMSA 6400.	23005	Cumberland County (pt.)
							23031	York County (pt.)

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
2160	DETR	MI	480, 481, 482, 483	2162	CMSA	Includes the PMSAs: Ann Arbor (0440) and Detroit (2160).	*Detroit, MI PMSA* 26087 Lapeer County 26093 Livingston County 26099 Macomb County 26115 Monroe County 26125 Oakland County 26147 St. Clair County 26163 Wayne County *Ann Arbor, MI PMSA* 26161 Washtenaw County
2640	FLIN	MI	484, 485	2640	MSA		26049 Genesee County
4040	LANS	MI	488, 489	4040	MSA		26037 Clinton County 26045 Eaton County 26065 Ingham County
6960	SAGI	MI	486, 487	6960	MSA		26017 Bay County 26111 Midland County 26145 Saginaw County
2240	DULU	MN	586, 587, 588	2240	MSA	Douglas County, WI is part of this MSA but has the ZIP code 548. The ZIP code 586 appears to pertain to Slope, Hettinger, Bowman, Adams, Billings, Golden Valley and Stark counties in south west ND	27137 St. Louis County, MN 55031 Douglas County, WI

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
5120	MINN	MN	550, 551, 553, 554	5120	MSA		27003 Anoka County, MN 27019 Carver County, MN 27025 Chisago County, MN 27037 Dakota County, MN 27053 Hennepin County, MN 27059 Isanti County, MN 27123 Ramsey County, MN 27139 Scott County, MN 27163 Washington County, MN 27171 Wright County, MN 55109 St. Croix County, WI
3760	KANS	MO	640, 641, 660, 661	3762	CMSA	Includes the PMSAs: Kansas City, KS (3755) and Kansas City, MO (3760).	*Kansas City, MO PMSA* 29037 Cass County, MO 29047 Clay County, MO 29095 Jackson County, MO 29107 Lafayette County, MO 29165 Platte County, MO 29177 Ray County, MO *Kansas City, KS PMSA* 20091 Johnson County, KS 20103 Leavenworth County, KS 20121 Miami County, KS 20209 Wyandotte County, KS

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
7040	SLOU	MO	620, 622, 630, 631	7042	CMSA	Includes the PMSAs: Alton-Granite City (0275); East St. Louis-Belleville (2285); and St. Louis (7040).	*St. Louis, MO PMSA* 17133 Monroe County, IL 29071 Franklin County, MO 29099 Jefferson County, MO 29183 St. Charles County, MO 29189 St. Louis County, MO 29510 St. Louis city, MO *Alton-Granite City, IL PMSA* 17083 Jersey County, IL 17119 Madison County, IL *East St. Louis-Belleville, IL PMSA* 17027 Clinton County, IL 17163 St. Clair County, IL
3560	JACK	MS	390, 391, 392	3560	MSA		28049 Hinds County 28089 Madison County 28121 Rankin County
880	BILL	MT	590, 591	880	MSA		30111 Yellowstone County
3040	GFAL	MT	594	3040	MSA		30013 Cascade County
4360	LINC	NB	683, 684, 685	4360	MSA	The state abbreviation should be NE	31109 Lancaster County
5920	OMAH	NB	680, 681, 515	5920	MSA	The state abbreviation should be NE	19155 Pottawattamie County, IN 31055 Douglas County, NE 31153 Sarpy County, NE 31177 Washington County, NE
1520	CHAR	NC	280, 281, 282, 297	1520	MSA		37025 Cabarrus County, NC 37071 Gaston County, NC 37109 Lincoln County, NC 37119 Mecklenburg County, NC 37159 Rowan County, NC 37179 Union County, NC 45091 York County, SC

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
3120	GREE	NC	270, 271, 272, 273, 274	3120 & 1300	MSA	The majority of the adjacent MSA of Burlington (1300) also has the ZIP code 272.	*Greensboro-Winston-Salem-High Point, NC MSA* 37057 Davidson County 37059 Davie County 37067 Forsyth County 37081 Guilford County 37151 Randolph County 37169 Stokes County 37197 Yadkin County *Burlington NC MSA* 37001 Alamance County
6640	RALE	NC	275, 276, 277	6640	MSA		37063 Durham County 37069 Franklin County 37135 Orange County 37183 Wake County
1010	BISM	ND	585	1010	MSA		38015 Burleigh County 38059 Morton County
2985	GFOR	ND	582	2985	MSA		38035 Grand Forks County
2520	FARG	ND	580, 581	2520	MSA	Clay county MN is in the MSA but has a ZIP code of 565.	27027 Clay County, MN 38017 Cass County, ND
4763	MANC	NH	030, 031	4760 & 5350	MSA	The SMSA number "4763" does not exist in the 1983 definitions, however the ZIP codes appear to refer to PMSA 4760. The adjacent MSA of Nashua also has the ZIP code 031.	*Manchester, NH MSA* 33011 Hillsborough County (pt.) 33013 Merrimack County (pt.) 33015 Rockingham County (pt.) *Nashua, NH PMSA* 33011 Hillsborough County 33015 Rockingham County (pt.)

Original Analysis				Re-analysis				
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis	
6453	PORT	NH	038, 039	6450	MSA	The SMSA number "6453" does not exist in the 1983 definitions, however the ZIP codes appear to refer to PMSA 6450.	23031 33015 33017	York County, ME (pt.) Rockingham County, NH (pt.) Stafford County, NH (pt.)
3640	JCIT	NJ	073, 076	3640	PMSA	The ZIP code 070 is also contained in this PMSA; 076 is in Bergen-Passaic PMSA (see PATE)	34017	Hudson County
5640	NEWA	NJ	070, 071, 072, 078, 079	5640	PMSA		34013 34027 34037 34039	Essex County Morris County Sussex County Union County
6040	PATE	NJ	075, 074	875	PMSA	Bergen-Passaic PMSA; Several cities in Bergen County have 076 ZIP codes	34003 34031	Bergen County Passaic County
8480	TREN	NJ	085, 086	8480	PMSA		34021	Mercer County
200	ALBU	NM	870, 871, 872	200	MSA		35001	Bernalillo County
4120	LVEG	NV	889, 890, 891	4120	MSA		32003	Clark County
6720	RENO	NV	894, 895	6720	MSA		32031	Washoe County
160	ALBA	NY	120, 121, 122, 123	160	MSA		36001 36039 36057 36083 36091 36093	Albany County Greene County Montgomery County Rensselaer County Saratoga County Schenectady County
1280	BUFF	NY	140, 141, 142, 143	1280	PMSA		36029	Erie County
2335	ELMI	NY	148, 149	2335	MSA		36015	Chemung County

Original Analysis				Re-analysis				
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis	
5380	NASS	NY	115, 117, 118, 119	5380	PMSA		36059 36103	Nassau County Suffolk County
5600	NYOR	NY	100, 101, 102, 103, 104, 105, 106, 107, 108, 110, 111, 112, 113, 114, 116	5600	PMSA	Rockland County is also in the CMSA but has the ZIP code 109	36005 36047 36061 36079 36081 36085 36087 36119	Bronx County Kings County New York County Putnam County Queens County Richmond County Rockland County Westchester County
6460	POUG	NY	125, 126	6460	MSA		36027	Dutchess County
6840	ROCH	NY	144, 145, 146	6840	MSA		36051 36055 36069 36073 36117	Livingston County Monroe County Ontario County Orleans County Wayne County
8160	SYRA	NY	130, 131, 132	8160	MSA		36053 36067 36075	Madison County Onondaga County Oswego County
8680	UTIC	NY	133, 134, 135	8680	MSA		36043 36065	Herkimer County Oneida County
80	AKRO	OH	442, 443	80	PMSA		39133 39153	Portage County Summit County
1320	CANT	OH	446, 447	1320	MSA		39019 39151	Carroll County Stark County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
1640	CINC	OH	451, 452, 470, 410	1640	PMSA		18029 Dearborn County, IN 21015 Boone County, KY 21037 Campbell County, KY 21117 Kenton County, KY 39025 Clermont County, OH 39061 Hamilton County, OH 39165 Warren County, OH
1680	CLEV	OH	440, 441	1680 & 4440	PMSAs	The ZIP code 440 includes Lorain County in adjacent PMSA of Lorain-yria (4440).	*Cleveland, OH PMSA* 39035 Cuyahoga County 39055 Geauga County 39085 Lake County 39103 Medina County *Lorain-Elyria, OH PMSA* 39093 Lorain County
1840	COLU	OH	430, 431, 432	1840	MSA		39041 Delaware County 39045 Fairfield County 39049 Franklin County 39089 Licking County 39097 Madison County 39129 Pickaway County 39159 Union County
2000	DAYT	OH	453, 454, 455	2000	MSA		39023 Clark County 39057 Greene County 39109 Miami County 39113 Montgomery County
3200	HAMI	OH	450	3200	PMSA		39017 Butler County
4800	MANS	OH	448, 449	4800	MSA		39139 Richland County
	SPRI	OH				Not included in the study since no ZIP codes were assigned.	

Original Analysis				Re-analysis				
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis	
8080	STEU	OH	439, 260	8080	MSA		39081	Jefferson County, OH
							54009	Brooke County, WV
							54029	Hancock County, WV
8400	TOLE	OH	434, 435, 436	8400	MSA		39051	Fulton County
							39095	Lucas County
							39173	Wood County
9320	YOUN	OH	444, 445	9320	MSA		39099	Mahoning County
							39155	Trumbull County
5880	OKLA	OK	730, 731	5880	MSA	Pottawatomie county is in this MSA but has a ZIP code of 748.	40017	Canadian County
							40027	Cleveland County
							40083	Logan County
							40087	McClain County
							40109	Oklahoma County
							40125	Pottawatomie County
6440	PORT	OR	970, 971, 972	6440	PMSA		41005	Clackamas County
							41051	Multnomah County
							41067	Washington County
							41071	Yamhill County
240	ALLE	PA	180, 181, 182, 183	240	MSA		34041	Warren County, NJ
							42025	Carbon County, PA
							42077	Lehigh County, PA
							42095	Northampton County, PA
280	ALTO	PA	166	280	MSA		42013	Blair County
2360	ERIE	PA	164, 165	2360	MSA		42049	Erie County
3240	HARR	PA	170, 171	3240	MSA		42041	Cumberland County
							42043	Dauphin County
							42075	Lebanon County
							42099	Perry County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
3680	JOHN	PA	155, 159	3680	MSA		42021 Cambrian County 42111 Somerset County
4000	LANC	PA	173, 175, 176	4000	MSA		42071 Lancaster County
5745	NORT	PA	184, 185	42069	County	There is no MA called Northampton, PA; The MSA Scranton-Wilkes-Barre includes the ZIP codes 184 and 185 (both in Lackawanna county) but also includes 178 and 182, 183, and 187	42069 Lackawanna County
6160	PHIL	PA	189, 190, 191, 193, 194	6160	PMSA	The cities on the NJ side of this PMSA have the ZIP codes 080 and 081 which are not included here.	34005 Burlington County, NJ 34007 Camden County, NJ 34015 Gloucester County, NJ 42017 Bucks County, PA 42029 Chester County, PA 42045 Delaware County, PA 42091 Montgomery County, PA 42101 Philadelphia County, PA
6280	PITT	PA	150, 151, 152, 153, 154, 156	6282	CMSA	Includes the PMSAs: Beaver County (0845) with a ZIP code of 150 and Pittsburgh (6280).	*Pittsburgh, PA PMSA* 42003 Allegheny County 42051 Fayette County 42125 Washington County 42129 Westmoreland County *Beaver County, PA PMSA* 42007 Beaver County
6680	READ	PA	195, 196	6680	MSA		42011 Berks County
8050	SCOL	PA	168	8050	MSA		42027 Center County
7610	SHAR	PA	161	7610	MSA		42085 Mercer County
9140	WILL	PA	177	9140	MSA		42081 Lycoming County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
9280	YORK	PA	174	9280	MSA		42001 Adams County 42133 York County
6483	PROV	RI	027, 028, 029	6482	CMSA	The SMSA number "6483" does not exist in the 1983 definitions, however the ZIP codes appear to refer to PMSA 6480. Includes the PMSAs: Providence (6480: ZIP 028 and 029); Pawtucket-Woonsocket-Attleboro (6060: ZIP 027); and Fall River (2480: ZIP 027)	*Providence, RI PMSA* 44001 Bristol County (pt.) 44003 Kent County (pt.) 44005 Newport County (pt.) 44007 Providence County (pt.) 44009 Washington County (pt.) *Fall River, MA-RI PMSA* 25005 Bristol County, MA (pt.) 44005 Newport County, RI (pt.) *Pawtucket-Woonsocket-Attleboro, RI-MA PMSA* 25005 Bristol County, MA (pt.) 25021 Norfolk County, MA (pt.) 25027 Worcester County, MA (pt.) 44007 Providence County, RI (pt.)
1440	CHAR	SC	294	1440	MSA		45015 Berkeley County 45019 Charleston County 45035 Dorchester County
1760	COLU	SC	290, 291, 292	1760	MSA		45063 Lexington County 45079 Richland County
3160	GREE	SC	296	3160 & 0405	MSAs	Spartanburg County is also in this MSA but has a ZIP of 293. The ZIP code 296 includes the adjacent MSA of Anderson (0405).	*Greenville, SC MSA* 45045 Greenville County 45077 Pickens County 45083 Spartanburg County *Anderson, SC MSA* 45007 Anderson County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
1560	CHAT	TN	373, 374, 307	1560	MSA		13047 Catoosa County, GA 13083 Dade County, GA 13295 Walker County, GA 47065 Hamilton County, TN 47115 Marion County, TN 47153 Sequatchie County, TN
3840	KNOX	TN	377, 378, 379	3840	MSA		47001 Anderson County 47009 Blount County 47057 Grainger County 47089 Jefferson County 47093 Knox County 47155 Sevier County 47173 Union County
4920	MEMP	TN	380, 381, 386, 723	4920	MSA		05035 Crittenden County, AR 28033 De Soto County, MS 47157 Shelby County, TN 47167 Tipton County, TN
5360	NASH	TN	370, 371, 372	5360	MSA	ZIP codes 370 and 371 include Montgomery County in the adjacent MSA of Clarksville-Hopkinsville (1660) but the other city in that MSA, Hopkinsville KY has a ZIP code of 422.	*Nashville, TN MSA* 47021 Cheatham County 47037 Davidson County 47043 Dickson County 47147 Robertson County 47149 Rutherford County 47165 Sumner County 47187 Williamson County 47189 Wilson County *Clarksville-Hopkinsville, TN-KY MSA* 47125 Montgomery County, TN
840	BPAR	TX	776, 777	840	MSA		48199 Hardin County 48245 Jefferson County 48361 Orange County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
1880	CCHR	TX	783, 784	1880	MSA		48355 Nueces County 48409 San Patricio County
1920	DALL	TX	750, 751, 752, 753, 760, 761	1922 & 7640	CMSA & MSA	Denton county is in this CMSA but its ZIP code is 762. The majority of the adjacent MSA of Sherman-Denison (7640) also has the ZIP code 750. This CMSA includes the PMSAs: Dallas (1920) and Fort Worth-rlington (2800)	*Dallas, TX PMSA* 48085 Collin County 48113 Dallas County 48139 Ellis County 48257 Kaufman County 48397 Rockwall County *Fort Worth-Arlington, TX PMSA* 48251 Johnson County 48367 Parker County 48439 Tarrant County *Sherman-Denison, TX MSA* 48181 Grayson County
2320	EPAS	TX	798, 799	2320	MSA	The ZIP code 885 is also used for El Paso but was left out in this study.	48141 El Paso County
2920	GALV	TX	775	2920 & 1145	PMSAs	The ZIP code covers the PMSA Galveston-Texas City (2920) as well as the PMSA Brazoria (1145). It does not include the other PMSA in this CMSA, Houston	*Galveston-Texas City, TX PMSA* 48167 Galveston County *Brazoria, TX PMSA* 48039 Brazoria County
3360	HOUS	TX	770, 771, 772	3360	PMSA	The following counties are in the PMSA but are missing ZIP codes: Montgomery (773), Fort Bend and Waller (774). Most of Liberty Cty and part of Harris County are included in GALV (775)	48157 Fort Bend County 48201 Harris County 48291 Liberty County 48339 Montgomery County 48473 Waller County
4600	LUBB	TX	793, 794	4600	MSA		48303 Lubbock County
7200	SANG	TX	769	7200	MSA		48451 Tom Green County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
7240	SANT	TX	780, 781, 782	7240	MSA		48029 Bexar County 48091 Comal County 48187 Guadalupe County
8800	WACO	TX	766, 767	8800	MSA		48309 McLennan County
9080	WICH	TX	763	9080	MSA		48485 Wichita County
7160	SLAK	UT	840, 841, 842	7160	MSA		49011 Davis County 49035 Salt Lake County 49057 Weber County
1950	DANV	VA	245	1950 & 4640	MSAs	This ZIP code covers the adjacent MSAs of Danville (1950) and Lynchburg (4640), as well as Halifax County not included in an MSA.	*Danville, VA MSA* 51590 Danville city 51143 Pittsylvania County *Lynchburg, VA MSA* 51009 Amherst County 51031 Campbell County 51680 Lynchburg city *Rural, VA* 51 083 Halifax
5720	NORF	VA	233, 234, 235, 236	5720	MSA		51550 Chesapeake city 51073 Gloucester County 51650 Hampton city 51095 James City County 51700 Newport News city 51710 Norfolk city 51735 Poquoson city 51740 Portsmouth city 51800 Suffolk city 51810 Virginia Beach city 51830 Williamsburg city 51199 York County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
6760	RICH	VA	230, 231, 232	6760	MSA	Dinwiddie, Prince George and Chesterfield Counties in this MSA all have the ZIP code 238 which is not included.	51036 Charles City County 51041 Chesterfield County 51570 Colonial Heights city 51053 Dinwiddie County 51075 Goochland County 51085 Hanover County 51087 Henrico County 51670 Hopewell city 51127 New Kent County 51730 Petersburg city 51145 Powhatan County 51149 Prince George County 51760 Richmond city
6800	ROAN	VA	240, 241	6800	MSA		51023 Botetourt County 51161 Roanoke County 51770 Roanoke city 51775 Salem city
7600	SEAT	WA	980, 981	7600	PMSA	Everett and Snohomish which is also in this MSA has a ZIP of 982, other places in Snohomish County are accounted for.	53033 King County 53061 Snohomish County
7840	SPOK	WA	990, 991, 992	7840	MSA		53063 Spokane County
8200	TACO	WA	983, 984	8200 & 1150	PMSA & MSA	The ZIP code 983 includes both places within Tacoma PMSA as well as the adjacent MSA of Bremerton.	*Tacoma, WA PMSA* 53053 Pierce County *Bremerton, WA MSA* 53035 Kitsap County
2290	ECLA	WI	547	2290	MSA		55017 Chippewa County 55035 Eau Claire County
3800	KENO	WI	531	3800	PMSA	This PMSA is part of the CMSA of Chicago-Gary-Lake County	55059 Kenosha County

Original Analysis				Re-analysis			
SMSA	CITY	ST	ZIP	MA	Level	Comments	Counties Included in Sensitivity Analysis
4720	MADI	WI	535, 537	4720 & 3620	MSAs	The ZIP code 535 includes both Madison (4720) and the adjacent MSA of Janesville-Beloit (3620).	*Madison, WI MSA* 55025 Dane County *Janesville-Beloit, WI MSA* 55105 Rock County
5080	MILW	WI	530, 532	5080 & 7620	PMSA & MSA	The ZIP code 530 includes both the PMSA of Milwaukee and the adjacent MSA of Sheboygan.	*Milwaukee, WI PMSA* 55079 Milwaukee County 55089 Ozaukee County 55131 Washington County 55133 Waukesha County *Sheboygan, WI MSA* 55117 Sheboygan County
6600	RACI	WI	534	6600	PMSA		55101 Racine County
1480	CHAR	WV	250, 251, 252, 253	1480	MSA	ZIP code 252 appears to only refer to counties Jackson and Roane, outside of the MSA.	*Charleston, WV MSA* 54039 Kanawha County 54079 Putnam County, Rural Counties 54 035 Jackson County 54 067 Roane County
3400	HUNT	WV	255, 256, 257, 411, 412	3400	MSA	ZIP code 412 appears to only refer to KY counties Lawrence, Johnson and Martin which are all outside of the MSA.	21019 Boyd County, KY 21043 Carter County, KY 21089 Greenup County, KY 39087 Lawrence County, OH 54011 Cabell County, WV 54099 Wayne County, WV Rural Counties in KY 21 127 Lawrence, KY 21 115 Johnson, KY 21 159 Martin, KY