Excerpt from

Neighborhoods and Health:

Building Evidence for Local Policy

By Kathryn L.S. Pettit, G. Thomas Kingsley, and Claudia J. Coulton
With Jessica Cigna

May 2003

SUBMITTED TO
The Office of the Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services

SUBMITTED BY
The Urban Institute
2100 M Street NW, Washington DC

Delivery Order 19, Contract No. HHS-100-99-0003.

Part 2

Cross-Site Analysis

Section 7

APPROACH AND METHODOLOGY

This section begins with a description of the framework, drawn from relevant literature, that drives the work of our cross-site analysis. It notes what research of this kind can and cannot accomplish. It then describes the way we have structured the types of neighborhood conditions that we suspect influence health outcomes and, accordingly, the research hypotheses we will test. The section next describes the specific data—of different types and at different geographic levels—which we utilize in the research. Finally it presents the hypotheses themselves and reviews our methods of analysis in more detail. Specifics on analytic methods (bivariate and multivariate analysis) will be discussed in section 10.

FRAMEWORK

Specific Purposes

The notion than conditions of the social and physical environment in urban neighborhoods affect the lives of the residents has been the subject of speculation and scholarly research since the 1920s (Burgess 1925; Park 1929, 1936). This tradition of ecological research has demonstrated that neighborhood conditions matter, but clearly they are not the only things that matter.

Socioeconomic characteristics and behaviors of individuals and families, for example, also have an extremely important effect on their well-being, whatever their neighborhood of residence. In fact, the field offers stern warnings to researchers to avoid the “ecological fallacy” (i.e., concluding because there is a close statistical relationship between some neighborhood conditions and changing outcomes that the one is the cause of the other, without recognizing the role of other variables (e.g., family characteristics) that are not included directly in the analysis). Nonetheless, it is generally conceded that neighborhood conditions do have some effects independent of the influence of individual or family characteristics (Ellen and Turner 1997).

The limitations involved in obtaining individual and family level data are an important overall constraint on what this research could accomplish. Administrative data files with health indicators, such as those maintained by the NNIP partners, seldom contain (or at least permit agencies to release) detailed descriptive information on individuals and their families. Without such data, we cannot perform the type of multilevel analysis needed to explain more fully relationships between changes in neighborhood-level variables and health outcomes.

Still, there are other purposes of studying these relationships, and they can be of substantial practical value.[17] Ecological analyses can be used effectively in identifying potential problems and developing hypotheses about what is causing them, if not for drawing definitive conclusions on causation. Most basically, health agencies need to know about trends in the extent to which (and where) health problems are spatially concentrated if they are to develop efficient strategies for prevention and care. Knowing trends in the relationships between specific neighborhood conditions and health outcomes can provide valuable hints as to likely changes to spatial patterns in the future and, thereby, to specific programmatic opportunities those shifts may imply.

Categories: Overview

In this analysis, we have grouped the conditions likely to influence health outcomes (our independent variables) into three basic categories, as suggested by our review of the literature. First, we divide neighborhood-level conditions into two groups: *socioeconomic conditions* (e.g., income and poverty levels) that indirectly influence health, and *direct pathways* that can have more direct effects (e.g., environmental hazards or crime).

Neighborhood Conditions: Socio-economic Factors

As noted, our first set of indicators tries to capture the demographic and socioeconomic status characteristics of the neighborhood, and there is little doubt of important associations with health. In the most comprehensive review of the literature we identified, Ellen, Mijanovich, and Dillman (2001) recognize a broad consensus that “residents of socially and economically deprived communities experience worse health outcomes on average than those living in more prosperous areas . . . suffer from higher rates of heart disease, respiratory ailments and overall mortality.” Yet more seriously, their review suggests to them that:

---

[17] For more complete discussion of the appropriate uses of descriptive studies in the health field, see Grimes and Schulz 2002.

> "…neighborhoods may primarily influence health in two ways: first, through relatively short-term influences on behaviors, attitudes, and health care utilization, thereby affecting health conditions that are most immediately responsive to such influences; and second, through a longer term process of "weathering," whereby the accumulated stress, lower environmental quality, and limited resources of poorer communities, experienced over many years, erodes the health of residents in ways that make them more vulnerable to mortality to any given disease (Geronimus 1992)."

There have been numerous studies on linkages between neighborhood socioeconomic conditions and varying types of health outcomes. Researchers have found correlations between socioeconomic status (SES) and indicators of health-related behavior, such as smoking (Kleinschmidt, Hills, and Elliot 1995) and physical activity and body-mass index (Robert 1999). Relationships between low-SES neighborhoods and mental health problems have been examined by Aneshensel and Sucoff (1996)—a study of adolescents in Los Angeles County —and Katz, Kling, and Liebman (2000)—psychological benefits of moving to better neighborhoods in the Moving to Opportunity Program.

Perhaps the most work of this type has been done on associations between neighborhood socioeconomic characteristics and birth-related outcomes, including low-birth weight births (Collins and David 1990; Duncan and Laren 1990; O'Campo et al. 1997) and infant mortality (Collins and David 1992; Coulton and Pandy 1992; Guest, Almgren, and Hussey 1998). Research has also evidenced the association between distressed neighborhood conditions and lower early prenatal care rates among African-American mothers (Perloff and Jaffee 1999). Finally, Robert (1998) and Marmot et al. (1998) have found relationships between lower neighborhood SES measures and more chronic health problems (self-rated) of adults.

For this research, we have subdivided indicators in this broader category into three subgroups. The first is ***demographic***, covering data on race and age composition. The second, and probably most significant, is ***economic***. To represent the economic circumstances in a neighborhood, we use several variables, including average household income, the unemployment rate, and the overall poverty rate. The third subgroup includes measures related to ***social*** risks, such as the share of adults with no high school degree, the share of households receiving public assistance, and the share of families with children that are headed by females.

Neighborhood Conditions: Direct Pathways

Ellen, Mijanovich, and Dillman (2001) identify four pathways through which neighborhood can affect health more directly: (1) neighborhood institutions and resources; (2)

stresses in the physical environment; (3) stresses in the social environment; and (4) neighborhood-based networks and norms.

The first of these refers to the availability of neighborhood institutions and assets, including health care facilities, grocery stores, and reliable transit, which can affect individuals' ability to access health care, healthy food, or other health supports in their neighborhood or elsewhere. The second pathway refers to physical conditions or contaminants that directly affect people's health, such as environmental hazards or lead paint. The third touches on the mental or psychological stress from neighborhood conditions such as crime. The last refers to the social networks, support systems, and community expectations of a neighborhood that can help promote healthy behaviors.

The contextual data available to us cannot measure all of these mechanisms directly, but they can act as imperfect proxies for the underlying processes for three of them. For the first of the pathways identified, ***neighborhood institutions and resources***, we do not have data from any source that we believe can serve as adequate proxies.

However, we can access a set of indicators on housing quality that should act as a proxy for ***physical stressors*** in the neighborhood, dealing with the age of housing (older housing is more likely to have problems with lead paint, poor heating and plumbing systems, and run-down structures), the extent of overcrowding (overcrowded housing has been associated with less sanitary conditions and the spread of disease), and measures related to home values. Use of the latter is responsive to hedonic price modeling theory, which says that a home price represents a bundle of characteristics that have a particular value in the housing market. We include average home values and values of home purchase mortgage loans as measures of overall quality of the housing in the neighborhood.

Next, we use data on crime rates to represent ***social stressors*** in residents' lives. Data from local systems are available on total, violent, and property crime rates in the five cities. Based on the work of Zapata et al. (1992) and others, we would expect the violent crime rate to be a greater stressor than property crime and thus to have a stronger correlation with health outcomes.

The final pathway, ***neighborhood-based networks and norms***, is difficult to quantify. To investigate proxies for this concept, we selected several indicators of turnover and mobility within tracts. They are based on the assumption that in areas where people move around a lot, there is less opportunity to develop meaningful connections with neighbors or be integrated into neighborhood social networks. In this analysis, we used higher rates of renter-occupied housing, vacancies, and home purchase mortgages as proxies for weaker social ties. They will also have a greater proportion of people who lived in a different house in 1995, and a rapidly changing (growing or declining) population. Finally, we suggest that a greater number of home improvement mortgages is a positive sign of stability, indicating households who are committed to staying in their home and area.

Putting all of this together, we will be reporting on relationships between health indicators and independent variables in five major categories: (1) socioeconomic conditions; (2) physical stressors; (3) social stressors; and (4) social networks

DATA SOURCES AND DEFINITIONS

We now discuss the data assembled for this analysis. Basically, there were two types of sources: (1) the data systems operated by the participating local partners in NNIP and (2) national datasets with information at the census tract level. In our discussion of the former, we distinguish between data on health outcomes and those on contextual variables. We also note two types of national datasets: decennial census data, as presented in the Neighborhood Change Database (NCDB), and the Home Mortgage Disclosure Act data files.

Geographic Definitions

Most of the analysis in this report pertains to the “central city/county” in each study site; the central area within each metropolis for which we were able to obtain health-related data (see discussion below). For one of our sites, this was the central county, Cuyahoga County in metropolitan Cleveland. For the others, we use central city boundaries as our reference area, although it should be noted that the central cities are also counties in two of the sites: Denver County in metropolitan Denver and Marion County in metropolitan Indianapolis.

In some sections, we also present data for the metropolitan areas in full: the central city/county plus a number of additional surrounding counties. In these cases we use the 2000 boundaries of the Primary Metropolitan Statistical Area or the Metropolitan Statistical Area as defined by the federal government (Office of Management and Budget).

Data Systems Operated by Local NNIP Partners

Before we discuss individual variables, it should be helpful to say a few things about the data systems operated by NNIP’s local partners in general. All 20 of them have built or are building advanced GIS information systems with integrated, recurrently updated information on neighborhood conditions in their cities. This is a capacity that did not exist in any U.S. city in the 1980s. The breakthrough became possible because (1) most administrative records of government agencies (for example, on crimes or births) are now computerized; and (2) inexpensive GIS software now exists that can match the thousands of addresses in these records to point locations, and then add up area totals for small geographic areas (such as blocks or census tracts).

The indicators in their systems cover topics such as births, deaths, crime, health status, educational performance, public assistance, and property conditions. Operating under long-term data-sharing agreements with the public agencies that create the base records, they recurrently obtain new data, integrate them in their systems, and make them available to a variety of users for a variety of purposes. Their accomplishment demonstrates that, while never easy, it is quite possible today to overcome the past resistance of major public agencies to sharing their data in this way.[18]

Important for this research is what is known about the quality of their data. Urban Institute staff were generally familiar with their data holdings and the procedures they follow to ensure quality before this project began, and we found out more during the project about the specific data files sent to us for this work. All sites do follow regular procedures to check and clean the files they receive from public agencies, and the individual indicators they calculate and disseminate are documented (see web sites listed in annex A for the selected partners).

An advantage for us was that the specific data files we needed from them for this effort were among the highest quality in their systems: files on (1) vital statistics; (2) reported crimes; and (3) Aid to Families with Dependent Children/Temporary Assistance for Needy Families (AFDC/TANF) recipiency. For all of these, particularly the first two, the original data providers work under fairly tight guidelines as to data quality and use definitions that are reasonably standard for national reporting.

Local Data – Health Indicators

Virtually all NNIP partners obtain detailed vital statistics data (births and deaths) at the census tract level on an annual basis. A few maintain other health-related indicators (e.g., from records on immunizations and hospital admissions), but these indicators were not useable for our cross-site analysis because they were not uniformly available, let alone uniformly defined, in all five sites. Therefore, we relied solely on vital statistics data to develop health indicators for this component of the research. The selected measures are as follows:

---

[18] NNIP has developed a handbook that explains the histories, philosophies, and operating methods and techniques of the original NNIP partners—see *Building and Operating Neighborhood Indicators Systems: A Guidebook*, (Kingsley 1999).

Maternal and infant health indicators

- Percentage of low birth weight births of all births (low birth weight defined as less than 2,500 grams)
- Percentage of births where mother received early prenatal care (in first trimester)
- Teen birth rate (number of births to teens aged 15 to 19 per 1,000 women of that age group)

Mortality indicators

- Infant mortality rate (deaths of infants 0 to 12 months old per 1,000 live births in the same year)
- Age-adjusted mortality rate (total deaths per 100,000 population that would have occurred assuming local death rates by age category and the national percentage distribution of population in the same categories)[19]

We obtained all data needed to construct these indicators by census tract for all of the individual years from 1990 through 2000 for which the partners maintained the information (see table 7.1). Our work was generally guided by advice on constructing indicators based on vital records provided by Coulton (1998).

Table 7.1
Data Availability by Site

Health variables		
	Birth files	**Death Files**
Cleveland (Cuyahoga Co.)	1990-2000	1990-2000
Denver	1990-2000	1990-2000
Indianapolis	1990-2000	1990-2000
Providence	1995-2000	none
Oakland	1990-2000	1990-1999
Context Variables		
	Crime	**AFDC/TANF**
Cleveland (Cuyahoga Co.)	1990-2000	1992-2000
Denver	1990-2000	1999
Indianapolis	1992-2000	1998-2000
Providence	2000	1996, 1998, 2000
Oakland	1996-2000	none

[19] The mortality rates are age-adjusted to the year 2000 standard population. As to the age groupings, we used a standard categorization: less than 1; 1 to 14; 15 to 24; 25 to 44; 45 to 64; and 65 and over.

All data were provided to us by age and race/ethnicity categories within tracts. As to the latter, local designations in Denver, Oakland, and Providence were combined to form five categories in our analysis: Hispanic plus four groups of non-Hispanics, white, black, Asian, and other. Data from Cleveland and Indianapolis did not allow us to separate Hispanics from non-Hispanics within races, so the data for those sites are not comparable. While this must be kept in mind, we do not believe it much affects our inter-site comparisons, since there are comparatively few Hispanics in those two sites.

For the teen birth rate and the age-adjusted mortality rate, we constructed denominators from census data. To do so, we interpolated between 1990 and 2000 numbers in the relevant age categories in each tract (straight-line method) to create annual estimates to correspond to the years of the vital statistics data. Although we know change was not likely to be exactly uniform over the 1990s, this procedure seems generally reasonable and has the benefit of being consistent across sites.

We acquired the data in early 2002, after multiple conversations with staff in our partner organizations to clarify variable definitions, file specifications, and steps taken to clean the data. Where possible, we checked county-level totals and rates in the data provided to us by the partners with comparable measures for the same counties independently reported by the original local data providers to the National Center for Health Statistics. We found no unreasonable differences.

The "Rare Events" Issue

The last and most challenging step in data development was to consider how to reliably depict indicators derived from rare events (e.g., the number of low-birth weight births or infant deaths in one census tract in any given year). As Buescher (1997) explains:

> Most health care professionals are aware that estimates based on a random sample of a population are subject to error due to sampling variability. Fewer people are aware that rates and percentages based on a full population are also estimates subject to error. Random error may be substantial when the measure such as a rate or percentage, has a small number of events in the numerator. . . . A rate observed in a single year can be considered as a sample or estimate of the true underlying rate.

Buescher recommends computing a confidence interval (the interval within which we would expect the "true" rate to fall a certain percentage of the time) around the proportion or rate to help decide the size of the numerator needed for the analysis at hand. There is no hard and fast rule, but a numerator of 20 cases is often considered as an absolute minimum, since for any smaller number a 95 percent confidence interval will be wider than the rate itself. With this in mind, the numbers for infant mortality at the census tract level were found to be too small for use in this research.

To address this issue in neighborhood analysis, the data often have to be grouped to obtain larger numerators. There are two ways to do that. First, add the data for several years together for one census tract (e.g., present information for a multiyear period such as 1991–1993). Second, add the data for several tracts together for one year (e.g., present information for a new neighborhood aggregation of several tracts). Neither approach is better than the other is in general, and very small numbers may require that both approaches be used. The choice should be made based on the purposes of the analysis at hand (i.e., depending on whether you care more about a high level of geographic detail or a more finely grained examination of change over time).

Table 7.2 shows characteristics of the distribution of numerator sizes for different variables in our five sites. The data are shown first for census tracts and then for neighborhood clusters (clusters of adjacent tracts that each of the cities uses for planning purposes). These cluster definitions may be very useful for local planning and action, when policymakers and community members have a common understanding of the different areas. Unfortunately, they vary a great deal in size and how they were defined, making cross-site comparisons difficult. For example, the city of Cleveland has 224 tracts and 36 neighborhood clusters, for an average of 6.2 tracts per cluster, whereas Denver has only 1.8 tracts per cluster.

For total births and births to mothers with early prenatal care, a comparatively small share of the tracts had fewer than 20 events during the three-year period 1998–2000, whereas for low-birth weight births and births to teens, the share below 20 is very high. Switching to neighborhood clusters reduces all shares below 20 significantly, although the shares are still fairly high for low-birth weight births and births to teens in all cities except Cleveland.

It is one thing to use tract-level data in a regression (as we do in section 5), but it is quite another to present exact rates in a table form for individual tracts or even neighborhoods at these levels. Tract-level tables with three-year data might make sense for total births and births to mothers with early prenatal care, if a special symbol instead of a number were given for tracts with very small numerators. However, switching to the neighborhood level would certainly be advisable for rates of low-birth weight births and births to teens, and strong cautions would need to be stated.

Table 7.2
Ranges in Numbers of Events, Census Tracts and Neighborhood Clusters 1998-2000

	Cleveland (Cuyahoga Co.)	Denver (City/Co.)	Indiana-polis (Marion Co.)	Oak-land (City)	Provi-dence (City)
Census Tracts (Total number)	224	143	203	37	105
Total births, 1998-2000					
Median events/tract	105	163	153	150	154
% of tracts, < 20 events	13	7	2	3	3
Births to mothers with early prenatal care					
Median events/tract	75	128	115	112	134
% of tracts, < 20 events	18	7	2	3	4
Low birth weight births					
Median events/tract	10	14	14	13	12
% of tracts, < 20 events	83	62	73	68	78
Births to teens (age 15-19)					
Median events/tract	18	14	23	29	13
% of tracts, < 20 events	54	58	44	38	61
Neighborhood Clusters (NC) (Total no.)	36	79	48	15	44
Total births, 1998-2000					
Median events/NC	666	295	398	485	382
% of NCs, < 20 events	0	4	0	0	0
Births to mothers with early prenatal care					
Median events/NC	501	222	314	317	346
% of NCs, < 20 events	3	4	0	0	0
Low birth weight births					
Median events/NC	71	25	35	38	29
% of NCs, < 10 events	3	22	6	13	5
% of NCs, < 20 events	11	34	25	13	20
Births to teens (age 15-19)					
Median events/NC	112	30	53	85	49
% of NCs, < 10 events	3	28	4	20	9
% of NCs, < 20 events	3	35	19	27	27

In this research we employed both approaches at different stages. We avoid these problems in our analysis of time trends in section 4 by presenting data only for higher aggregations of time and geography. We group the data for three-year periods and present results for only two geographic aggregations of tracts in each city: (1) all high-poverty tracts and (2) all other tracts. In the bivariate and multivariate analyses (as will be explained more fully in section 5), we use tract-level data grouped for three-year periods. Each three-year tract average is an observation in the regression, and we believe the large number of observations (more than 8,000) and the multivariate methodology offset any random year-to-year variation.

Local Contextual Variables

The bulk of the demographic and contextual data needed for the cross-site analysis could be obtained from the census (see discussion below), but we felt that it would helpful to have two types of year-by-year information from local partners' data systems in addition.

The first is reported crime. We obtained tract-level data for Part I crimes (as uniformly defined by the FBI) from all sites. The reporting of Part I crimes is guided by the standards set by the Uniform Crime Reporting (UCR) program. The UCR provides a nationwide view of crime based on the submission of statistics by law enforcement agencies throughout the country. Part I crime consists of eight offenses—murder, forcible rape, robbery, aggravated assault, arson, burglary, larceny-theft, and motor vehicle theft. Murder, rape, robbery, and aggravated assault are crimes against persons. Arson, burglary, larceny-theft, and auto theft are crimes against property.

Data are provided for the following summary indicators for years as noted in table 7.1:[20]

- Violent crimes per 1,000 population
- Property crimes per 1,000 population
- Total Part I crimes per 1,000 population

The second topic is federal welfare (AFDC/TANF) recipiency. In this case, Oakland is the only one of our sites that has been unable to obtain data on the topic. For the others, three provided individual record data and one provided household data. We have calculated the following indicators:[21]

---

[20] Oakland's crime data are available only for police beat areas, but the staff of our partner institution there created tract-level estimates for the purposes of this study.

[21] For all reported crime and AFDC/TANF indicators, population denominators for specific years are again estimated by interpolating between 1990 and 2000 census figures.

- Individuals receiving AFDC/TANF as a percentage of total population (Cleveland, Denver, and Indianapolis)
- AFDC/TANF cases as a percentage of total households (Providence)

NNIP partners maintain other contextual indicators that could have potentially been useful for this analysis. For instance, Denver, Oakland, and Providence have more extensive education data, while Cleveland and Providence have more information about property conditions. Because we did not have data on these topics uniformly across at least three or four sites, however, we decided not to try to incorporate a wider range of variables in this work.

National Data Files: The Census and the Neighborhood Change Database

Before NNIP-type data systems were set up, tract-level data from the U.S. censuses were virtually the only nationally comparable indicators of neighborhood conditions in America. For our type of analysis, however, the data made available directly by the U.S. Bureau of the Census have a problem: all is not held constant from one census to the next. Some variable definitions and, more important, about 35 percent of tract boundaries change between censuses, so the data are not directly comparable over time.

To remedy this, the Rockefeller Foundation funded the Urban Institute and GeoLytics, Inc., to go back over the definitions and the data (using block data where possible to make adjustments) and achieve comparability. The result was the Neighborhood Change Data Base (NCDB), the only database that contains nationwide census data at the tract level with tract boundaries and variables that are consistently defined across the four U.S. censuses from 1970 through 2000.[22]

In this research, we rely on the NCDB as a primary source of data for measuring 1990–2000 conditions and trends in most of the topical categories introduced earlier in this section. At the neighborhood level, they include data on demographic, economic, social, and housing characteristics and neighborhood stability. The NCDB is also our source for measures of economic health and segregation at the metropolitan level.

Since the sites' year-by-year data on health and contextual indicators are based on tract boundaries as defined for the 1990 census, we conduct all of these analyses using 1990 tract boundaries (i.e., weights developed in the NCDB were used to enable us to present 2000 census data for 1990-defined tracts).

---

[22] To find out about the structure and contents of the NCDB, visit http://www.urban.org/nnip/ncua/ncdb and http://www.geolytics.com.

National Data Files: Home Mortgage Disclosure Act

The Federal Reserve annually releases files on home mortgage applications by census tract nationwide, as required by the Home Mortgage Disclosure Act (HMDA).[23] The files contain records on individual applications, including the census tract identifier, loan amount, race and income of applicant, purpose of loan (purchase/home improvement, owner/renter occupied), and whether the application was approved or denied. To prepare for this analysis, we compiled the individual loan data for originated mortgages (those both approved by the banks and accepted by the borrower) from 1995 through 1999 by tract. We then constructed indicators for the number and average value of loans, broken down by those used for home purchase and those used for home improvement. These indicators act as proxies for neighborhood investment and home values.

RESEARCH HYPOTHESES

Static Hypotheses

Considering our potential range of independent variables within the categories noted earlier, we have formed the following hypotheses about these associations at any point in time.

Socioeconomic hypotheses

- Census tracts with a majority nonwhite population and higher levels of immigrants will have higher levels of mortality and poor maternal and infant health outcomes than majority white census tracts.
- Low-income census tracts, as measured by poverty rate and average income, will be associated with poorer scores on the mortality and the maternal and infant health measures than higher-income tracts.
- Census tracts with higher overall social risks (e.g., lower education and employment levels, higher rates of welfare or public assistance recipiency) will have poorer scores on the mortality and the maternal and infant health measures than those with lower risks.

Physical stressors

- Census tracts with poor housing quality, as measured by age of the housing, overcrowded units, and low home values, will have higher levels of mortality and poorer maternal and infant health outcomes than stronger and more stable tracts.

[23] For documentation on HMDA data files, visit http://ffiec.gov.

Social stressors

- Census tracts with high total, violent, or property crime rates will have poorer scores on the mortality and the maternal and infant health measures than safer tracts.

Social networks

- Census tracts with less stable populations, as measured by renter occupancy, vacancy rate, and mobility rate, will have higher levels of mortality and poorer maternal and infant health outcomes than stronger and more stable tracts. Tracts with less change in total or minority population or a higher rate of home improvement or refinancing loans will have better mortality and birth outcomes.

Dynamic Relationships

The discussion so far has focused on relationships between indicators at a point in time. Since we had indicators over a period of several years, we were able to examine how these relationships changed over time. We tested all of the hypotheses above to see if the associations remained constant throughout our analysis period. The time periods examined depended on the data source. We looked at 1990 and 2000 census data, 1995–2000 HMDA data, and varied dates for local contextual variables. Generally, we hypothesized that the relationships held over the 1990s; that is, as the level of one indicator that was negatively related to health (such as the crime rate) rose, the health outcomes worsened.

However, there are three cases in which we think that the relationships will continue to be in the same direction but will be weaker at the end of the decade than at the beginning. First, we know that racial disparities, while still considerable, are decreasing for some maternal and health outcomes, so we suspect that the relationship between high-minority tracts and poor health outcomes may be somewhat reduced (Keppel 2002). Second, Medicaid, Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), and other public programs have been expanded over the past decade, enabling more pregnant women to access prenatal care. Low-income tracts should therefore be less correlated (although still positively) with the percentage of births to mothers receiving late or no prenatal care.

Finally, we believe two factors have altered the relationship between low-birth weight births and the income level of the tract. The better care that low-income women are receiving will reduce the number of low-birth weight births and allow more of the remaining low-birth weight babies to survive. On the higher end of the income scales, more affluent women are delaying having children, thereby increasing their chances of having a low-birth weight baby. As with the first two, we believe the relationship will remain significant and positive, but at a lesser magnitude than earlier in the decade.

In sum, our hypotheses about changing relationships between health and demographic variables are as follows:

- The correlation between high-minority tracts and poor birth and mortality outcomes will remain positive, but has decreased over the 1990s.
- The correlation between low-income tracts and births with late or no prenatal care will remain positive, but has decreased over the 1990s.
- The correlation between low-income tracts and high rates of low-birth weight births will remain positive, but has decreased over the 1990s.

METHODOLOGY

Tables, Graphs, and Mapping Analysis

We have been working with a large amount of information, and our challenge has been to develop statistical and graphic methods to test these hypotheses and to cogently display the key findings. We begin our analysis in sections 8 and 9 with a series of tables and graphs. Section 8 tells the story of the changing context of urban change in the 1990s in our five sites. Tables there cover comparative metropolitan characteristics, and then go on to contrast demographic and nonhealth conditions and trends in high-poverty neighborhoods with those in other urban neighborhoods over the decade.

Section 9 begins by similarly contrasting health conditions and trends in high-poverty tracts with other tracts in the five cities. Differences between high-minority tracts and low-minority tracts within each city are also discussed. Using these divisions, we examine the extent to which these noncontiguous groupings of tracts are helpful in revealing health disparities among different types of neighborhoods. From the tabular information, we identify interesting stories to depict in graphic format. We include line charts illustrating the change over time and differences between types of neighborhoods across cities.

Also in section 9, we present illustrative maps of the most informative indicator values and the change in levels over the time. High-poverty areas (tracts with greater than 30 percent poverty) are indicated by a dot pattern laid over colors that indicate patterns for the health indicators.

Correlation and Regression Analyses

In section 10, we calculate the correlation coefficients among the health and contextual variables across time. These matrices allow us to examine the magnitude and direction of the connections among variables. With our longitudinal data files, we examine how some of these relationships have changed over the 1990s—whether they are growing stronger or weaker. These results will be used to confirm or reject our hypotheses and to suggest the most meaningful variables to map together. Full correlation matrices are included in annex C.

We use multivariate analysis to analyze health outcomes (e.g., low birth weight rates) as dependent variables. Independent variables will include some area characteristics as control variables (e.g., sociodemographic characteristics), as well as the community characteristics that are being investigated as possible causes for the health problem (such as crime rates, housing quality, or other factors).

Again, as discussed above, relationships that are statistically significant will not definitively identify causal relationships but rather will be used in discussions with communities as ways to develop hypotheses regarding what could be contributing to an identified community problem.