MEASURING SPRAWL AND ITS IMPACT

Volume I

by

Reid Ewing
Rutgers University

Rolf Pendall
Cornell University

Don Chen
Smart Growth America

1 INTRODUCTION

Across the nation, growing numbers of communities are discovering links between urban sprawl and a wide range of problems, from traffic and air pollution to central city poverty and the degradation of scenic areas. As more civic leaders take steps to ameliorate these costs, they are in increasing need of meaningful information about the characteristics, extent and consequences of sprawl.

To help meet these needs, Smart Growth America (SGA) has sponsored this groundbreaking research by Rutgers University Professor Reid Ewing and Cornell University Professor Rolf Pendall. It represents a rigorous effort to measure the characteristics of sprawl and their impacts on quality of life. In this study, sprawl is defined as low-density development with residential, shopping and office areas that are rigidly segregated; a lack of thriving activity centers; and limited choices in travel routes. These features constitute *four factors* that can then be measured and analyzed: 1) Residential density; 2) Neighborhood mix of homes, jobs, and services; 3) Strength of centers, such as business districts; and 4) Accessibility via the street network. All of these are well-established descriptors of urban sprawl in the relevant academic literature, but this study represents the first effort to attempt to measure sprawl in all of these dimensions.

The heart of this project is an extensive database that allows for both the careful measurement of urban sprawl as well as the assessment of its relationship to a wide variety of quality-of-life indicators. The database contains two sets of variables. The first set includes 22 variables grouped into the four factors that characterize sprawl. The second set of data includes dozens of indicators of community quality of life, including everything from how much people drive every day to the consumption of farmland and forests. This report is the first of several that will assess the impact of sprawl on these important outcomes.

This research is significant for two main reasons. First, it is by far the most comprehensive attempt to define and quantify urban sprawl in the U.S. Some studies have defined sprawl simply in terms of the amount of land used as the population grows, but ignoring the form in which that growth occurs. This study shows that sprawl is not just growth, but is a specific, and dysfunctional, style of growth. By evaluating metropolitan growth patterns based on four factors, this research presents a highly detailed portrait of sprawl that will enable decision-makers to target their growth management strategies more effectively. Second, and perhaps more importantly, this study analyzes how growth patterns and affect everyday things that people value. In other words, the researchers have demonstrated that sprawl is a real, measurable phenomenon, and it has real, measurable consequences for people.

This first volume presents sprawl measures for 83 of the largest metropolitan areas in the United States and examines the relationships between urban sprawl and transportation-related measures, including vehicle miles traveled, traffic fatalities, the extent of walking and public transit use, roadway congestion and air quality. Future volumes will address how sprawl may be influencing other outcome measures, such as the decline of central cities, the loss of open space, the degradation of scenic areas, and impacts on public health. Also, some data will be examined at the county level to explore the variation of development patterns within different metropolitan areas.

2 OVERVIEW

Two features distinguish this study from earlier investigations of sprawl. First, this effort defines and operationalizes sprawl more completely and for a larger set of metropolitan areas than have earlier studies. Second, we relate sprawl to a more extensive set of outcomes (in this and follow-up reports) than has any study to date.

2.1 Characterization of Sprawl

In the most comprehensive literature synthesis to date, Burchell et al. list as "defining characteristics of sprawl" three distinct types: spatial patterns, root causes, and main consequences of sprawl, 10 of them in all.[1] Spatial patterns on Burchell et al.'s list are: "low density," "unlimited outward expansion," "land uses spatially segregated," "leapfrog development," and "widespread commercial strip development." Two causes of sprawl also make the list: "no central ownership or planning" and "highly fragmented land-use governance." Finally, three consequences of sprawl appear on the list: "transport dominance by motor vehicles," "great variance in local fiscal capability," and "reliance on filtering for low-income housing."

The most basic decision that has to be made in developing a sprawl index is whether to limit the index to spatial patterns, or to include causes and consequences as well. We are persuaded by the logic of Galster et al. "Conceptually, a thing cannot simultaneously be what it is and what causes it or what it causes. If sprawl is to be a useful concept for describing something important

---

[1] Robert Burchell et al., *Costs of Sprawl Revisited: The Evidence of Sprawl's Negative and Positive Impacts*, Transit Cooperative Research Program, Transportation Research Board, Washington, D.C., 1998, Table 12.

that occurs in urban areas, it must first be reduced to some objective conditions or traits."[2] In this study, we define sprawl solely in terms of spatial patterns. Separately, once defined and measured, sprawl is related to consequences in the areas of transportation, travel, and air quality.

As for spatial patterns, a literature review by Ewing finds *poor accessibility* the common denominator of sprawl—with the following patterns most often identified in the literature: scattered or leapfrog development, commercial strip development, uniform low-density development, or single-use development (with different land uses segregated from one another, as in bedroom communities).[3] In scattered or leapfrog development, residents and service providers must pass vacant land on their way from one developed use to another. In classic strip development, the consumer must pass other uses on the way from one store to the next; it is the antithesis of multipurpose travel to an activity center. Of course, in low-density, single-use development, everything is far apart due to large private land holdings and segregation of land uses.

The second big decision in developing a sprawl index is exactly which patterns should qualify as sprawl. In this study, the decision is largely dictated by data availability. Because we are attempting to measure sprawl for metropolitan areas across the United States, data has to be available from national sources. From national sources, sprawl may be characterized by:

- Low Development Density
- Segregated Land Uses
- Lack of Significant Centers

In addition to these land use characteristics, national data are available on the backbone of the transportation system, the region's street network. While the technical literature on sprawl focuses on land use patterns that produce poor regional accessibility, poor accessibility is also a product of fragmented street networks that separate urban activities more than need be. When asked, planners now routinely associate sprawl with sparse street networks as well as dispersed land use patterns. Thus to the land use characteristics of sprawl can be added:

- Poor Street Accessibility

Figure 1. Sprawling Development Patterns

---

[2] George Galster, Royce Hanson, Michael Ratcliffe, Harold Wolman, Stephan Coleman, and Jason Freihage, "Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept," *Housing Policy Debate*, Vol. 12, no. 4, 2001, p. 685.

[3] Reid Ewing, "Is Los Angeles-Style Sprawl Desirable?" *Journal of the American Planning Association*, Vol. 63, No. 1, Winter 1997, pp. 107-126.

Low Density and Single Use Development

Uncentered Strip Development

Scattered and Leapfrog Development

Sparse Street Network

2.2 Relation to Outcomes

National data are also available on outcome measures reflecting the quality-of-life and sustainability of metropolitan development, and on control variables that might account for these outcomes independent of sprawl. The purpose of this study is to see if outcomes are linked to sprawl, controlling for other influences. For this release, the following travel and transportation outcomes are analyzed for relationships to sprawl:

- Vehicle Ownership
- Commute modes
- Commute times
- Vehicle miles traveled
- Congestion
- Fatal accidents
- Air quality

For most travel and transportation outcomes, sprawling regions perform less well than compact ones. This is true of everything from transit use to traffic fatalities. The exceptions are commute times and congestion, which do not clearly favor compactness over sprawl.

Subsequent releases will deal with the effects of sprawl on:

- Public health
- Open space
- Public service costs
- Consumer expenditures
- Housing affordability
- Racial segregation

3 PREVIOUS ATTEMPTS TO MEASURE SPRAWL

3.1 Studies Simply Measuring Sprawl[4]

3.1.1 USA Today

The sprawl index to receive the most attention, despite its limitations, was developed by *USA Today*.[5] The *USA Today* index assigned a score to each of 271 metropolitan areas based on two density-related measures:

- Percentage of a metro area's population living in urbanized areas. For the years in question, the Census Bureau defined "urbanized" as those parts of a metro with 1,000 or more residents per square mile.

- Change in the percentage of metropolitan population living in urbanized areas between 1990 and 1999.

Metropolitan areas were ranked 1 through 271 on each measurement (with lower numbers representing less sprawl). The two rankings were summed to produce each metro area's sprawl score. The highest possible score was 542, the lowest 2. The advantage of the *USA Today* index is its simplicity, which makes it easy to explain. The big disadvantage is its total reliance on density as an indicator of sprawl, and density measured in a way that fails to distinguish between development at low suburban densities (as low as 1,000 persons per square mile, something less than one dwelling unit per acre) and development at high urban densities. Based on this index, USA Today declared:

"Los Angeles, whose legendary traffic congestion and spread-out development have epitomized suburban sprawl for decades, isn't so sprawling after all. In fact, Portland, OR, the metropolitan

---

[4] Sprawl has been measured in other ways for individual metropolitan areas. This literature survey is limited to studies which, like this one, use a comparative index to rank metros in terms of sprawl. For examples of individual area studies, see Cameron Speir and Kurt Stephenson, "Does Sprawl Cost Us All? Isolating the Effects of Housing Patterns on Public Water and Sewer Costs," *Journal of the American Planning Association*, Vol. 68, No. 1, Winter 2002, pp. 56-70; and Lance Freeman, "The Effects of Sprawl on Neighborhood Social Ties: An Exploratory Analysis," *Journal of the American Planning Association*, Vol. 67, No. 1, Winter 2001, pp. 69-77.
[5] *USA Today*, February 22, 2001.

area that enacted the nation's toughest anti-growth laws, sprawls more." Indeed, according to USA Today's index, Los Angeles is less sprawling than even the New York metropolitan area.

3.1.2 Sierra Club

In a report titled *The Dark Side of the American Dream: The Costs and Consequences of Suburban Sprawl*, the Sierra Club ranked U.S. metropolitan areas on the degree to which they sprawl.[6] Sprawl was defined as "low-density development beyond the edge of service and employment, which separates where people live from where they shop, work, recreate and educate—thus requiring cars to move between zones."

Metros were subjectively rated as more or less sprawling based on population shifts from city to suburb, growth of land area vs. growth of population, time wasted in traffic, and loss of open space. Sprawl was thus defined not only by its characteristics but its effects. Among the largest metros (1 million or more people), Atlanta, St. Louis, and Washington, D.C. were rated most sprawling. Among medium size metros (500,000-1,000,000 population), Orlando, Austin, and Las Vegas shared that distinction.

3.1.3 Galster et al.

Galster et al. developed the most complex and multi-faceted sprawl index to date.[7] Sprawl was characterized in eight dimensions: density, continuity, concentration, clustering, centrality, nuclearity, mixed use, and proximity. The condition, sprawl, was defined as pattern of land use that has low levels in one or more of these dimensions. Variables representing causes and consequences of sprawl, such as fragmented governance and auto dependence, were explicitly excluded from the definition.

Each dimension was operationally defined and six of the eight were quantified for 13 urbanized areas. New York and Philadelphia ranked as the least sprawling of the 13, and Atlanta and Miami as the most sprawling. The main drawback of Galster et al.'s index is its availability for only 13 areas. Also limiting is its analysis of urbanized areas rather than metros, since the most sprawling development in many metros occurs outside urbanized areas. Finally, like all other attempts to measure sprawl so far, this study was unable to quantify one very important dimension of sprawl—the segregation of different land uses at the expense of accessibility.

3.2 Studies Measuring Sprawl and Relating It To Outcomes

3.2.1 Kahn

[6] Sierra Club, *The Dark Side of the American Dream: The Costs and Consequences of Suburban Sprawl*, Challenge to Sprawl Campaign, College Park, MD, undated.
[7] Galster et al., op. cit., pp. 681-717.

Kahn explored one potential benefit of sprawl, increased housing affordability and greater equality of housing opportunity across racial lines.[8] Using 1997 American Housing Survey data, Kahn measured housing consumption for blacks and whites in metropolitan areas characterized as more or less sprawling. Housing consumption was represented by number of rooms, unit square footage, homeownership rates, and year of construction. For his measure of sprawl, Kahn drew upon his research with Glaeser (see below). Sprawl was represented by the degree of employment decentralization in a metro area, specifically, by the proportion of metropolitan employment located more than 10 miles from the central business district. If all employment were located inside a 10-mile ring around the CBD, Kahn's "sprawl level" would be zero. If all were located outside the 10-mile ring, the sprawl level would be 1. As it is, values of this index varied from 0.196 for Portland to 0.786 for Detroit.

The most obvious drawback of this sprawl measure is the failure to consider residential development patterns. Almost as important is the failure to consider the multi-centered employment patterns characteristic of large metropolitan areas. The final drawback is the bias against larger metros, which by virtue of size alone, are likely to have a larger portion of their employment beyond 10 miles from the central business district. By this definition, a small metropolitan area less than 20 miles across would be completely sprawl-free.

Based on this characterization of sprawl, Kahn found that the housing differential between blacks and whites is narrower in more sprawling areas, at statistically significant levels for two measures of housing consumption. Sprawl was said to increase black housing consumption by making low-cost suburban housing available to everyone and by freeing up inner city housing for blacks as whites move to the suburbs to take advantage of cheap housing.

3.2.2 Downs

In a 1999 article in *Housing Policy Debate*, Downs presented a general outline of his research on sprawl and its effects on urban decline.[9] His conclusion: No meaningful and significant statistical relationship exists between specific traits of sprawl and measures of urban decline. Concentrated urban poverty, which Downs views as the source of urban decline, would occur with or without sprawl.

In Chapter 13 of *The Cost of Sprawl Revisited*, Downs made his methodology explicit.[10] Again, he tested for statistically significant relationships between suburban sprawl and urban decline, and found none. Sprawl was defined in terms of an assortment of land use patterns, root causes of these patterns, and specific consequences of these patterns. Thus, Downs' conception of sprawl failed to distinguish causes and consequences from characteristics of sprawl.

---

[8] Matthew Kahn, "Does Sprawl Reduce the Black/White Housing Consumption Gap?" *Housing Policy Debate*, Vol. 12, No. 1, 2001, pp. 77-86.
[9] Anthony Downs, "Some Realities About Sprawl and Urban Decline," *Housing Policy Debate*, Vol. 4, No. 4, 1999, pp. 955-974.
[10] Robert Burchell et al., *Costs of Sprawl Revisited—2000*, Transit Cooperative Research Program, Transportation Research Board, Washington, D.C., 2002, Chapter 13.

The following traits were measured for urbanized areas, specifically for 162 urbanized areas that had 1990 populations of 150,000 or more. All data came from the 1990 U.S. census.

- Land area of the urbanized area
- Population density of the urbanized area outside the central city or cities
- Ratio of central city population density to density of the urbanized fringe
- Percentage of the total metropolitan area population living outside the urbanized area
- Percentage of the total metropolitan area population living within the central city or cities
- Percentage of urbanized area commuters who drive alone or in carpools
- Number of separate jurisdictions that control land use per 100,000 metro area residents
- Ratio of poor central city residents to poor suburban residents

To obtain a *sprawl score*, variable values were normalized on a scale of 0 to 100 and then averaged. Measured this way, 6 of the 20 most sprawling urbanized areas are located in the Midwest or Northeast. Several of these older cities—such as Hartford, Scranton, and Harrisburg—get high composite sprawl scores because they have very low percentages of their total metropolitan area populations within central city boundaries. Other midwestern cities get high sprawl scores because they have many local governments per 100,000 residents. By this measure, among the least sprawling metros are Los Angeles, Phoenix, Tucson, San Diego, and Denver.

In addition to mixing characteristics, causes, and effects of sprawl, Downs' index suffers from: reliance on political, and hence economically arbitrary, boundaries of central cities to define centeredness; reliance on the urbanized area definition of 1,000 residents per square mile to define the worst of all sprawl. In this last respect, Downs' index is subject to the same criticism as *USA Today's* (see above).

3.3 Studies Measuring Sprawl and Exploring Causes

3.3.1 Glaeser et al.

Glaeser et al. related sprawl to the degree of decentralization of employment using data from the U.S. Department of Commerce's Zip Code Business Patterns for 1996.[11] Zip code business patterns data were extracted from the Standard Statistical Establishments List, a file of all single and multi-establishment companies listed by zip code and firm size.

For the 100 largest U.S. metropolitan areas, the share of overall metropolitan employment within a three-mile ring of the Central Business District was computed, as were the shares inside and outside a 10-mile ring. The share within three miles reflects the presence or absence of a well-defined employment core, while the share beyond 10 miles captures the extent of job sprawl. Metros were then divided into four categories, based on values of these indices. Dense

[11] Edward Glaeser, Matthew Kahn, and Chenghuan Chu, *Job Sprawl: Employment Location in U.S. Metropolitan Areas*, Center for Urban & Metropolitan Policy, The Brookings Institution, Washington, D.C., July 2001.

employment metros like New York have at least one quarter of their employment within three miles of the city center. Centralized employment metros like Minneapolis-St. Paul have between 10 and 25 percent of employment within three miles of the city center, and more than 60 percent within 10 miles. Decentralized employment metros like Washington D.C. have 10 to 25 percent of employment within the three-mile ring, and less than 60 percent within 10 miles. Finally, extremely decentralized employment metros like Los Angeles have less than 10 percent of their employment within the three-mile ring.

To explain differences across metros, Glaeser et al. related their measure of sprawl to the age of the metro area (year in which the primary city was founded) and to the degree of political fragmentation within the metropolitan area (number of local jurisdictions within their boundaries). They found no relationship to age but a statistically significant relationship between political fragmentation and job decentralization.

3.3.2 Pendall

Pendall sought to explain the incidence of sprawl for large metropolitan areas in terms of land values, metropolitan political organization, local government spending, traffic congestion, and various local land use policies.[12] Among land use policies, adequate public facilities requirements, which force new development to pay its own way, were found to discourage sprawl, while low-density zoning and building caps were associated with more sprawl. Among control variables, high valued farmland and expensive housing reduced sprawl, while jurisdictional fragmentation increased it.

Pendall's measure of sprawl is strictly related to density. Population was divided by urban acreage to obtain density estimates for 1982 and 1992, with urban acreage taken from the U.S. Department of Agriculture's Natural Resources Inventory. To measure increases in sprawl over time, estimates of population change between 1982 and 1992 were divided by estimates of change in urban land during the same period. By this measure, Los Angeles, San Francisco, and San Diego were the most compact metros in 1992, while Milwaukee, Atlanta, Cleveland, and Denver were the most sprawling. Baltimore, Los Angeles, and San Diego grew in the most compact manner over the decade, while Boston, Cincinnati, and Minneapolis-St. Paul grew in the least.

3.3.3 Fulton et al.

Building on Pendall's earlier work, Fulton et al. studied urban land consumption relative to population change for every U.S. metropolitan area.[13] If land is consumed at a faster rate than

---

[12] Rolf Pendall, "Do Land-Use Controls Cause Sprawl?" *Environment and Planning B*, Vol. 26, No. , 1999, pp.
[13] William Fulton, Rolf Pendall, Mai Nguyen, and Alicia Harrison, *Who Sprawls Most? How Growth Patterns Differ Across the U.S.*, Center for Urban & Metropolitan Policy, The Brookings Institution, Washington, D.C., July 2001.

population is growing, sprawl is said to be increasing. As with Pendall's earlier work, this concept of sprawl is strictly density-related.

By this criterion, the West is home to some of the least sprawling metropolitan areas in the nation. By contrast, the Northeast and Midwest are in some ways the nation's biggest sprawl problems since they add few new residents, yet consume large amounts of land. In this study, Honolulu and Los Angeles were rated most compact in 1997, and Las Vegas and Phoenix (often characterized as sprawling badly) were both in the top 20 in compactness. Las Vegas and Phoenix were first and third in density gain over the 15 years studied, 1982 to 1997.

The report also examined the causes of sprawl. Metropolitan areas tend to consume less land for urbanization—relative to population growth—when they are growing rapidly in population, rely heavily on public water and sewer systems, and have large immigrant populations. Metropolitan areas tend to consume more land for urbanization—again, relative to population growth—if they are compact to begin with and have fragmented local governance.

3.4 Recap of Past Studies

The most notable feature of past studies (with few exceptions like that of Galster et al.) is the failure to define sprawl in all its complexity. Density is relatively easy to measure, and hence serves as the sole indicator of sprawl in several studies. This flies in the face of both the technical literature and popular conceptions of sprawl.

Another notable feature, related to the first, is the wildly different sprawl ratings given to different metros by different analysts. With the exception of Atlanta, which always seems to rank as one of the worst, the different variables used to operationalize sprawl lead to very different results. In one study, Portland is ranked as most compact and Los Angeles is way down the list. In another their rankings are essentially reversed. In a third study, certain Northeastern metros are characterized as sprawling, in a fourth they are relatively compact.

A third notable feature is how little attention has been paid to the impacts of sprawl. With the exception of a few studies focusing on individual impacts, the literature is nearly devoid of empiricism. Sprawl is presumed to have negative consequences, or presumed to be free of them, depending on the ideological bent of the author.

4 OPERATIONALIZING SPRAWL

Sprawl, and its antithesis compact development, are constructs. Constructs are theoretical abstractions such as "intelligence" in education, "utility" in economics, and "biodiversity" in ecological studies. They must be *operationalized* to be investigated empirically. This means that they must be represented by variables that can be objectively measured. Operational variables are seldom true and complete representations of the underlying constructs to which they relate. Ordinarily, they capture some features of the construct but neglect other features and incorporate certain irrelevant features. Also, operational variables are subject to measurement and sampling errors. Thus, multiple variables are usually required to capture the essence of constructs.

Previous attempts to operationalize "sprawl" with a variable or two include *USA Today's* proportion of metropolitan population living outside the urbanized area, and Kahn's proportion of metropolitan employment more than 10 miles from the central business district. A variable or two cannot adequately represent the inherent complexity of sprawl.

The various dimensions of development that distinguish compact development from sprawl are also constructs. Development density, land use mix, degree of centering, and street accessibility can be measured in many ways, none of which perfectly captures the underlying construct. In this study, multiple variables are reduced to a few factors that represent the various dimensions of sprawl, and these factors are then combined to produce an overall sprawl index. Sprawl factors and indices are derived for the largest metropolitan areas in the United States.

4.1 Metropolitan Coverage

Our sample of US metropolitan areas consists of the 101 largest metropolitan statistical areas (MSAs), consolidated metropolitan statistical areas (CMSAs), and New England county metropolitan areas (NECMAs). These were the largest as of 1990, our reference year for defining metropolitan boundaries.

Outside New England, metropolitan areas are defined by county boundaries. Within New England, they are defined by town and city boundaries, and separately, by county boundaries. For consistency's sake, and also to tap into the largest number of databases, NECMAs became our units of analysis in New England, rather than MSAs or CMSAs.

As the study progressed, we deemed primary metropolitan statistical areas (PMSAs) to be more logical units of analysis than entire CMSAs. The extreme example, New York CMSA, consists of nine diverse PMSAs. Thus, our sample of CMSAs was disaggregated into PMSAs. This disaggregation occurred for all but the NECMAs, since PMSA equivalents (that is aggregations of counties into smaller metropolitan units) are not defined for NECMAs.

Within our sample of 101 MSAs/CMSAs/NECMAs are 139 MSAs/PMSAs/NECMAs. Ultimately, we limited the sample to 83 metros for reasons of metro comparability and data availability. Smaller metros seem fundamentally different from large ones when it comes to land use patterns. They are more likely to be monocentric, for example, while large metropolitan are likely to be polycentric. It seems wrongheaded to compare the degree of sprawl in Madison to the degree of sprawl in Milwaukee or Chicago, particularly since the big metros appear to sprawl less than the small ones when sprawl is measured in terms of density and its correlates. This fact runs counter to public perceptions of sprawl, an important consideration if our sprawl index is to have *face validity*.

Also, the availability of data drops off as metropolitan population declines, and sample sizes shrink for those land use and outcome measures based on samples. For instance, the American

Housing Survey (AHS) identifies metropolitan area of residence for respondents from 95 of 103 metropolitan areas with populations over one half million as of year 2000; land use measures can be extracted from the AHS for these 95 metropolitan areas but not the others. By contrast, the AHS identifies metro of residence for only 22 of the 36 smaller metropolitan areas in our original sample. And sample sizes drop to a couple dozen or less for some AHS variables. So several important variables would be missing or unreliable for a large subset of the sample if all metros were included. The final sample is further limited by data availability from the Census Transportation Planning Package (CTPP) and Glaeser et al.'s study of Zip Code Business Patterns (Business Patterns).

Table 1 summarizes data availability by metropolitan area for the original sample. Metros in the final sample are highlighted in bold type. The final sample includes 83 metros with populations over one half million as of 2000. In that year, these metros collectively were home to more than 150 million Americans, more than half of the entire U.S. population.

Table 1. Data Availability for Major Metropolitan Areas

	CTPP	Business Patterns	AHS
Akron, OH PMSA	yes	yes	yes
Albany-Schenectady-Troy, NY MSA	yes	yes	yes
Albuquerque, NM MSA	yes	yes	yes
Allentown-Bethlehem-Easton, PA-NJ MSA	yes	yes	yes
Anaheim-Santa Ana, CA PMSA	yes	yes	yes
Ann Arbor, MI PMSA	yes	no	no
Atlanta, GA MSA	yes	yes	yes
Augusta, GA-SC MSA	yes	no	yes
Aurora-Elgin, IL PMSA	yes	no	no
Austin, TX MSA	yes	yes	yes
Bakersfield, CA MSA	yes	no	yes
Baltimore, MD MSA	yes	yes	yes
Baton Rouge, LA MSA	yes	yes	yes
Beaver County, PA PMSA	yes	no	yes
Bergen-Passaic, NJ PMSA	yes	no	yes
Birmingham, AL MSA	yes	yes	yes
Boston-Lawrence-Salem-Lowell-Brockton, MA NECMA	yes	yes	yes
Boulder-Longmont, CO PMSA	yes	no	yes
Brazoria, TX PMSA	yes	no	no
Bridgeport-Stamford-Norwalk-Danbury, CT NECMA	yes	yes	yes
Buffalo, NY PMSA	yes	yes	yes

Canton, OH MSA	yes	yes	yes
Charleston, SC MSA	yes	no	yes
Charlotte-Gastonia-Rock Hill, NC-SC MSA	yes	yes	no
Chattanooga, TN-GA MSA	yes	yes	yes
Chicago, IL PMSA	yes	yes	yes
Cincinnati, OH-KY-IN PMSA	yes	yes	yes
Cleveland, OH PMSA	yes	yes	yes
Colorado Springs, CO MSA	yes	yes	yes
Columbia, SC MSA	yes	yes	yes
Columbus, OH MSA	yes	yes	yes
Dallas, TX PMSA	yes	yes	yes
Dayton-Springfield, OH MSA	yes	yes	no
Daytona Beach, FL MSA	yes	no	yes
Denver, CO PMSA	yes	yes	yes
Des Moines, IA MSA	yes	yes	yes
Detroit, MI PMSA	yes	yes	yes
El Paso, TX MSA	yes	yes	yes
Flint, MI MSA	yes	no	yes
Fort Lauderdale-Hollywood-Pompano Beach, FL PMSA	yes	yes	yes
Fort Wayne, IN MSA	yes	yes	yes
Fort Worth-Arlington, TX PMSA	yes	yes	yes
Fresno, CA MSA	yes	yes	yes
Galveston-Texas City, TX PMSA	yes	no	no
Gary-Hammond, IN PMSA	yes	yes	yes
Grand Rapids, MI MSA	yes	yes	yes
Greensboro-Winston-Salem-High Point, NC MSA	yes	yes	yes
Greenville-Spartanburg, SC MSA	yes	yes	yes
Hamilton-Middletown, OH PMSA	yes	no	no
Harrisburg-Lebanon-Carlisle, PA MSA	yes	yes	no
Hartford-New Britain-Middletown-Bristol, CT NECMA	yes	yes	yes
Honolulu, HI MSA	yes	yes	yes
Houston, TX PMSA	yes	yes	yes
Indianapolis, IN MSA	yes	yes	yes
Jackson, MS MSA	yes	no	yes
Jacksonville, FL MSA	yes	yes	yes
Jersey City, NJ PMSA	yes	yes	yes

Johnson City-Kingsport-Bristol, TN-VA MSA	yes	no	yes
Joliet, IL PMSA	yes	no	no
Kansas City, MO-KS MSA	yes	yes	yes
Kenosha, WI PMSA	no	no	no
Knoxville, TN MSA	yes	yes	yes
Lake County, IL PMSA	no	no	yes
Lakeland-Winter Haven, FL MSA	yes	no	no
Lancaster, PA MSA	yes	yes	yes
Lansing-East Lansing, MI MSA	yes	no	yes
Las Vegas, NV MSA	yes	yes	yes
Little Rock-North Little Rock, AR MSA	yes	yes	yes
Lorain-Elyria, OH PMSA	yes	no	no
Los Angeles-Long Beach, CA PMSA	yes	yes	yes
Louisville, KY-IN MSA	yes	yes	no
Madison, WI MSA	yes	yes	yes
Mcallen-Edinburg-Mission, TX MSA	yes	no	yes
Melbourne-Titusville-Palm Bay, FL MSA	yes	no	yes
Memphis, TN-AR-MS MSA	yes	yes	yes
Miami-Hialeah, FL PMSA	yes	yes	yes
Middlesex-Somerset-Hunterdon, NJ PMSA	yes	no	yes
Milwaukee, WI PMSA	yes	yes	yes
Minneapolis-St. Paul, MN-WI MSA	yes	yes	yes
Mobile, AL MSA	yes	no	yes
Modesto, CA MSA	yes	no	yes
Monmouth-Ocean, NJ PMSA	yes	no	yes
Nashville, TN MSA	yes	no	yes
Nassau-Suffolk, NY PMSA	yes	no	yes
New Bedford-Fall River-Attleboro, MA NECMA	yes	no	no
New Haven-Waterbury-Meriden, CT NECMA	yes	yes	yes
New Orleans, LA MSA	yes	yes	yes
New York, NY PMSA	yes	yes	yes
Newark, NJ PMSA	yes	yes	yes
Niagara Falls, NY PMSA	yes	no	no
Norfolk-Virginia Beach-Newport News, VA MSA	yes	yes	yes
Oakland, CA PMSA	yes	yes	yes
Oklahoma City, OK MSA	yes	yes	yes

Omaha, NE-IA MSA	yes	yes	yes
Orange County, NY PMSA	yes	no	no
Orlando, FL MSA	yes	yes	yes
Oxnard-Ventura, CA PMSA	yes	yes	yes
Philadelphia, PA-NJ PMSA	yes	yes	yes
Phoenix, AZ MSA	yes	yes	yes
Pittsburgh, PA PMSA	yes	yes	yes
Portland, OR PMSA	yes	yes	yes
Providence-Pawtucket-Woonsocket, RI NECMA	yes	yes	yes
Racine, WI PMSA	yes	no	no
Raleigh-Durham, NC MSA	yes	yes	yes
Richmond-Petersburg, VA MSA	yes	yes	no
Riverside-San Bernardino, CA PMSA	yes	yes	yes
Rochester, NY MSA	yes	yes	yes
Sacramento, CA MSA	yes	yes	yes
Saginaw-Bay City-Midland, MI MSA	yes	no	no
St. Louis, MO-IL MSA	yes	yes	yes
Salt Lake City-Ogden, UT MSA	yes	yes	yes
San Antonio, TX MSA	yes	yes	yes
San Diego, CA MSA	yes	yes	yes
San Francisco, CA PMSA	yes	yes	yes
San Jose, CA PMSA	yes	yes	yes
Santa Barbara-Santa Maria-Lompoc, CA MSA	yes	no	yes
Santa Cruz, CA PMSA	yes	no	no
Santa Rosa-Petaluma, CA PMSA	yes	no	yes
Scranton-Wilkes-Barre, PA MSA	no	yes	yes
Seattle, WA PMSA	yes	yes	yes
Springfield, MA NECMA	yes	yes	yes
Stockton, CA MSA	yes	no	yes
Syracuse, NY MSA	yes	yes	yes
Tacoma, WA PMSA	yes	yes	yes
Tampa-St. Petersburg-Clearwater, FL MSA	yes	yes	yes
Toledo, OH MSA	yes	yes	yes
Trenton, NJ PMSA	yes	no	yes
Tucson, AZ MSA	yes	yes	yes
Tulsa, OK MSA	yes	yes	yes

Vallejo-Fairfield-Napa, CA PMSA	yes	yes	yes
Vancouver, WA PMSA	yes	no	no
Vineland-Millville-Bridgeton, NJ PMSA	yes	no	no
Washington, DC-MD-VA MSA	yes	yes	yes
West Palm Beach-Boca Raton-Delray Beach, FL MSA	yes	yes	yes
Wichita, KS MSA	yes	yes	yes
Wilmington, DE-NJ-MD PMSA	yes	yes	no
Worcester-Fitchburg-Leominster, MA NECMA	yes	yes	yes
York, PA MSA	yes	yes	no
Youngstown-Warren, OH MSA	yes	yes	yes

We defined metro areas by their 1990 boundaries, both for 2000 and 1990. A consistent set of metro boundaries was required to compare the degree of sprawl over time and the relationship between sprawl and outcomes at the two points in time. Data availability made it easier to drop urbanizing counties from 2000 metro area boundaries than to add then rural counties to 1990 metro area boundaries. As a practical matter, the use of 1990 boundaries should have minor effects on sprawl statistics since the recently added rural counties will have populations too small to appreciably raise or lower metropolitan averages.

4.2 Sprawl Measures

In this study, we operationalized sprawl as follows:

1) Many variables were combined into four factors representing density, land use mix, degree of centering, and street accessibility.

We accomplished this via principal components analysis, an analytical technique which takes a large number of correlated variables and extracts a small number of factors that embody the common variance in the original data set. The principal component selected to represent each construct was the one explaining the greatest variation in the original data set. We reasoned that the single factor which captures the greatest amount of variance among multiple density variables is likely to be a valid and reliable measure of density. Likewise for the constructs land use mix, degree of centering, and street accessibility.

The extracted factors are weighted combinations of the original variables. The greater the correlation between an original variable and a factor, the more weight the original variable is given in the overall factor score. The more highly correlated the original variables, the more variance is captured by a single factor. In this regard, some of our factors are more successful in capturing common variance than are others.

2) Individual factor scores were converted to a scale with a mean value of 100 and standard deviation of 25.

Factor scores derived with principal components analysis had a mean value of 0 and a standard deviation of 1 for sampled metropolitan areas in 2000. The linear transformation performed in this step did not affect rankings of metros nor relative positions on the sprawl scale. It simply made all values positive and hence familiar to lay people used to indices of this type (IQ and SAT scores, for example). Also, by creating an index of only positive values, the transformation gave us the ability to test nonlinear relationships between sprawl and outcomes (see below).

3) Standardized factor scores were summed and then adjusted for the population of the metropolitan area to obtain an overall sprawl index, also on a scale with a mean value of 100 and a standard deviation of 25.

In some of the technical literature on sprawl, and nearly all of the popular literature, sprawl is defined in terms of the size of a metropolitan area. Big metros are deemed more sprawling than small ones, simply by virtue of geographic size or population. At the same time, the largest metropolitan areas tend to have higher land values due to heightened competition for central locations, and with the higher land values go higher mean densities and intensities of development and other attributes associated with compact development. So we attempt to control for size in our final sprawl index by comparing the degree of sprawl that actually exists to what might be expected for a metropolitan area of a given size. This gives us a measure of overall sprawl that is independent of population size. The sprawl effect on quality-of-life outcomes can thus be easily distinguished from the size effect on quality-of-life outcomes.

The operational variables that together make up each dimension of sprawl are defined below, along with the data sources from which they came. The loadings of operational variables on individual factors are also shown. The Appendix contains detailed descriptions of the data sources tapped in this study.

4.2.1 Density Factor

Residential density is on everyone's list of sprawl indicators. To assess the degree of sprawl at the metropolitan level, average density can be computed for the urban sections collectively. Alternatively, densities can be computed for subareas, and the degree of metropolitan sprawl judged by the proportion of the metro above or below threshold densities.

Seven variables constitute the *density factor* developed for this study. The first four variables came from the U.S. Censuses of 1990 and 2000. Census tracts with very low densities, less than 100 persons per square mile, were excluded from the calculation of these variables to eliminate rural areas, desert tracts, and other undeveloped tracts that happen to be located within metro area boundaries.[14]

---

[14] Tracts with areas less than 0.001 square miles were also excluded. This was done because tracts with such small areas (including areas of zero) were felt to represent special situations, such as crews of vessels. We also felt that calculated densities for tracts with such small areas were likely to be misleading or meaningless.

dens gross population density in persons per square mile
l1500p percentage of population living at densities less than 1,500 persons per square mile, a low suburban density
g125cp percentage of population living at densities greater than 12,500 persons per square mile, an urban density that begins to be transit-supportive
dgcent estimated density at the center of the metro area derived from a negative exponential density function[15]

We derived one density variable from the U.S. Department of Agriculture's Natural Resources Inventory (NRI), by dividing metropolitan population by an estimate of urban and built-up land area. NRI provides land use data at 5-year intervals, including how much land is in urban uses. Thus, NRI allows us to estimate independent measures of urban area density (independent of the census, which uses population density and jurisdictional boundaries to establish thresholds between urban and non-urban areas). 1997 densities from NRI were taken as representative of the end of the decade, while 1987 densities were used to represent the beginning. This gave us a 10-year interval for calculating changes in densities, equivalent to the decennial census.

urbdn gross population density of urban lands

One density variable was derived from the national microdata sample of the American Housing Survey.[16] The national survey is conducted every two years. To reduce sampling error, data were pooled for 1997 and 1999 to represent the end of the decade, and for 1989, 1991, and 1993 to represent the beginning of the decade. The one density variable, average lot size of single family dwellings, comes as close to a net density measure as possible with a national data set. A

---

[15] Central density is just the intercept of a negative exponential density function. The function assumes the form:

$$D_i = D_o \exp(-b\, d_i).$$

where:

D_i = the density of census tract i
D_o = the estimated density at the center of the metropolitan area
b = the estimated density gradient or rate of decline of density with distance
d_i = the distance of the census tract from the center of the principal city

The function was estimated as follows. The principal cities of the metro areas were identified as the first-named cities in the 1990 definitions of those areas. Their centers were determined by locating central business district tracts within the principal cities as specified in the 1980 STF3 file. 1980 designations were adopted because central business districts were not designated in 1990 or 2000. The means of the latitudes and longitudes of the centroids of those central business district tracts were taken as the metropolitan centers. The distances from the centers to all tracts were calculated using an ArcView script. The Equidistant Conic projection method was used in the calculation of distances. Finally, a negative exponential density function was fit to the resulting data points to estimate the intercept and density gradient.

[16] AHS geography does not exactly conform to metro area definitions used in this study. Specifically, the AHS adopts MSA and CMSA boundaries for New England metros, and includes Niagara Falls in the Buffalo metro area and Vancouver in the Portland metro area.

weighted average value was used to adjust for different probabilities of sample selection in the original data set.

lot weighted average lot size in square feet for single family dwellings

The final component of the *density factor* relates to population centers identified by the Claritas Corporation from 1990 and 2000 censuses. Population centers are local density maxima to which other grid cells relate.[17] Their spheres of influence may cross metro area boundaries. For example, Jersey City had no population centers of its own in either census year, but instead fell within the spheres of influence of population centers in New York and Newark. A population center density variable was computed for each metro area as the weighted average density for all population centers within the given metro area. The average densities were weighted by the resident populations in the sphere of influence of each population center.

dncen weighted density of all population centers within a metro area

Principal components were extracted from this set of density-related variables, and the principal component that accounted for the greatest variance became our *density factor*. Factor loadings (that is, correlations of these variables with the *density factor*) are shown in Table 2. The 0 at the end of the variable name refers to 2000 (while a 97 refers to 1997 etc.). Again, the factor loadings are for 2000, as this is our base year.

The eigenvalue of the *density factor* is 4.57, which means that this one factor accounts for more of the variance in the original dataset than four of the individual variables combined. To be exact, the *density factor* accounts for almost two-thirds of the total variance in the data set. As expected, two of the variables load negatively on the *density factor*, those being the percentage of population living at less than 1,500 persons per square mile (*l1500p0*) and the average lot size of single-family dwellings (*lot9799*). The rest load with positive signs. Thus, for all component variables, higher densities translate into higher values of the density factor.

Table 2. Variable Loadings on the *Density Factor*

---

[17] Claritas divided the US into 900,000 cells of 1/30 degree longitude and latitude, or approximately 4 square miles each (where the exact area depends on the latitude). Densities were computed for all them, and local density maxima defined as cells whose densities are greater or equal to those of all the grid cells surrounding them or in the second ring around them (approximately a 5-mile radius). A local density maximum was treated as a population center for another cell if a route could be constructed from the latter to the former, traveling cell by cell in any of eight possible directions along the grid, in which the density of each successive cell always increased or remained equal, and that route was shorter than that to all other competing local density maxima. Following Claritas' lead, we used a low density threshold to qualify a 9 grid cell area as a population center—850 persons per sq mi. Defined this way, the most dense population centers are at the hearts of big cities, while the least dense are very rural. The trip northward from Washington DC takes you, for example, through the population centers of Silver Spring, Bethesda, Cabin John, Potomac, Clinton, and Upper Marlboro, with local density maxima declining along the way. For more on the Claritas database, see David Miller and Ken Hodges, "A Population Density Approach to Incorporating an Urban-Rural Dimension into Small Area Lifestyle Clusters," paper presented at the Annual Meeting of the Population Association of America, Miami, FL, 1994.

dens0	0.89
l1500p0	-0.69
g125cp0	0.94
dgcent0	0.90
urbdn97	0.94
lot9799	-0.30
dncen0	0.81
eigenvalue	4.57
% of variance	65.31

4.2.2 Mix Factor

Three types of mixed-use measures are found in the "land use impacts on travel" literature: those representing relative balance between jobs and population within subareas of a region; those representing the diversity of land uses within subareas of a region; and those representing the accessibility of residential uses to nonresidential uses at different locations within a region.[18] All three types were estimated for metropolitan areas in our sample and became part of the *mix factor*.

We derived the first three mixed-use variables from the national microdata samples of the American Housing Survey.[19] Samples were pooled and weighted as described previously.

ecom percentage of residents with businesses or institutions within _ block of their homes

shop percentage of residents with satisfactory neighborhood shopping within 1 mile
sch percentage of residents with a public elementary school within 1 mile

We derived three additional mixed-use variables from the Census Transportation Planning Package (CTPP) for 1990.[20] CTPP is the only census product that summarizes data by place of work as well as by place of residence; it alone permits us to measure the degree of balance between employment and population (jobs and residents) for subareas of metros, as well as the

---

[18] Reid Ewing and Robert Cervero, "Travel and the Built Environment," *Transportation Research Record 1780*, 2001, pp. 87-114.
[19] One additional mixed-use variable was available from AHS, the percentage of residents with commercial activity on the same property as their home. It correlates highly with some of the other mixed-use variables, lending them legitimacy. This variable was dropped from the *mix factor* because it is only available for 1990 and has such small values that even a fraction of a percentage change has a big effect on factor scores.
[20] Insofar as possible, CTPP data were made to conform to 1990 census definitions of metropolitan areas. Every metro is a separate story. In all cases where MPO study areas extended beyond metro boundaries, CTPP data files were pared back to equal or approximate 1990 census boundaries. In cases where MPO study areas were less extensive than 1990 census boundaries, we had no option but to use the MPO boundaries.

degree of employment mixing for subareas. For most metros in the sample, the subareas are traffic analysis zones (TAZs); for a few they are census block groups or census tracts.

Until the 2000 CTPP is released, the 1990 CTPP provides our best estimates of the degree of land-use balance and mixing within metropolitan areas. Given the relatively slow rate of change in metropolitan land-use patterns, and the use of weighted measures of balance and mix (weighted by population and employment), 1990 values should be reasonable proxies for conditions in 2000.

We derived two balance variables from the CTPP. One measures the degree of balance within TAZs between jobs and residents, where balance equals 1 for TAZs with the same ratio of jobs-to-residents as the metro as a whole; 0 for TAZs with only jobs or residents, not both; and intermediate values for intermediate cases.[21] Another variable, analogous to the first, measures the degree of balance between population-serving jobs and residents; sectors considered population-serving are retail, personal services, entertainment, health, education, and other professional and related services.

We also derived a job mix variable. The mix variable equals 1 for TAZs with equal numbers of jobs in each sector; 0 for TAZs whose jobs are concentrated in a single sector; and intermediate values for intermediate cases.[22] This type of variable, derived from an entropy formula, has become common in the land use-travel literature. The sectors considered in this case were the same as for the second job-resident balance variable, that is, retail, personal services, entertainment, health, education, and other professional and related services.

---

[21] The equation used to calculate job-resident balance was:

$$\sum_{i=0}^{i=n} (1 - (ABS(J_i - JP*P_i))/(J_i + JP*P_i)) * ((J_i + JP*P_i)/(TJ + JP*TP))$$

where:

i = TAZ number (usually a traffic analysis zone)
n = number of TAZs in the metro area
J = jobs in the TAZ
JP = jobs per person in metropolitan area
P = residents in the TAZ
TJ = total jobs in the metropolitan area
TP = total residents in the metropolitan area

[22] The equation for this measure is:

$$\sum_{i=1}^{n} -\sum_{j} ((P_j * LN(P_j))/LN(j)) * (RJ_i / TRJ)$$

where:

i = TAZ number
n = number of TAZs in the metro area
j = number of sectors
P_j = proportion of jobs in sector j
RJ = number of retail and total personal services jobs in the TAZ
TRJ = total number of retail and total personal services jobs in the metropolitan area

CTPP variables were weighted by population and employment of TAZs and normalized by adjusting for average TAZ size. The larger the TAZ, the greater the apparent degree of balance and mix regardless of actual development patterns. The increase in balance and mix with size appears to follow a logarithmic curve; thus, to normalize, absolute values were divided by the natural log of jobs plus residents per TAZ in each metro. The resulting mixed-use variables were:

nbal job-resident balance
nrbal population-serving job-resident balance
nrent population-serving job mix (entropy)

Principal components were extracted from the set of mix-related variables, and the principal component that accounted for the greatest variance became our *mix factor*. Loadings of these variables on the *mix factor* are shown in Table 3. While all variables have positive relationships to the *mix factor*, as they should, this construct was not as fully operationalized as was density. The first principal component extracted, our *mix factor*, accounts for only a little over one third of the combined variance, the equivalent of two operational variables. One variable, population-serving job mix (*nrent*), is only marginally represented by the *mix factor*.

Table 3. Variable Loadings on the *Mix Factor*

ecom9799	0.60
shp9799	0.36
sch9799	0.52
nbal9	0.85
nrbal9	0.87
nrent9	0.13
eigenvalue	2.25
% of variance	37.47

4.2.3 Centers Factor

Metropolitan centers are concentrations of activity that provide agglomeration economies, support alternative modes and multipurpose trip making, create a sense of place in the urban landscape, and otherwise differentiate compact metros from sprawling ones.[23] Centeredness can exist with respect to population or employment, and with respect to a single dominant center or multiple subcenters. The technical literature associates compactness with centers of all types, and sprawl with the absence of centers of any type.

Six operational variables made up the *centers factor*. Two came from the U.S. Censuses of 1990 and 2000. One was just the coefficient of variation in tract densities, defined as the standard deviation of census tract densities divided by the mean density. The more variation in densities

[23] Reid Ewing, "Is Los Angeles-Style Sprawl Desirable?" *Journal of the American Planning Association*, Winter 1997, pp. 107-126.

around the mean, the more centering and/or subcentering exists within the metro area. The other census variable was the density gradient moving outward from the metropolitan center, estimated with a negative exponential density function. The faster density declines with distance from the center, the more centered (in a monocentric sense) the metropolitan area will be.[24]

coefvr coefficient of variation of population density across census tracts (standard deviation divided by mean density)
dggrad density gradient (rate of decline of density with distance from the center of the metro area)

We represented the degree of centralization of employment within the metropolitan area by two variables borrowed from the work of Glaeser et al. (see literature review above). For the 100 largest U.S. metropolitan areas, they calculated the share of overall metropolitan area employment within a three-mile ring of the Central Business District, the share of metropolitan area employment within a 10-mile ring of this spot, and the share beyond the 10-mile ring. All were measured for 1996. Two of the three variables became part of our *centers factor*:

l3emp percentage of metropolitan population less than 3 miles from the CBD
g10emp percentage of metropolitan population more than 10 miles from the CBD

The last two variables contributing to the *centers factor* came from Claritas databases.[25] Claritas identified population centers and their spheres of influence. Each block group within a metropolitan area was related to: a population center in the same metro; a population center in a different metro; a population center outside all metropolitan areas; or no population center at all. Most of the population of Akron, for example, relates to medium density centers in the Akron metropolitan area; a portion on the north side relates to higher density centers in Cleveland, part of the same CMSA; a portion on the east side relates to centers in Youngstown; and a little bit does not relate to centers at all.

From this database we derived two additional variables with clear relationships to centeredness:

popcen percentage of the metropolitan population relating to centers or subcenters within the same MSA or PMSA
rdnap1c ratio of the weighted density of population centers within the same MSA or PMSA to the highest density center to which a metro relates

Factor loadings are shown in Table 4. The *centers factor* has the expected relationships to all its component variables. It is positively related to all but two variables: the density gradient (*dggrad*) and the percentage of employment more than 10 miles from the center (*g10emp*), both of which assume higher values in decentralized metro areas. As with the *mix factor*, the first principal component accounts for only about one-third of the variance in the original data set. In this sense, the construct of metropolitan centers is not as fully operationalized as is the construct

---

[24] See footnote 15 for information on how the gradient was estimated.
[25] The Claritas databases are described in footnote 17.

of density. Two operational variables, in particular, are only marginally represented by the *centers factor*, those being *coefvr* and *popcen*.

Table 4. Variable Loadings on the *Centers Factor*

coefvr0	0.21
dggrad0	-0.74
l3emp96	0.76
g10emp96	-0.76
popcen0	0.17
dnap1c0	0.48
eigenvalue	2.01
% of variance	33.49

4.2.4 Streets Factor

Street networks can be dense or sparse, interconnected or disconnected, straight or curved. Blocks carved out by streets can be short and small, or long and large. Sparse, discontinuous, curvilinear networks creating long, large blocks have come to be associated with the concept of sprawl, while their antithesis is associated with compact development patterns.

There is no practical way, from national data sources, to quantify the degree of connectedness or curvature in metropolitan street networks. However, from U.S. Census TIGER files, block lengths can be determined. And from the U.S. Census Summary Files, block size is known. To a degree, block size captures not only the length of block faces but the extent to which streets are interconnected, as suburban superblocks with branching streets ending in cul-de-sacs may appear fairly dense and short-blocked, but are still large in total area.

Initially, from Census TIGER files, we tallied street centerline miles and street segments for entire counties, and from these, computed average block lengths. The resulting network measure was inflated by large portions of many metropolitan counties that are undeveloped. Therefore, we recalculated average block lengths considering only the streets within urbanized area boundaries for 1990 and 2000. The resulting network measure is more representative of the places where most residents live and work. Changes in the criteria used to define urbanized areas between the 1990 and 2000 censuses mean that this network measure is not entirely equivalent for the two census years. When boundary files become available for 1990 urbanized areas using 2000 criteria (in early 2003), consistent area definitions will be applied.

We tabulated block sizes and derived two measures for each metropolitan area. One was the average block size, the other the proportion of blocks one hundredth of a square mile or less in size (which is a traditional urban block, a little more than 500 feet on a side). Reviewing the data, it became obvious that huge rural tracts could distort averages and should be excluded from the calculation. A ceiling of 1 square mile, the size of a standard section and large superblock, was established for this purpose. This resolved the same issue as above, that many metropolitan areas contain large rural tracts unrepresentative of the places most residents live and work.

The following variables became components of the *streets factor*.[26] Factor loadings are shown in Table 5. All variables have the expected relationships, whether positive or negative, to this factor. Block length is described as "approximate" because not all street segments in the TIGER files end at intersections.

bklnu approximate average block length in the urbanized portion of the metro
bksz average block size in square miles (excluding blocks > 1 square mile)
smbk percentage of small blocks (< 0.01 square mile)

Table 5. Variable Loadings on the *Streets Factor*

bkln0u	-0.83
bksz0	-0.86
psmbk0	0.92
eigenvalue	2.28
% of variance	76.0

4.2.5 Four Factors

The result of all this effort is a set of four sprawl factors. As already described, the four factors were standardized on scales with mean values of 100 and standard deviations of 25. The factors are:

denfac0 *density factor* for 2000 (a weighted combination of 7 density variables)
mixfac0 *mix factor* for 2000 (a weighted combination of 6 mixed-use variables)
cenfac0 *centers factor* for 2000 (a weighted combination of 6 center-related variables)
strfac0 *streets factor* for 2000 (a weighted combination of 3 street-related variables)

The four factors represent a *balanced scorecard* of sprawl indicators, measuring independent dimensions of the phenomenon. Density and mix, while correlated, are very different constructs. Centeredness and street accessibility are as well.

4.2.6 Overall Sprawl Index

The next issue we had to wrestle with was how to combine the four factors into a single sprawl index. A priori, there is no "right" way to do so, only ways that have more or less face validity.

Should the four factors be weighted equally, or should one or another given more weight than the others? Density has certainly received more attention as an aspect of sprawl than has, say, street accessibility. However, beyond play in the literature, we could think of no rationale for differential weights. All factors contribute to the accessibility or inaccessibility of different

[26] One additional variable, intersection density, correlates highly with the other street variables, lending them legitimacy. It was dropped from the streets factor because it is only available for 1992.

development patterns, none presumptively more than the others. Depending on their values, all move a metropolitan area along the continuum from compact development to sprawl. Thus they were simply summed, in effect giving each dimension of sprawl equal weight in the overall index.

The final and most difficult issue was whether to, and how to, adjust the resulting sprawl index for metropolitan area size. As metro areas grow, so do their labor and real estate markets, and their land prices. Their density gradients accordingly shift upward, and other measures of compactness (street density, for example) follow suit. The simple correlation between the sum of the four sprawl factors and the population of the MSA or CMSA is 0.4, significant at .001 probability level. Thus, the largest metro areas, perceived as the most sprawling by the public, actually appear less sprawling than smaller metros when sprawl is measured strictly in terms of the four factors, with no consideration given to size.

Some of the technical literature on sprawl includes size in the definition.[27] Certainly, sheer geographic size is central to popular notions of sprawl. Despite their relatively high densities, metro areas such as Los Angeles and Phoenix, and even Chicago and Philadelphia, are perceived as sprawling because they "go on forever." A sprawl index that disregarded this aspect of urban form would never achieve face validity.

Accordingly, we sought a method of transforming the sum of the four sprawl factors into a sprawl index that would be *neutral* with respect to population size. This transformation was accomplished by regressing the sum of the four sprawl factors on the population of the metro area (either MSA population, or CMSA population for metros that are part of larger consolidated metropolitan statistical areas). The standardized residuals (difference between actual and estimated values divided by the standard deviation of the difference) became our overall measure of sprawl. Given the way it was derived, this index is uncorrelated with population. Metro areas that are more compact than expected, given the population of their MSA or CMSA, have positive values. Metro areas that are more sprawling than expected, again given the population of their MSA or CMSA, have negative values. This adjustment for population size still leaves the sprawl index highly correlated with the sum of the four component factors ($r = 0.92$).

As with the individual sprawl factors, we transformed the overall sprawl index (*index*) into an index with a mean of 100 and a standard deviation of 25. This was done for the sake of consistency and ease of understanding. With this transformation, the more compact metro areas have *index* values above 100, while the less compact have values below 100.

4.3 Sprawl Ratings for 2000

Table 6 presents sprawl ratings for metro areas in 2000. The individual dimensions of sprawl are displayed in columns two through five, and the overall sprawl index appears in the last column. Figure 2 provides the same data in graphic form.

The highest ratings on the *density factor* go mostly to the central PMSAs of large CMSAs. The New York PMSA is in a class by itself, having a factor score more than five standard deviations above the mean. While the smaller Jersey City PMSA ranks second, this is followed by other large PMSAs: San Francisco, Los Angeles, Chicago, and Miami PMSAs. Also high on the *density factor* are secondary PMSAs of these same CMSAs: Anaheim, San Jose, Newark, Oakland, and Ft. Lauderdale. Their large housing and labor markets drive up the bid rent curves of these CMSAs, making accessible central locations particularly valuable. Valuable land is naturally developed at higher densities, as housing producers and consumers both seek to minimize expensive land inputs. The simple correlation of the *density factor* for 2000 with the population of the MSA or PMSA is high (r = .614).[28] The simple correlation with the population of the MSA or CMSA is even higher (r = .653).

At the bottom of density rankings are medium-size metros in the Southeast, in ascending order: Knoxville, TN; Greenville--Spartanburg, SC; Greensboro--Winston-Salem--High Point, NC; Columbia, SC; Raleigh--Durham, NC; and Birmingham, AL. These are places whose growth has mostly occurred during the automobile era, and has been without topographic or water-related constraints that restrict development elsewhere in the Sunbelt. Still, the clustering of low densities in this particular region is striking and requires further investigation.

The *mix factor* is moderately correlated with the *density factor* (r = .443). This is to be expected, as higher densities naturally support mixed uses. At very high densities, where walking rivals the automobile as the mode of choice, a fine-grained mix of uses is required for reasonable walking distances. Yet, the *mix factor* is clearly capturing something distinct from density.

The metros with the greatest degree of land-use mixing are medium-sized and mostly concentrated in the Northeast. In descending order, the top five are: Jersey City, NJ; New Haven, CT; Providence, RI; Oxnard, CA; and Bridgeport, CT. The densest metros, New York, San Francisco, Los Angeles, Chicago, and Miami rank only, respectively, 8th, 35th, 12th, 28th, and 39th on the *mix factor*. The bottom of the mixed-use ranking is more in line with expectations, given the low densities of these same places. In ascending order they are: Raleigh--Durham, Riverside--San Bernardino, Greensboro--Winston-Salem--High Point, Greenville--Spartanburg, and West Palm Beach, FL.

The *centers factor* is the only factor to bear no relationship to density. The simple correlation with the *density factor* is only 0.06. In this sense it makes a unique contribution to our characterization of sprawl.

The *centers factor* measures two distinct conditions: the focus of development on downtown and central city, and the presence of subcenters within the metropolitan area. The former dominates the latter in the resulting rankings. With the exception of New York, the metros scoring highest on this factor are medium-sized and monocentric. In descending order, they are: Honolulu; Columbia, SC; Springfield, MA; and Providence. Others in the top 10 include Colorado Springs,

---

[27] Burchell et al., 1998, *op. cit.*

[28] This correlation is for 2000 densities and population, but as explained before, 1990 metro boundaries are used throughout this analysis.

Omaha, NE, and Wichita, KS. Other than New York, the only large, multi-centered metro near the top is San Francisco.

The metro areas ranking lowest on this factor mostly fall within the sphere of larger metropolitan areas. Among them are: Vallejo--Fairfield--Napa, CA; Riverside--San Bernardino, CA PMSA; Oakland, CA PMSA; and Gary--Hammond, IN PMSA. Two of the bottom 10 stand on their own, but have exceptionally weak downtowns: Tampa--St. Petersburg--Clearwater and Detroit. Los Angeles, whose downtown is also weak, just misses the bottom 10 in this ranking.

The *streets factor* is highly correlated with the *density factor* (r = .67). This is to be expected, since higher population and employment densities require more street capacity to meet travel demands. However, the relationship is not a simple linear one since high population and employment densities are associated with lower VMT per capita, and various strategies to mitigate congestion such as a shift to relatively efficient one-way street operation.[29] Thus, even the *streets factor* is distinct from the *density factor*.

The very highest ratings on the *streets factor* belong to older metropolitan areas: New York, Jersey City, San Francisco, and New Orleans. Behind them come some younger metropolitan areas that are developing at relatively high densities within their urbanized areas: Ft. Lauderdale, Anaheim, and Miami. At the other end of the spectrum are low density metros in the Northeast and Southeast. Syracuse, Rochester, and Hartford metros rank at or near the bottom. Atlanta, Greenville--Spartanburg, and Greensboro--Winston-Salem--High Point rank close behind.

As for the overall sprawl index (*index*), a few metro areas are compact in all dimensions. New York, San Francisco, Boston, and Portland fall into this group. They rank near the top in overall score. Others near the top, despite one factor score below average, include Jersey City, Providence, Honolulu, and Omaha. A few areas sprawl badly in all dimensions. These include Atlanta, Raleigh--Durham, Greensboro--Winston-Salem--High Point, NC, and Riverside--San Bernardino, CA. They rank at or near the bottom in overall score.

Table 6. Sprawl Ratings for 2000 in Order of Increasing Sprawl

	density factor	***mix factor***	***centers factor***	***streets factor***	***overall index***
New York	242.5	129.8	144.6	154.9	177.3
Jersey City	195.7	172.9	98.7	166.8	161.9
Providence-Pawtucket-Woonsocket	99.1	140.5	140.3	135.9	153.4
San Francisco	155.2	107.3	128.6	139.8	146.5
Honolulu	116.5	84.3	167.3	114.3	140.0
Omaha	96.4	119.3	132.3	104.6	128.2
Boston-Lawrence-Salem-Lowell-Brockton	113.6	124.4	109.4	119.1	126.8

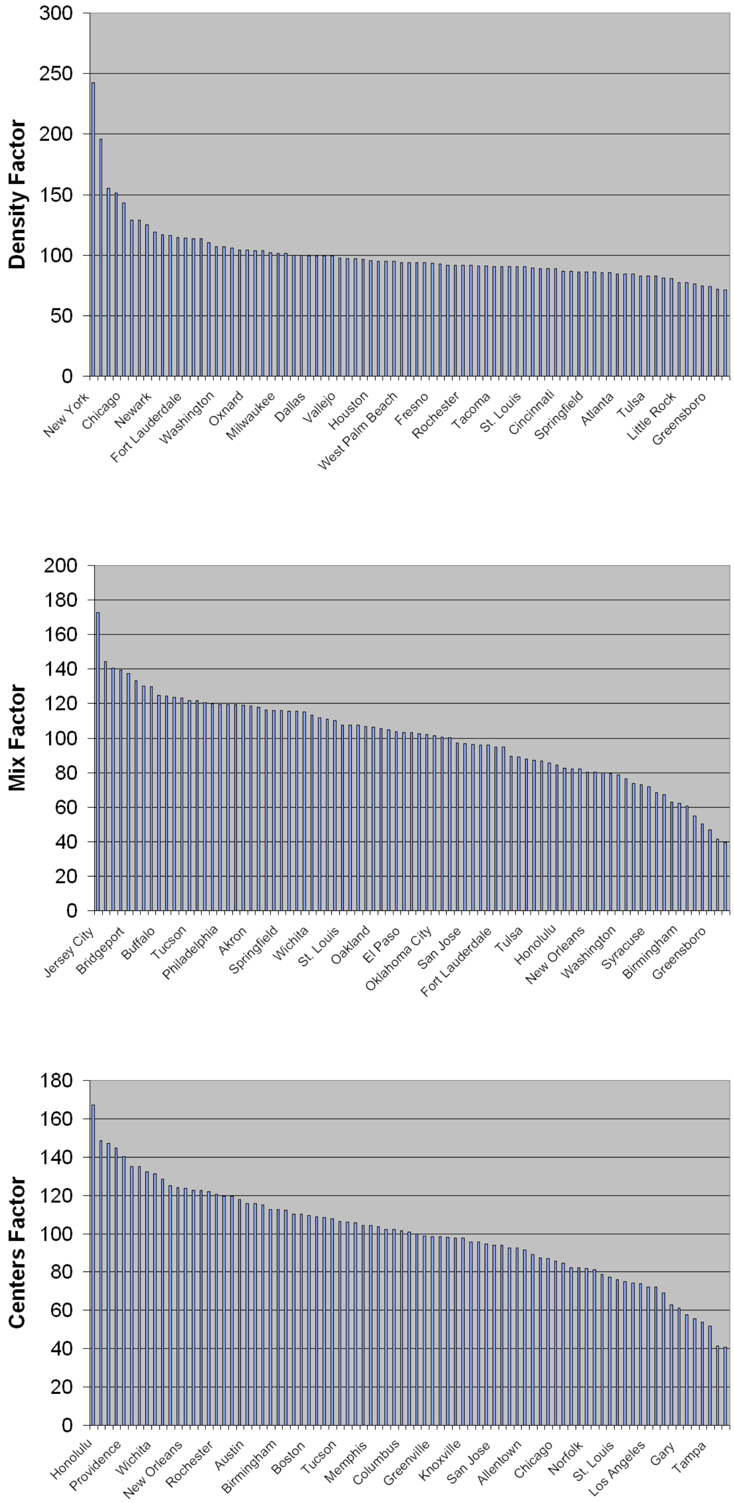
[29] Reid Ewing, "Sketch Planning a Street Network," *Transportation Research Record 1722*, 2000, pp. 75-79.

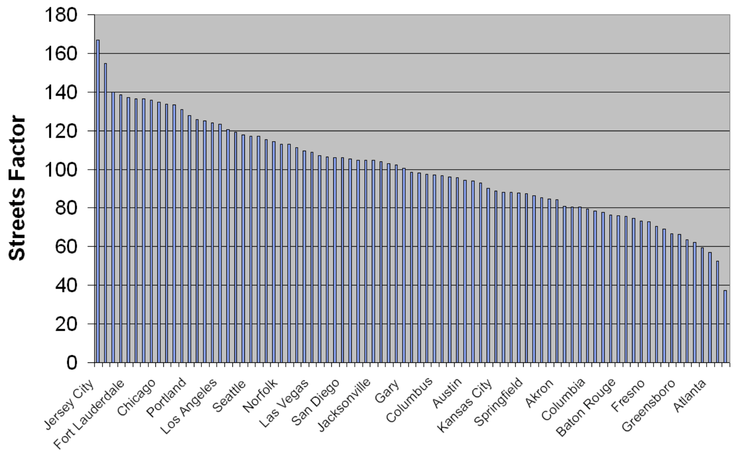
Portland	101.3	102.3	121.8	128.0	126.0
Miami-Hialeah	129.1	104.7	92.7	136.4	125.5
New Orleans	105.9	80.4	123.7	138.6	125.2
Denver	103.7	115.7	108.9	125.7	125.1
Albuquerque	97.0	103.7	124.0	117.8	124.3
Colorado Springs	91.2	119.0	135.2	96.7	124.3
Allentown-Bethlehem-Easton	86.2	133.4	91.7	131.0	123.9
Springfield	86.3	115.7	148.6	87.3	122.4
Chicago	142.9	115.1	85.8	134.9	121.1
Buffalo	102.1	124.7	135.2	70.6	119.0
Milwaukee	101.4	117.9	117.7	93.9	117.2
El Paso	100.1	103.4	119.5	102.3	117.1
Baltimore	104.3	106.8	115.6	105.2	115.8
Philadelphia	114.7	119.5	95.9	113.0	112.5
Phoenix	106.8	116.0	92.6	107.2	110.9
Salt Lake City-Ogden	99.5	103.2	93.8	117.0	110.9
Fresno	93.5	130.1	112.6	73.0	110.2
Austin	89.0	111.9	115.8	94.4	110.2
Wichita	84.4	113.1	131.4	78.6	110.0
San Jose	124.8	96.6	93.9	125.2	109.6
Tucson	90.4	121.8	106.4	88.0	109.1
Fort Lauderdale-Hollywood-Pompano Beach	113.9	94.7	75.0	137.2	108.4
San Antonio	95.0	100.6	108.4	103.0	107.7
Toledo	91.3	119.6	112.2	77.6	107.1
New Haven-Waterbury-Meriden	91.6	144.3	78.9	86.5	106.9
Pittsburgh	90.4	86.8	104.5	124.2	105.9
Tacoma	90.8	85.6	122.7	111.2	105.8
Akron	86.8	118.7	119.5	84.2	105.8
Las Vegas	110.0	80.1	99.8	108.8	104.7
Sacramento	99.1	110.9	87.4	98.4	102.6
San Diego	113.4	105.4	74.4	106.0	101.9
Los Angeles-Long Beach	151.5	123.1	72.4	123.3	101.8
Seattle	103.6	79.4	98.0	117.1	100.9
Tulsa	82.7	88.0	115.0	96.2	99.1
Oakland	116.6	106.3	57.6	133.4	98.8
Anaheim-Santa Ana	128.8	121.5	72.1	136.4	97.2

Orlando	93.8	60.8	103.5	120.6	96.4
Cincinnati	88.8	95.8	110.2	85.4	96.1
Minneapolis-St. Paul	94.7	94.7	107.8	87.7	95.9
Norfolk-Virginia Beach-Newport News	95.0	87.2	82.0	113.1	95.7
Grand Rapids	82.7	115.7	110.3	63.7	95.2
St. Louis	90.3	107.4	76.2	106.0	94.5
Columbia	74.6	67.1	147.3	79.5	94.2
Indianapolis	89.3	96.2	102.4	84.5	93.8
Houston	95.3	110.1	87.0	95.6	93.3
Memphis	88.9	97.0	104.2	76.5	92.2
Cleveland	99.7	107.4	100.9	66.8	91.8
Kansas City	90.9	100.0	89.0	88.8	91.7
Jacksonville	85.6	72.9	102.1	104.6	91.6
Columbus	91.5	76.5	101.5	97.2	91.2
Washington	106.9	78.7	97.8	98.0	90.9
Worcester-Fitchburg-Leonminster	81.2	82.3	122.7	74.5	90.5
Baton Rouge	80.8	95.9	106.2	76.2	90.2
Birmingham	77.1	62.2	112.5	104.0	88.0
Tampa-St. Petersburg-Clearwater	93.6	80.0	51.9	133.6	86.3
Oklahoma City	84.5	101.3	95.6	69.1	85.7
Hartford-New Britain-Middletown-Bristol	86.3	119.4	84.6	59.6	85.3
Albany-Schenectady-Troy	82.9	89.3	98.5	73.2	83.4
Little Rock-North Little Rock	77.5	68.3	105.9	88.2	82.4
Newark	118.9	120.4	82.2	115.4	81.4
Syracuse	85.8	72.0	124.9	52.6	80.4
Detroit	97.3	102.5	63.0	93.0	79.6
Vallejo-Fairfield-Napa	97.4	116.3	40.9	109.7	78.5
Dallas	99.5	82.6	81.1	90.2	78.4
Rochester	91.4	82.3	120.7	37.2	78.1
Gary-Hammond	86.4	123.7	61.2	100.5	77.5
Fort Worth-Arlington	90.3	89.1	73.9	97.5	77.4
Oxnard-Ventura	103.9	139.4	55.5	106.5	75.3
Knoxville	71.2	62.9	97.8	75.5	68.9
Bridgeport-Stamford-Norwalk-Danbury	92.5	137.5	94.8	80.7	68.6
West Palm Beach-Boca Raton-Delray Beach	94.0	54.7	53.9	104.7	67.9
Greenville-Spartanburg	71.9	50.4	98.5	62.1	58.8

Atlanta	84.5	73.7	82.3	57.0	57.9
Raleigh-Durham	76.2	39.5	77.2	80.8	54.5
Greensboro-Winston-Salem-High Point	74.2	46.7	69.1	66.3	47.1
Riverside-San Bernardino	93.5	41.5	41.4	80.5	14.7

Figure 2. Sprawl Ratings for 2000 in Order to Increasing Sprawl





4.4 Sprawl Trends from 1990 to 2000

We derived sprawl factors for 1990 by applying factor weights for 2000 to variable values for 1990. This produced sprawl factors for 1990 entirely consistent with the sprawl factors for 2000.

Thanks to consistent operational definitions, it became possible to compare the degree of metropolitan sprawl in 1990 and 2000, and to judge whether sprawl was getting better or worse over the decade. Table 7 presents changes in sprawl ratings between 1990 and 2000 for the three factors that lend themselves to such comparisons.

For one factor, it was not possible to measure changes over time. Of the six variables that make up the *mix factor*, three come from the 1990 CTPP and cannot be updated to 2000 until late 2002 when the new CTPP is released. The other three come from the AHS and are based on samples so small as to produce sizable sampling errors. Thus, we are left with three factors for which changes can be measured with a degree of consistency.

The sample of metros divides into three groups, a small number are *more* sprawling with respect to all three factors; a larger number are *less* sprawling with respect to all three factors; and the majority are becoming more sprawling in some respect but less in another. The "mores" include Akron, Ft. Worth, and Tampa, a hard group to categorize. The "lesses" are mostly fast growing metros concentrated in the West and Florida, including Anaheim, Phoenix, and Orlando. The mixed group spans the continent. Many are losing density and becoming less centered, but fall into the mixed category due to a rise in the *streets factor* accompanying subdivision of land. The phenomenon of sprawl, when measured in multi-dimensional terms over a span of years, is far more complex than most of the technical literature, and all of the popular literature, makes it out to be.

Table 7. Changes in Sprawl Ratings on Three Factors in Order of Increasing Sprawl

	change in density factor	***change in centers factor***	***change in streets factor***	***overall***

New York	5.04	-.04	3.00	mixed
Jersey City	1.14	-5.26	2.18	mixed
Providence-Pawtucket-Woonsocket	-3.04	14.04	.15	mixed
San Francisco	.55	-1.78	1.25	mixed
Honolulu	-9.37	-2.14	3.36	mixed
Omaha	-.99	-12.32	1.65	mixed
Boston-Lawrence-Salem-Lowell-Brockton	-2.33	.45	.05	mixed
Portland	.19	-12.31	-.59	mixed
Miami-Hialeah	3.49	-9.53	-5.86	mixed
New Orleans	-5.68	-2.33	2.98	mixed
Denver	-.09	2.66	7.52	mixed
Albuquerque	.75	-.29	10.66	mixed
Colorado Springs	1.60	-.20	7.96	mixed
Allentown-Bethlehem-Easton	-2.94	-5.76	-6.79	mixed
Springfield	-3.50	-1.44	8.42	mixed
Chicago	1.03	-1.60	8.64	mixed
Buffalo	-7.13	.93	1.18	mixed
Milwaukee	-4.01	-.88	1.24	mixed
El Paso	-2.55	5.09	-7.46	mixed
Baltimore	-3.22	-4.44	3.61	mixed
Philadelphia	-3.61	-1.33	-2.55	more
Phoenix	7.16	1.99	6.38	less
Salt Lake City-Ogden	1.82	-1.45	13.54	mixed
Fresno	-1.38	.46	-3.20	mixed
Austin	.30	-.83	8.06	mixed
Wichita	-6.01	3.03	6.11	mixed
San Jose	5.33	2.10	5.44	less
Tucson	2.64	7.74	-2.93	mixed
Fort Lauderdale-Hollywood-Pompano Beach	3.99	-5.59	5.18	mixed
San Antonio	-3.99	1.87	5.40	mixed
Toledo	-2.65	-1.69	3.49	mixed
New Haven-Waterbury-Meriden	-2.19	-.62	-1.70	more
Pittsburgh	-2.53	-1.94	5.91	mixed
Tacoma	2.75	-6.73	.25	mixed
Akron	-2.50	-3.13	-.73	more
Las Vegas	9.26	-33.32	13.43	mixed
Sacramento	1.50	.86	5.69	less
San Diego	1.62	-2.25	4.58	mixed
Los Angeles-Long Beach	4.11	-.42	6.13	mixed
Seattle	.21	-1.83	-1.81	mixed
Tulsa	-2.01	-2.38	13.16	mixed
Oakland	-1.52	-3.16	.52	mixed
Anaheim-Santa Ana	4.90	2.80	13.15	less
Orlando	4.66	.42	6.33	less
Cincinnati	-3.78	-5.51	8.69	mixed
Minneapolis-St. Paul	-2.42	-1.20	6.38	mixed

Norfolk-Virginia Beach-Newport News	-.99	-2.22	9.62	mixed
Grand Rapids	.77	1.54	2.34	less
St. Louis	-3.71	-7.21	6.90	mixed
Columbia	-4.53	-6.58	14.13	mixed
Indianapolis	-.29	-4.67	16.98	mixed
Houston	-1.12	1.08	5.83	mixed
Memphis	-4.02	-5.34	5.94	mixed
Cleveland	-4.27	-2.71	5.03	mixed
Kansas City	-.69	-4.47	10.27	mixed
Jacksonville	.95	.93	12.05	less
Columbus	-1.00	1.92	20.16	mixed
Washington	-3.79	-4.16	6.60	mixed
Worcester-Fitchburg-Leonminster	-3.79	4.30	-.48	mixed
Baton Rouge	.33	-1.55	.39	mixed
Birmingham	-4.90	-5.34	14.58	mixed
Tampa-St. Petersburg-Clearwater	-.67	-.74	-.26	more
Oklahoma City	-.93	.74	19.33	mixed
Hartford-New Britain-Middletown-Bristol	.63	-8.15	3.29	mixed
Albany-Schenectady-Troy	-4.03	-.86	1.45	mixed
Little Rock-North Little Rock	-.19	-1.99	9.48	mixed
Newark	2.00	-2.09	2.33	mixed
Syracuse	-4.61	-1.71	-2.83	more
Detroit	-5.63	-.95	3.74	mixed
Vallejo-Fairfield-Napa	-.68	3.22	5.19	mixed
Dallas	4.28	1.42	10.20	less
Rochester	1.55	.85	-.33	mixed
Gary-Hammond	-.36	-9.18	8.36	mixed
Fort Worth-Arlington	-.59	-.24	-4.43	more
Oxnard-Ventura	3.35	.70	2.33	less
Knoxville	-1.71	-3.70	9.71	mixed
Bridgeport-Stamford-Norwalk-Danbury	.32	-1.31	-2.49	mixed
West Palm Beach-Boca Raton-Delray Beach	2.02	2.34	-3.85	mixed
Greenville-Spartanburg	2.13	-.95	3.26	mixed
Atlanta	3.05	-3.49	11.58	mixed
Raleigh-Durham	-2.77	.90	14.40	mixed
Greensboro-Winston-Salem-High Point	-.54	.31	4.68	mixed
Riverside-San Bernardino	4.76	-1.18	8.84	mixed

4.5 Illustrating Differences

The richness of the sprawl database gives us the ability to analyze urban form at different levels of detail. Tucson, Arizona, and Ft. Lauderdale, Florida have similar overall sprawl scores for the year 2000: Tucson at 109, and Ft. Lauderdale at 108. Both are a bit more compact than average for their size. Yet they arrive at these scores in very different ways.

Tucson has large blocks and very low-density housing. Tucson's score for street accessibility is 88, ranking it 29th most sprawling in terms of its street layout. One indication of poor street accessibility is the size of its blocks: in Tucson only 45 percent of blocks are less than a hundredth of a square mile, or about 500 feet on a side. Tucson's housing is also extremely spread out: the metro area scored 90 on the residential *density factor* in part because its average urban density is only 1,767 persons per square mile, one of the lowest of all metros in our sample. Yet being a standalone metro, Tucson's growth has remained focused on its own centers (rather than relating to centers in neighboring counties, as in Ft. Lauderdale); and the presence of mountains ringing the Tucson valley has kept nearly all employment within 10 miles of downtown. In degree of centering, Tucson gets an above-average score of 106. Tucson also does well in its mix of homes, offices, stores, and other uses, scoring 121 on this scale.

Ft. Lauderdale's blocks are smaller than Tucson's, and its housing is denser. Ft. Lauderdale scores 137 on the *streets index*; 68 percent of its blocks are less than one hundredth of a square mile, one of the highest percentages in our sample. It also has higher-than-average residential density, with an average urban density of 4,837 persons per square mile, way above average for our sample. But, offsetting these factors, Ft. Lauderdale's degree of centering is below average; the metro area scored just 75 on this measure, making it the 14th most sprawling place in this regard. It has a weaker than average downtown for its size, few significant subcenters, and more than a third of its population relating to centers outside the metropolitan area. Only 15 percent of its employment falls within a 3-mile ring of the central business district. It also places homes and workplaces farther apart than average, scoring 94 on the *mix factor*.

5 Investigating Outcomes of Sprawl

Ultimately, sprawl must be judged by its outcomes. No development pattern is inherently good or bad. It is negative outcomes, such as high vehicle miles traveled or severe congestion, that make one development pattern superior to another. We are fortunate to have among our outcome measures recently released journey-to-work data from the 2000 U.S. Census, and just released congestion data from the Texas Transportation Institute.

5.1 Statistical Method

Correlational studies, of which this is one, cannot be used to establish cause-effect relationships between dependent and independent variables. But they can establish statistically significant associations between variables, a necessary condition for causality. If studies, in addition, control for other influences on dependent variables, and still find strong associations with independent variables, then it becomes easier to justify the contention that one variable causes or contributes to another.

Given the aggregate nature of this analysis, the statistical method of choice, used to test for significant relationships, is multiple regression analysis. We tested for significant relationships by running a series of regressions for travel and transportation outcomes in 2000. In the first set of regressions, we regressed each outcome measure on the *overall sprawl index* and a standard set of control variables to establish the existence of a relationship between sprawl and the outcome. In the second set of regressions, we regressed each outcome measure on the full set of sprawl factors (all four) and a standard set of control variables to explore the nature of the relationship between sprawl and the particular outcome.

The challenge in this kind of research is to control for *confounding influences*. These are variables that are not of primary interest, and may not even be measured, but influence outcomes in ways that may confound results. Multiple regression analysis captures the independent effect of each variable on the outcome of interest, controlling for the effects of all other variables in the regression equation. The use of multiple regression analysis allows us to control for confounding influences, provided that they are measured and included in the regression equation.

As already noted, associations that emerge from the regressions do not necessarily mean that one thing causes another, only that the two are related. On theoretical grounds, we might expect causality to flow in *both directions* for some of our outcome measures. Consider the outcome measure "average auto ownership." Due to lack of alternatives to travel by automobile, a sprawling development pattern would be expected to produce high auto ownership rates. At the same time, higher auto ownership rates may beget sprawl by encouraging more and longer auto trips by household members, who have less need to share autos or coordinate trip making if households have more vehicles at their disposal. Dual causality could also apply to sprawl and the outcome measure "transit mode share," since compact development patterns are both cause and effect of high transit usage. Subsequent research will experiment with simultaneous equation structures in which sprawl and outcomes are estimated jointly, each serving as an independent variable for the other.

5.2 Outcome Measures

Outcomes attributed to sprawl are summarized by Burchell et al.[30] Those related to travel and transportation became the dependent variables in our analyses. Several travel and transportation outcome measures were derived from the U.S. Censuses of 1990 and 2000.

transhr	percentage of commuters using public transportation (including taxi)
walkshr	percentage of commuters walking to work
mntime	mean journey-to-work time in minutes

From the Texas Transportation Institute (TTI) mobility database came traffic congestion data for 1990 and 2000. TTI data apply to urbanized areas rather than metros. They are available for 60 urbanized areas corresponding to our final sample of 83 metropolitan areas. Several of the urbanized areas incorporate multiple metropolitan areas, and some take in far more territory than the largest metropolitan area in the corresponding CMSA. The entire New York area, for example, is lumped together in the TTI database. For 55 urbanized areas, the correspondence between urbanized area and metro area is close enough to retain these cases for subsequent analysis.[31]

---

[30] Burchell et al., 1998, *op. cit.*

[31] The urbanized areas dropped from the sample were: New York-Northeastern NJ; Chicago-Northwestern IN; San Francisco-Oakland; Los Angeles; and Dallas-Ft. Worth.

TTI measures are computed rather than measured in the field, and so are no better than the formulae upon which they are based. Of the several measures of congestion in the TTI database, the one that is most intuitively understandable was selected for analysis.

dlycap annual hours of delay per capita

From the Federal Highway Administration's Highway Performance Monitoring System (HPMS) came VMT (vehicle mile traveled) data for 1990 and 2000. Like TTI data, HPMS data apply to urbanized areas rather than metros. We had to piece together values for metro areas from the urbanized areas that make them up. VMT and population were estimated for each urbanized area that has land within a given metro, with estimates based on the proportion of an urbanized area's total land area which falls within metropolitan boundaries. These estimates were summed over all urbanized areas in a given metro, and a weighted average VMT per capita thereby derived. For example, the Dayton-Springfield, OH metropolitan area contains all of the Springfield urbanized area, and nearly all of the Dayton urbanized area. The final VMT per capita estimate for this metro area included all of the Springfield VMT and population, and 96 percent of the Dayton VMT and population.

Pieced together this way, HPMS data were available for 77 metros in our final sample of 83 metros. The correspondence between urbanized areas and metro areas was close enough to retain 72 of these urbanized areas for subsequent analysis.[32]

vmtcap daily VMT per capita

From the National Highway Traffic Safety Administration's Fatal Accident Reporting System (FARS) came highway fatality data for 1990 and 2000. Because data are available for all counties in the U.S., fatal accident rates can be computed for metro areas exactly as defined in 1990, the reference year for our metro area definitions.

facap annual fatal highway accidents per 100,000 persons

While not strictly transportation-related, the final outcome measure relates to the maximum ozone level in the metro area, a criteria pollutant closely linked to motor vehicle use.[33] Ozone was selected for analysis over carbon monoxide because the former manifests itself regionally rather than only in local hot spots. Values are for 1990 and 1999, the latter being the most recent year for which metropolitan trend data are available.

oz8h fourth highest daily maximum 8-hour average ozone level

---

[32] For 1990, there were 75 urbanized areas to begin with and 70 after five were dropped. For both 1990 and 2000, the five dropped were the same for HPMS data as for TTI data. See footnote 31.

[33] EPA tracks trends in air quality based on actual measurements of pollutant concentrations in the ambient (outside) air at monitoring sites across the country. Monitoring stations are operated by state, tribal, and local government agencies as well as some federal agencies, including EPA. Trends are derived by averaging direct measurements from these monitoring stations on a yearly basis.

Notably absent from this list are outcome measures from the Nationwide Personal Transportation Survey (NPTS). 1995 is the most recent year for which NPTS is currently available. Until an expanded version of NPTS, conducted in 2001 and called the National Household Travel Survey, is released in late 2002, we defer the analysis of mode shares for all trip purposes, outcome measures that are only available from this one source.

5.3 Control Variables

The following variables were used to control for influences on travel other than those of the built environment:

metpop	metropolitan area population (MSA or PMSA)
hhsize	average household size for the metro area
pwkage	percentage of population of working age in the metro area (20-64 years)
pcinc	per capita income in the metro area

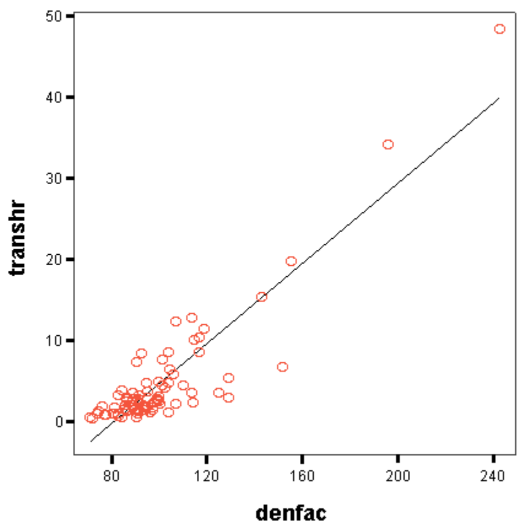
Transportation outcomes are arguably influenced in part by conditions beyond metropolitan area boundaries, especially for PMSAs that are parts of CMSAs. This is one reason why we standardized the sprawl index with respect to CMSA population, where applicable.

5.4 Outliers

Figure 3 presents a plot of one outcome, transit mode share on the journey to work, versus one sprawl factor, the *density factor*. In this plot, there are obvious outliers, having much higher transit mode shares and much higher densities than the other metros. These data points are outliers with respect not only to transit mode share, but also to most travel and transportation outcomes. As such they may exercise undue influence over the slopes of the regression lines. They also may make relationships between outcomes and density look stronger than they actually are.

Which cases, if any, should be dropped from the sample? Ultimately, following the rule of thumb that cases with leverage values around 0.2 or higher are problematic, the two most outlying cases with respect to the *density factor*, New York and Jersey City, were dropped from the sample. No other sprawl factor was similarly skewed toward outlying data points. Hence no other sprawl factor required exclusion of additional cases.

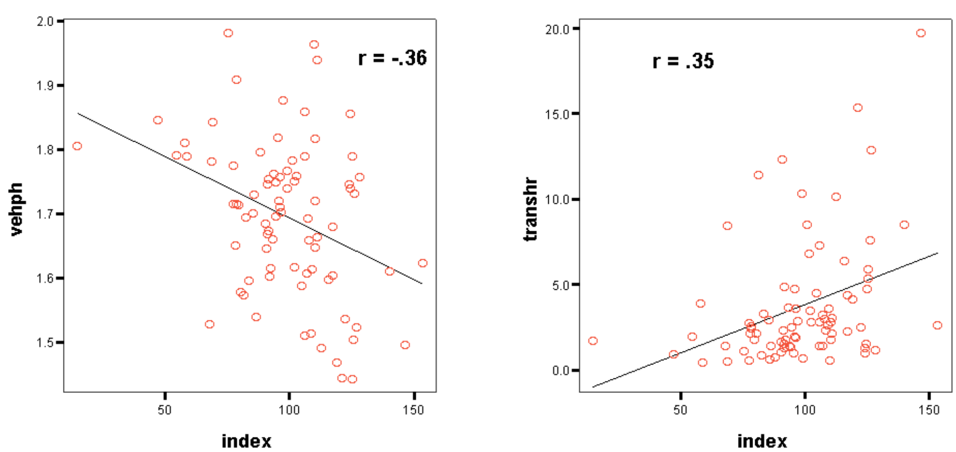
Figure 3. Plot of Transit Commute Share vs. *Density Factor*

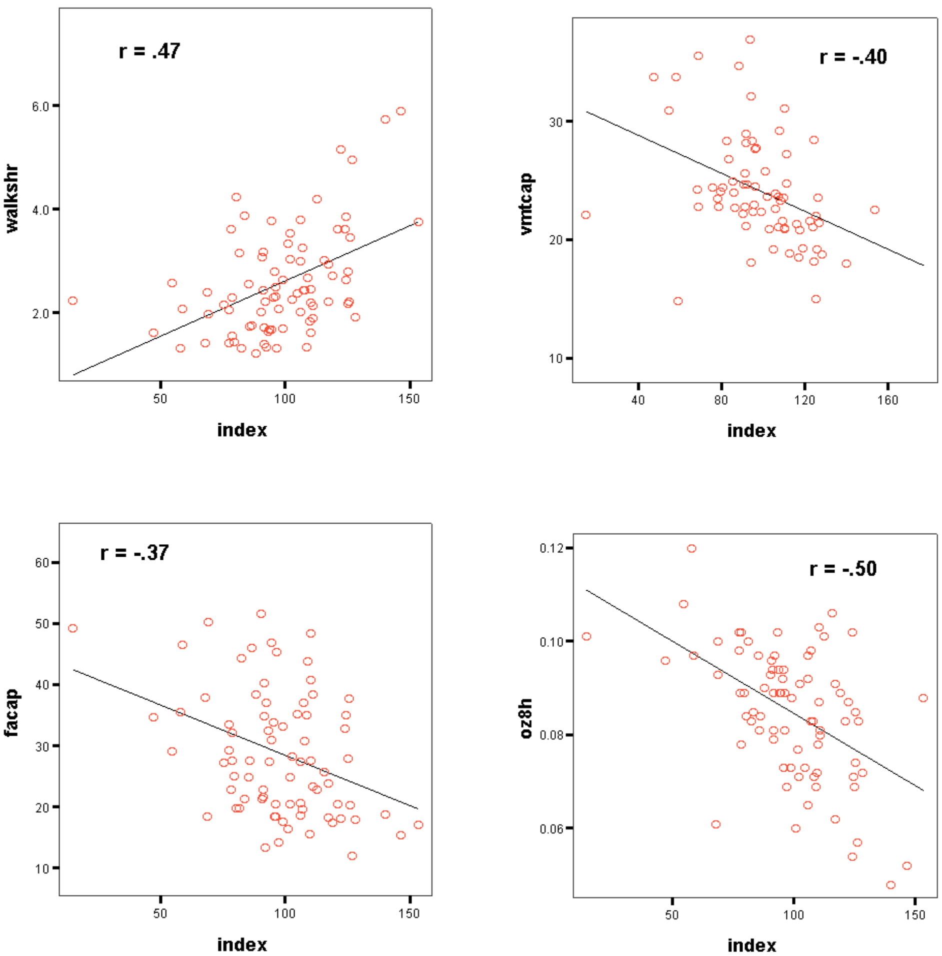


5.5 Associations with the Overall Sprawl Index (2000)

Our first pass analytically focused on the *overall sprawl index*. First we did scatterplots to check the assumption of linearity between the index and different outcomes. Without the two outlying metros, all *significant* relationships except perhaps that of the index and transit mode share appear fundamentally linear (see Figure 4).

Figure 4. Scatterplots, Regression Lines, and Simple Correlation Coefficients for Selected Outcomes





Then we regressed each outcome on the *index* plus the standard set of control variables. Results for 2000 are presented in Table 8. Regression coefficients and t-statistics appear across from their respective independent variables (with the t-statistics in parentheses). Adjusted R^2 statistics appear at the bottom of the table.

The *overall sprawl index* shows strong and statistically significant relationships to six outcome variables. All relationships are in the expected directions. As the *index* increases (sprawl decreases), average vehicle ownership, daily VMT per capita, annual traffic fatality rate, and maximum ozone level decrease to a significant extent. At the same time, shares of work trips by transit and walk modes increase to a significant extent.

The significance of these relationships rivals, or in some cases, actually exceeds that of the sociodemographic control variables. The *index* is the only variable that rises to the level of statistical significance for walk share of work trips and maximum ozone level, and has the strongest association to daily VMT per capita. It has secondary, but still highly significant, associations with average vehicle ownership, transit share of work trips, and annual traffic fatality rate.

Obviously, these relationships are not independent of each other. The lower vehicle ownership in dense metropolitan areas contributes to higher mode shares for alternatives to the automobile. These, in turn, contribute to lower VMT, which contributes to lower traffic fatalities and ozone levels. Due to different data sources, units of analysis, and sample sizes, it would be treacherous to model the causal paths among these outcome variables. But, intuitively, they should be related as indicated.

The *index* is not significantly related to either average commute time or annual traffic delay per capita. Both outcomes are a function primarily of metropolitan area population, and secondarily of other sociodemographic variables. Big metros generate long trips to work and high levels of traffic congestion. After controlling for population size and other sociodemographic variables, sprawl does not appear to have a marginal relationship to either outcome.

Why not? It has been argued that dispersal of jobs and housing allows residents to live closer to their workplaces than if jobs were more concentrated. It has also been argued that dispersal of jobs and housing eases traffic congestion by dispersing origins and destinations. These effects, if dominant, would lead to shorter work trips and less congestion in sprawling metro areas. But dispersal of jobs and housing may also result in jobs-housing imbalance, and significantly more VMT per capita than with more compact urban form. These forces may be canceling each other out, leaving no relationship between sprawl and average commute time or annual traffic delay per capita.

Table 8. Transportation Outcomes vs. Overall Sprawl Index (2000)

	Transportation Outcomes							
	vehph	***transhr***	***walkshr***	***mntime***	***dlycap***	***vmtcap***	***facap***	***oz8h***
constant	-0.11835	-19.84	-0.529	-0.291	-103.3	-0.500	39.30	0.162
index	-0.00189 (-4.2)***	0.0565 (4.8)***	0.0217 (4.8)***	-0.0170 (-1.6)	-0.0266 (-0.6)	-0.0782 (-3.6)***	-0.177 (-4.3)***	-0.0003 (-4.9)***
metpop	-3.4E-08 (-4.6)***	8.82E-07 (4.7)***	3.16E-08 (0.4)	9.48E-07 (5.4)***	2.84E-06 (3.1)**	3.74E-07 (0.8)	-8.6E-07 (-1.3)	1.51E-09 (1.5)
hhsize	0.306 (5.3)***	0.959 (0.6)	-0.128 (-0.2)	4.10 (2.9)**	11.96 (2.4)*	-2.51 (-0.9)	-3.08 (-0.6)	-0.0123 (-1.5)
wkage	0.0215 (3.9)***	0.0516 (0.4)	0.00456 (0.1)	0.144 (1.1)	1.18 (2.4)*	0.707 (2.8)**	0.767 (1.5)	-6.8E-05 (-0.1)
pcinc	-6E-07 (-0.2)	0.000476 (5.9)***	4.3E-05 (1.4)	0.00027 (3.6)***	0.000807 (2.2)*	-0.00016 (-1.0)	-0.00129 (-4.5)***	-6.1E-07 (-1.4)
adjusted R^2	0.43	0.59	0.22	0.51	0.52	0.21	0.33	0.24

* .05 probability level
** .01 probability level
*** .001 probability level

5.6 Associations with Individual Sprawl Factors (2000)

Our second pass analytically included all four sprawl factors. Each outcome was regressed on the four factors plus the standard set of control variables. Results for 2000 are presented in Table 9. Regression coefficients and t-statistics appear across from their respective independent variables (with the t-statistics in parentheses). Adjusted R^2 statistics appear at the bottom of the table.

The *density factor* has the strongest and most significant relationship to travel and transportation outcomes. It has a significant inverse relationship to average vehicle ownership, VMT per capita, traffic fatality rate, and maximum ozone level, and a significant direct relationship to public transportation and walk shares of commute trips. With the exception of the traffic fatality rate, all relationships are significant at the 0.01 probability level or beyond.

To illustrate the strength of density relationships, a 25-unit increase in the *density factor* (one standard deviation on the density scale) is associated with a 0.13 drop (25 x –0.00534) in average vehicles per household. That is, controlling for other factors, each standard deviation increase in density has the average household shedding 0.13 cars. With a range on the *density factor* of 3.4 standard deviations (excluding the two outlying metros, New York City and Jersey City), density alone is associated with nearly a one-half vehicle difference per household between high density and low density areas.

As another illustration of density's importance, a 25-unit increase in the *density factor* is associated with a 2.95 percentage point rise (25 x 0.118) in public transportation mode share on the journey to work. That is, controlling for other factors, each standard deviation increase in density increases public transportation mode share by almost 3 percentage points. With a range on the *density factor* of 3.4 standard deviations (again, excluding the outliers), density alone is associated with a 10 percentage point increase in public transportation use between high density and low density areas.

The *centers factor* has the next most significant environmental influence on travel and transportation outcomes. It is inversely related to annual delay per capita and traffic fatality rate, and is directly related to public transportation and walk shares of commute trips. These associations are in addition to (and independent of) those of density, which is controlled in the same equations.

With the exception of walk mode share for work trips, the relationships between degree of centering and outcomes are not as strong as the relationships between density and outcomes. Take the relationship between vehicle ownership and degree of centering. The degree of centering apparently affects the viability of other modes and the efficiency of auto use, which in turn affect vehicle ownership. But a 25-unit increase in the *centers factor* (one standard deviation on the centers scale) is associated with only a 0.03 drop (25 x –0.00117) in average vehicles per household, less than a fourth the change associated with the *density factor*.

Or consider the relationship between degree of centering and public transportation mode share on the journey to work. A 25-unit increase in the *centers factor* (one standard deviation on the centers scale) is associated with only a 0.88 percentage point rise (25 x 0.035) in public transportation mode share, just over one third the change associated with the *density factor*.

The *mix factor* is significant in only two cases, as a mitigating influence on travel time to work and an aggravating influence on the maximum ozone level. The latter relationship is just barely significant at the conventional level, and may be spurious. It does not show up in the 1990 regression analysis. Alternatively, it may be a real relationship, as a fine-grained mix may encourage more short vehicle trips, and hence more cold starts and hot soaks contributing to air pollution.

The big surprise is that land use mix does not significantly affect public transportation or walk mode shares for commute trips. There are two possible explanations, related to one another. Perhaps land use mix has not been successfully operationalized due to problems with the underlying datasets from the AHS and CTPP. Problems include the small samples sizes for some metros included in the AHS, and the imperfect correspondence between our metro area boundaries and those applied to CTPP data. Alternatively, land use mix may have been successfully operationalized but at a scale inappropriate for walk trips. Depending on the metropolitan area, CTPP uses traffic analysis zones, census block groups, or census tracts as its units of analysis. For two of three AHS variables, mixed use is measured in terms of the presence of activities within one mile of home. The geographic areas encompassed by these measures of land use mix may be too large, particularly in a suburban context, to distinguish walkable places from those that are not.

The *streets factor* was significant in two cases, albeit just barely and with unexpected signs. Average travel time for commute trips, and annual traffic delay per capita, are directly related to the *streets factor*. This runs counter to the expectation that higher values of this factor, which correspond to finer meshed street networks, would lead to shorter travel times and less delay. The potential for shorter trips is one argument (made by New Urbanists and others) for development of dense, interconnected street networks.

Perhaps the reason for this counterintuitive result is that the additional intersections in metros with dense street grids translate into more total delay, most delay being occasioned at intersections rather than on the stretches between them. Conventional traffic engineers have always argued as much. Another possibility is that the TTI delay measure, which is computed rather than measured, has sufficient error attached to it as to distort its relationships to street network measures. In any case, street patterns appear to be much less important than land use patterns as correlates of travel and transportation outcomes.

As for the control variables, they usually enter with the expected signs, often at significant levels. For example, average vehicle ownership rises with household size and percentage of working age population. The utility of owning an extra vehicle would be expected to increase with both sociodemographic variables, and from our results, it apparently does. Another example: The level of congestion, measured by annual delay per capita, increases with metropolitan area population, average household size, and percentage of working age residents. All of these relationships make intuitive sense.

Readers may have noticed anomalies in the results. While average commute time and annual delay per capita are not significantly related to the *overall sprawl index*, they are significantly related to individual sprawl factors. One possible explanation: Individual sprawl factors pull these outcomes in opposite directions, the *streets factor* being directly related to both outcomes while the *mix factor* is inversely related to average commute time and *centers factor* is inversely related to annual delay per capita. The two effects may cancel each other out.

Table 9. Outcomes vs. Sprawl Factors (2000)

	Transportation Outcomes							
	vehph	***transhr***	***walkshr***	***mntime***	***dlycap***	***vmtcap***	***facap***	***oz8h***
constant	-0.382	-14.11	0.566	4.77	-119.4	2.24	46.88	0.112
denfac	-0.00534 (-4.7)***	0.118 (3.9)***	0.0315 (2.6)**	-0.0245 (-0.9)	-0.110 (-0.9)	-0.215 (-3.0)**	-0.240 (-2.1)*	-0.0006 (-3.8)***
mixfac	0.000659 (1.5)	-0.00924 (-0.8)	0.00046 (0.1)	-0.0242 (-2.2)*	0.00728 (0.2)	0.00023 (0.0)	-0.0832 (-1.9)	0.00012 (2.0)*
cenfac	-0.00117 (-2.7)**	0.0351 (3.0)**	0.0199 (4.3)***	-0.0181 (-1.6)	-0.110 (-2.2)*	-0.0462 (-2.0)	-0.0945 (-2.1)*	-0.00012 (-1.9)
strfac	0.000492 (0.9)	0.00347 (0.2)	-0.00272 (-0.5)	0.0424 (3.2)**	0.130 (3.0)**	0.0128 (0.5)	0.0363 (0.7)	-0.00014 (-2.0)
metpop	-1.5E-08 (-1.7)	4.64E-07 (2.0)*	-1.7E-08 (-0.2)	8.53E-07 (4.0)***	2.05E-06 (2.2)*	8.72E-07 (1.6)	-2.2E-07 (-0.3)	4.27E-09 (3.5)***
hhsize	0.412 (7.0)***	-1.68 (-1.1)	-0.678 (-1.1)	4.32 (3.0)**	14.77 (2.9)**	1.76 (0.6)	3.75 (0.6)	0.00305 (0.4)
pwkage	0.0246 (4.5)***	-0.0207 (-0.1)	-0.0268 (-0.5)	0.0576 (0.4)	1.47 (3.0)**	0.667 (2.4)*	0.511 (0.9)	0.00047 (0.6)
pcinc	4.06E-06 (1.2)	0.00036 (4.0)***	2.6E-05 (0.7)	0.00029 (3.4)***	0.00075 (2.0)	3.01E-06 (0.0)	-0.00091 (-2.6)*	6.22E-08 (0.1)
adjusted R^2	0.56	0.67	0.36	0.61	0.63	0.28	0.36	0.40

* .05 probability level
** .01 probability level
*** .001 probability level

5.7 Validation (1990)

Thanks to the availability of complete datasets for 1990, it is possible to validate results for 2000. Recall that sprawl measures for 1990 are entirely consistent with measures for 2000. So are outcome measures (with the exception of VMT per capita, which applies to a slightly different set of urbanized areas). To validate the 2000 results, outcome measures for 1990 were regressed on the *overall sprawl index* for 1990 and, separately, on the four individual sprawl factors for 1990. These regressions controlled for the same sociodemographic variables as in 2000.

We first report on relationships between outcomes and the *overall sprawl index*. Comparing regression equations for 1990 to 2000, all but the constant terms are surprisingly similar. That is, the strength of relationships (magnitude of coefficients), direction of relationships (sign of coefficients), and significance of relationships (t-statistics of coefficients) are comparable for most independent variables in the two years. In particular, the *overall sprawl index* for 1990 is significantly related, in the expected directions, to average vehicle ownership, transit share of commute trips, walk share of commute trips, daily VMT per capita, annual traffic fatality rate, and maximum ozone level.

The *overall sprawl index* for 1990 is not significantly related to either average commute time or annual delay per capita. This too agrees with results for 2000.

Table 10. Transportation Outcomes vs. Overall Sprawl Index (1990)

	vehph	***transhr***	***walkshr***	***mntime***	***dlycap***	***vmtcap***	***facap***	***oz8h***
constant	-0.125	-22.09	-4.61	-3.51	-100.5	7.86	82.20	0.107
index	-0.00214 (-4.5)***	0.0635 (5.4)***	0.0246 (4.8)***	-0.00693 (-0.8)	0.00240 (0.1)	-0.0780 (-3.8)***	-0.196 (-4.0)***	-0.00033 (-5.2)***
metpop	-3.8E-08 (-4.3)***	1.07E-06 (4.8)***	-3.6E-08 (-0.4)	9.45E-07 (5.7)***	1.55E-06 (1.8)	-2.8E-07 (-0.5)	-1.4E-07 (-0.2)	3.28E-09 (2.7)**
hhsize	0.334 (4.6)***	0.294 (0.2)	0.363 (0.5)	4.04 (3.0)**	15.58 (3.2)**	-4.14 (-1.4)	4.86 (0.7)	0.0128 (1.3)
wkage	0.0195 (3.4)**	0.162 (1.1)	0.0730 (1.2)	0.171 (1.6)	0.730 (2.0)	0.649 (2.8)**	-0.233 (-0.4)	-0.00036 (-0.5)
pcinc	2.71E-06 (0.5)	0.00051 (3.8)***	2.2E-05 (0.4)	0.00027 (2.7)**	0.00157 (3.2)**	-0.0004 (-1.8)	-0.00168 (-3.1)**	-4.5E-07 (-0.6)
adjusted R^2	0.42	0.55	0.21	0.51	0.47	0.22	0.27	0.29

* .05 probability level
** .01 probability level
*** .001 probability level

Next we report on relationships between outcomes and individual sprawl factors. Here we discover some interesting differences. The *density factor* appears to be a less important factor in certain outcomes in 1990 than in 2000. In 1990, it does not prove significantly related to walk share of commute trips, VMT per capita, or the fatal accident rate, whereas it is significant for all three in 2000. It remains the most significant environmental variable in equations for average vehicle ownership and transit share of commute trips.

As if filling a void, the *centers factor* proves more significant in 1990 than in 2000, and overall, surpasses density as environmental variable most closely associated with travel and transportation outcomes. It is the most significant environmental correlate with walk share of commute trips, annual delay per capita, and VMT per capita, and emerges as a significant correlate with average commute time. In all cases the degree of centering has the expected, favorable relationship to outcomes.

The *mix factor* proves significantly related to average commute time in 1990, as in 2000, and emerges as significantly related to the fatal accident rate. Indeed, it is the most important environmental correlate with the latter in 1990. Its relationship to both outcomes is inverse, as expected.

Finally, the *streets factor* remains the one variable with unexpected relationships to transportation outcomes, in 1990 as in 2000. It is directly related to average commute time and annual delay per capita, both at significant levels. A possible explanation for these relationships was offered previously. The fact that the relationships are so similar for 1990 and 2000 suggests that this is not a statistical fluke, but a phenomenon that requires further study.

Table 11. Outcomes vs. Sprawl Factors (1990)

	transhr	*walkshr*	*mntime*	*dlycap*	*vmtcap*	*facap*	*oz8h*
constant	-12.78	-2.73	-4.01	-144.6	5.26	86.58	0.067
denfac	0.136 (4.1)***	0.0153 (1.0)	0.0182 (0.7)	0.135 (1.5)	-0.0446 (-0.6)	0.0184 (0.1)	-0.00071 (-3.8)***
mixfac	-0.0117 (-1.0)	0.00515 (1.0)	-0.0190 (-2.2)*	-0.0116 (-0.4)	-0.00221 (-0.1)	-0.174 (-3.3)**	6.92E-05 (1.0)
cenfac	0.0400 (3.2)**	0.0243 (4.3)***	-0.0220 (-2.4)*	-0.165 (-4.6)***	-0.0607 (-2.3)*	-0.136 (-2.5)*	-5.7E-05 (-0.8)
strfac	-0.00441 (-0.3)	0.000522 (0.1)	0.0252 (2.3)*	0.0990 (3.2)**	-0.0461 (-1.7)	-0.00184 (0.0)	-7.5E-05 (-0.9)
metpop	5.93E-07 (2.2)*	1.7E-08 (0.1)	6.43E-07 (3.2)**	-6E-07 (-0.8)	-3.4E-07 (-0.5)	-9.3E-07 (-0.8)	6.91E-09 (4.4)***
hhsize	-2.17 (-1.2)	0.0156 (0.0)	3.78 (2.9)**	15.29 (3.7)***	-2.90 (-0.9)	8.74 (1.1)	0.0278 (2.7)**
pwkage	-0.00847 (-0.1)	0.0142 (0.2)	0.210 (2.0)*	1.66 (5.3)***	0.765 (2.9)**	-0.340 (-0.5)	-2.8E-05 (0.0)
pcinc	0.00041 (2.8)**	4.51E-05 (0.7)	0.00018 (1.6)	0.00080 (2.0)	-0.00041 (-1.5)	-0.00149 (-2.3)*	8.16E-07 (1.0)
adjusted R^2	0.64	0.27	0.61	0.73	0.23	0.30	0.37

* .05 probability level
** .01 probability level
*** .001 probability level

6 CONCLUSION

The relationships found between urban sprawl and the quality of life outcomes show that traffic and transportation-related problems appear to increase in more sprawling areas. Even when controlling for income, household size, and other variables, people drive more, have to own more cars, breathe more polluted air, face greater risk of traffic fatalities, and walk and use transit less in places with more sprawling development patterns. While these findings may seem obvious, this is the first study to explicitly measure sprawl and explicitly relate sprawl, so measured, to an important set of transportation outcomes. This study suggests that if Houston, for example, were only somewhat more compact, thousands more people would walk to work, residents would drive less, and children would breathe cleaner air.

Generalizing to other transportation-related outcomes, these findings suggest that even after controlling for numerous demographic factors, urban sprawl has a major influence on energy (gasoline) consumption and other outcomes that are tied to vehicle-miles traveled. Future reports will further quantify the costs in health, safety, time and money associated with this phenomenon.

Policy Recommendations

Even for metropolitan regions that appear relatively compact, urban sprawl is a serious problem because of its strong association with numerous societal problems. For the nation's most sprawling regions, it is even more urgent to devise strategies that can reduce sprawl. Advocates and practitioners associated with the Smart Growth movement have devised a wide array of techniques and policies to manage growth and help regions avoid haphazard sprawl. (For a list of resources containing a broad range of smart growth policy recommendations, see Appendix X.) The following recommendations, however, are focused on the specific issues examined in this report, namely the four factors and the transportation outcomes measures.

This study found strong evidence that at the regional scale, increased residential density has the potential to diminish the need to own and drive automobiles, which in turn can help protect air quality and reduce traffic fatalities, while increasing the share of commuters who use transit or walk. That is not a prescription for high rises in every neighborhood – far from it. The research indicates that even modest increases in average density, from one or two houses per acre (roughly the size of a football field) to as few as six or seven, can offset the negatives examined in this report.

There are many strategies that can result in attractive communities with higher densities. Some of these strategies tend to fall under the general heading of community economic development. At the same time, the development of compact, walkable neighborhoods is gaining momentum in the real estate market, with growing numbers of retiring baby boomers expressing a preference for in-town living, greater conveniences and a stronger sense of community.

Policy Recommendation #1: Reinvest in Neglected Communities and Housing Opportunities

For decades, thousands of community-based organizations have sought to use policy and financing tools to improve the quality of life in distressed communities. These tools include state and local low-income housing tax credit, the Community Reinvestment Act, Community Development Block Grants, state affordable housing trust funds, and a whole range of state and local programs. Such strategies infuse badly needed resources into long neglected neighborhoods and may reverse the abandonment of such neighborhoods. To reduce the impacts of sprawl, these reinvestment and housing programs should at least be maintained at current funding levels and preferably increased. In particular, a federal proposal to create a national affordable housing trust fund should be enacted into law.

Policy Recommendation #2: Rehabilitate Abandoned Properties

A related strategy is the rehabilitation of individual abandoned properties, be they old vacant buildings, tax-delinquent homes, empty historic buildings, or other potentially useful properties. New Jersey, for example, passed a new rehabilitation code to facilitate the restoration of older buildings. Such measures have led to a large increase in rehab investment in New Jersey cities, and have been adopted by Maryland, Rhode Island and other states. Other states have reformed tax foreclosure laws and initiated improved inventory and tracking systems to more quickly identify negligent owners of abandoned properties and transfer them to new investors.

Policy Recommendation #3: Encourage New Development or Redevelopment in Already Built Up Areas

Smart growth is not about stopping growth or even slowing growth; rather it is about focusing growth in places where it can properly be accommodated. Chief among those would be areas that already are within the urban footprint. Most metro regions contain ample redevelopment opportunities, which may include old industrial sites (brownfields), empty shopping malls (greyfields), and vacant lots. Such properties tend to have existing infrastructure (roads, water, sewer and other utilities), are large enough to accommodate entire new neighborhoods with a mix of homes, shops, offices, civic buildings and parks, linked together by a grid of streets and sidewalks.

Policy Recommendation #4: Create and Nurture Thriving, Mixed-Use Centers of Activity
This study found that strong urban and suburban downtowns and other centers of activity are associated with fewer traffic fatalities, lower vehicle mileage, and more transit use and walking to work. As such, the fostering of such centers is an essential smart growth strategy. One of the most promising approaches to accomplishing this is to concentrate mixed-income housing, shops and offices around train stations and bus stops, which is commonly referred to as transit-oriented development (TOD).
Another important strategy involves rezoning to permit multifamily housing in and around the jobs-rich "edge cities". This can make it possible for more people to live near work while also introducing the residents needed to support neighborhood retail.

Policy Recommendation #5: Support Growth Management Strategies
The low scores for the overall Sprawl Index (indicating more sprawl) were associated with more driving, vehicle ownership, traffic fatalities, peak ozone levels, and lower levels of transit use and walking to work. Key strategies for curbing sprawl include planning and zoning tools that help regions better manage growth. Portland, Oregon has developed one oft-cited model, wherein a regional growth framework is established and managed by an elected regional council in concert with local governments. Another method is the strategic preservation of prime farmland, sensitive environmental lands, forests and other green spaces, in conjunction with careful planning for development in designated areas.

Policy Recommendation #6: Craft Transportation Policies that Complement Smarter Growth
In the coming year, Congress will consider the reauthorization of the nation's transportation law, the Transportation Equity Act for the 21st Century (TEA-21). This reauthorization is not only the means by which states receive federal gas tax dollars for much needed transportation projects, but it is also the main federal opportunity to improve the interaction between local and regional development plans and transportation planning and programming. In keeping with the previous five recommendations, this reauthorization should:

- Support "fix-it-first" state and federal transportation infrastructure policies, which favor the maintenance of existing streets and highways over the construction of new ones,
- Prioritize and increase funding that serves community development goals in lower-income neighborhoods,
- Create incentives for transit-oriented development, particularly mixed-use development and mixed-income housing, and

- Maintain important funding programs for historic preservation, walking and cycling facilities, and Main Street and streetscape improvement projects.

In addition, the new law should include resources that enable communities to better coordinate transportation and land use, including:

- Funds to support more sophisticated scenario planning for both corridors and regions,
- Better predictive models that cover not only transportation outcomes but also community impacts, and
- Tools for improved community involvement in the planning process.

APPENDIX -- DATA SOURCES

U.S. Census of Population and Housing

Many of our density and centeredness measures were derived from census tract data in the STF3 files for 1980 and 1990, extracted from the Geolytics *Census CD 1980* and *Census CD+Maps* CDROMS for 1980 and 1990, and the Census 2000 Redistricting Data, downloaded from the Census ftp website (the state geographic header files). State and county FIPS codes and the 100 percent population counts were extracted for each year. In addition, for 1990 and 2000, the land area of the tract and the latitude and longitude for the internal point in the tract were extracted.

First, census tracts were selected for the 101 largest MSAs/CMSAs/NECMAs as defined in 1990. The selection was accomplished by creating a correspondence file between MSAs/PMSAs/CMSAs and counties, extracted from the 1990 STF3 files, and modifying this file to include the counties in the NECMAs as reported in the *1991 State and Metropolitan Area Databook.* In New England, MSAs and PMSAs are defined using towns and not counties as the building blocks. Census tracts can be split by MCD (town) boundaries and hence by the MSA and PMSA boundaries. The parts of split census tracts were not available and were not used here. Rather, whole census tracts for the three years were selected for each MSA and PMSA based upon whether the centroids or internal points of the tracts fell within the MSA and PMSA boundaries.

Next, tracts with areas less than 0.001 square miles were deleted. This was done because tracts of such small area (including areas of zero square miles) were felt to represent special situations, such as crews of vessels. Also, it was felt that the calculation of densities for tracts with such small areas was likely to be misleading or meaningless.

Finally, the population densities of the tracts were calculated and tracts with population densities of less than 100 persons per square mile were excluded. This was done because such tracts were seen as representing essentially nonurbanized portions of metropolitan areas that should not be included in the calculation of urban sprawl measures.

The measures associated with the density gradient were calculated in the following manner. For each PMSA or MSA, the location of the central business district was identified for the first-named city in the area. No density gradient measures were calculated for the eight PMSAs identified only by the name or names of the counties making up the PMSA, with no principal city being identified. These PMSAs are:

- Beaver County, PA PMSA
- Bergen--Passaic, NJ PMSA
- Brazoria, TX PMSA
- Lake County, IL PMSA
- Middlesex--Somerset--Hunterdon, NJ PMSA
- Monmouth--Ocean, NJ PMSA
- Nassau--Suffolk, NY PMSA
- Orange County, NY PMSA

Census Transportation Planning Package

The Census Transportation Planning Package (CTPP) is a set of special tabulations from the decennial census designed for transportation planners. CTPP is a cooperative effort sponsored by the State Departments of Transportation under a pooled funding arrangement with the American Association of State Highway and Transportation Officials (AASHTO). Initiated with the 1970 Census, CTPP is the only census product that summarizes data by place of work as well as place of residence and provides information on travel flows between home and work.

To measure land use balance and mix, data were required for subdivisions of metropolitan areas. CTPP contains splits of metropolitan areas by traffic analysis zone (TAZ), block group, or census tract. Metropolitan planning organizations got to choose the units of subdivision. Of the 101 metropolitan areas in our original sample, 92 were divided by TAZ, six by block group, and three by census tract.

Population and employment by sector (18 sectors) were extracted from CTPP. The 18 employment sectors were grouped into five categories:

- retail
- finance, insurance, and real estate
- business services
- personal services: personal services, entertainment, health, education, and other professional, and related services
- all other sectors: agriculture, mining, construction, manufacturing non-durables, manufacturing durables, transportation, communication, wholesale, public administration, and armed forces

In CTPP, population data are stored separately from the employment data, requiring the merging of the files by subdivision identification code. This process resulted in many subdivisions that contained housing but no jobs or vice versa. CTPP staff assured us that these apparent mismatches were valid.

MPO planning areas only occasionally coincide with MSA and PMSA boundaries. In some cases, planning areas take in more territory (entire CMSAs or additional urbanized area outside MSAs). In other cases, they take in less territory (only urbanized portions of metros). When they take in less territory, the entire set of TAZs was used to calculate mixed-use measures. When they take in more territory, the set of TAZs had to be pared back to metro area boundaries before calculations could be performed.

The extraction of only relevant data (just TAZs within the metro area) involved selecting TAZs within metropolitan counties on CTPP relational maps, exporting population and employment data for these TAZs to Excel files, and attempting to match metro area populations for 1990. Because county boundaries are not always obvious on CTPP maps, the process of delineating metros and extracting relevant data involved trial and error. Eventually, perfect or near-perfect matches (within 5 percent population difference) were achieved for 74 or the 83 metros in our final data set.[34]

Natural Resources Inventory

In recent years, the U.S. Department of Agriculture (DOA) has made available for research purposes its National Resources Inventory (NRI). The NRI, a spatial survey conducted every five years, includes data on major categories of land use for non-federally owned lands, some 75 percent of the country's land base.

NRI captures data from statistically sampled locations on land cover, land use, soils, water bodies, and other natural features. Samples are located in all counties and parishes of the 50 states. Data for the 1997 NRI were collected for about 800,000 sample points, using photo-interpretation and other remote sensing methods. Data are considered statistically reliable enough for national, statewide, and multi-county use. The 1997 NRI database contains data for

---

[34] Special treatment was required for many metro areas. Data for Greenville and Spartanburg were stored separately and had to be combined. Data for Allentown and Wilmington were divided between state CDs and also had to be combined. Portland and Los Angeles used slightly different identification codes for population and employment data. The Portland MPO provided information that helped us match identification codes for both metropolitan areas. For two metropolitan areas, Providence and Hartford, CTPP did not provide data specifically for the metropolitan area, but rather for the complete states of Rhode Island and Connecticut. The identification codes from CTPP contained enough information to allow us to strip out data for relevant counties in Providence. Documents from the Connecticut MPO helped us identify TAZs within the Hartford metropolitan area. For two metropolitan areas, New Bedford and Worcester, outside TAZs had to be added to the base set to achieve population matches. For one metropolitan area, Scranton, CTPP contained insufficient data to allow the calculation of balance and mix measures. Of the four counties within the metro area, CTPP contains data for only one county constituting 30 percent of metropolitan area population. Scranton was therefore dropped from the final sample.

four points in time (1982, 1987, 1992, and 1997) that are comparable and consistent, and that can be used in year-to-year comparisons.

In NRI, lands are assigned to the urban and built-up category if they have the following land covers/land uses: residential, industrial, commercial, and institutional land; construction sites; public administrative sites; railroad yards; cemeteries; airports; golf courses; sanitary landfills; sewage treatment plants; water control structures and spillways; small parks (less than 10 acres) within urban and built-up areas; and highways, railroads, and other transportation facilities if they are surrounded by urban areas. Also included are tracts of less than 10 acres that do not meet the above definition but are completely surrounded by urban and built-up land.

Washington, D.C. and many independent cities (most in Virginia) are incorporated into NRI county totals. Data from other sources had to be aggregated for these counties and county equivalents to produce a consistent dataset.

American Housing Survey

The American Housing Survey (AHS) is conducted by the U.S. Census Bureau for the Department of Housing and Urban Development. It began in 1973 as the Annual Housing Survey, and is now collected every other year for a national sample and about every four years for each of 46 selected metropolitan areas. The AHS provides data on housing, household characteristics, equipment, fuels, recent movers, and neighborhood quality. The last of these is of particular interest to us.

Our datasets consist of the national microdata samples for the years 1989, 1991, and 1993 (pooled to represent the beginning of the decade) and 1997 and 1999 (pooled to represent the end of the decade). The national survey consists of a fixed sample of about 50,000 homes, plus 5,000 newly constructed units each year. Surveying the same units year after year provides insights into how homes and households change over time. Our pooling of data for multiple years should tend to reduce the error associated with individual responses.

Data are mostly for MSAs and PMSAs, as defined at the time of the survey. This creates inconsistencies with our other datasets. Some inconsistencies result from changes in MSA boundaries during the 1990s, others from the use of MSA or PMSA boundaries rather than NECMA boundaries to represent New England metros. In two cases, Buffalo-Niagara Falls and Portland-Vancouver, data are provided for CMSAs rather than PMSAs and cannot be disaggregated.[35]

The computer files contain raw data from each interview. In adding up cases, one needs to "weight" each case for the probability of selection. We did this.

---

[35] The error associated with the use of CMSAs should not be large, as about 80 percent of the CMSA is made up of the larger PMSA in each case.

Zip Code Business Patterns

Glaeser et al.'s measures of job sprawl were incorporated into our sprawl factors. Their source of data was the U.S. Department of Commerce's Zip Code Business Patterns for 1996. Zip Code Business Patterns data are extracted from the Standard Statistical Establishments List, a file of all single and multi-establishment companies created by the U.S. Census Bureau. Data are provided on the total number of establishments, employment, and payroll for more than 40,000 zip codes nationwide. The number of establishments is broken down into nine employment-size categories by detailed industry for each zip code.

Glaeser et al. considered only zip codes that lie inside the largest metropolitan areas. They measured the proportion of employment within rings around the CBD of each metro. Locations of CBDs were taken from the 1982 Economic Censuses Geographic Reference Manual, which identifies CBDs by tract number. GIS (geographic information system) software was used to calculate the distance from the centroid of each zip code to the centroid of the corresponding CBD. Glaeser et al. argued that while not ideal, zip code referenced data offer the best micro-geographic evidence on employment location in the U.S. to date.

TIGER/Line Files

The TIGER (Topologically Integrated Geographic Encoding and Referencing) files are a digital database of geographic features, such as streets, railroads, rivers, lakes, and political boundaries, covering the entire United States. The database contains such information about streets as their name, latitude and longitude, address ranges, and length. TIGER was developed at the Census Bureau to support the mapping and related geographic activities required by the decennial census. In order for the public to use the information in the TIGER database, the U.S. Census Bureau releases periodic extracts known as the TIGER/Line files. The most recent version is the Redistricting Census 2000 TIGER/Line.

To make use of these data, a user must have Geographic Information System (GIS) software that can import TIGER/Line data. With the appropriate software, a user can produce digital street maps and generate measures of street density and block length. In this study, shape files were opened in Arcview, and the statistics function was used to obtain miles of roadway and number of roadway segments. For 2000, segment lengths were already in miles. For 1990, lengths given in decimal degrees had to be converted to miles.[36]

---

[36] Once street data were entered into a spreadsheet, a logical inconsistency was discovered: aggregate 2000 street lengths were less for most counties than aggregate 1990 street lengths. An investigation to determine the source of the error quickly focused upon the 1990 street-length data as originally reported in the *First Street* CD-ROM database. Checking individual streets, east-west streets were reported to be longer in 1990 than 2000, and the disparity was greater at northern than southern latitudes. Expert opinions were sought from *ESRI's* technical staff. None could give a satisfactory explanation for the problem, but suspicion has fallen upon the datum used by *First Street* when compiling the themes. *ESRI* staff suggested we run the *Calcapl* script, which recalculates lengths given any datum or projection one chooses. This was done, using the *Albers Equal Area Conic (conterminous United States)* projection,

After processing street data for entire counties, we became concerned that the presence of large rural tracts in certain metropolitan areas gave a false picture of the street networks actually utilized by residents. County street data were therefore geo-processed by electronically intersecting (or overlaying) the urbanized area layer from the U.S. Census Bureau with the TIGER street shape files (so as to "cut off" streets at the urbanized area boundary). This generated an intersect theme for each county, which was then processed in the same manner as the countywide themes to obtain network densities and average block lengths for urbanized areas only. This was accomplished for both 1990 and 2000 urbanized areas and street networks.

7 ACKNOWLEDGEMENTS

Measuring Sprawl and Its Impact is the product of three years of research. The authors wish to acknowledge numerous individuals who generated or processed data upon which our sprawl and outcome measures are based. Key contributors were John Ottensmann of Indiana University, who processed Census data and helped design the sprawl measures, and David Miller of the Claritas Corporation, who provided proprietary data on centers and subcenters. Alex Zakrewsky of Global Geometrics worked with the TIGER files, and William Dolphin of Rutgers University also processed census data. The following Rutgers students worked with such datasets as the American Housing Survey and Census Transportation Planning Package: Robert Diogo, Danny Knee, Rachael Kennedy, Kurt Paulsen, Jee Shin, and Yue Wu.

The authors also gratefully acknowledge the study's four academic peer reviewers: Robert Cervero of U.C. Berkeley, Randall Crane of U.C.L.A., Susan Handy of U.C. Davis, and Jonathan Levine of the University of Michigan. Others who helped us review this report include Deron Lovaas, Elizabeth Humphrey, Lee Epstein, Scott Bernstein, Kaid Benfield, and Niki Mitchell.

Barbara McCann and David Goldberg helped write the summary version of the report, and developed and edited all the supporting materials. John Bailey worked with SGA's partners to arrange the national and local releases. Linda Bailey and Michelle Ernst helped analyze the findings and prepared charts and tables for the final reports. John Goldener did the design and layout of the reports, and Marquis Clayton prepared the online version. The authors thank them for their important contributions.

This research was made possible by generous support from the Surdna Foundation, the William and Flora Hewlett Foundation, the George M. Gund Foundation, the David and Lucile Packard Foundation, the H. M. Jackson Foundation, the Turner Foundation, and the U.S. Environmental Protection Agency.

---

with map units reporting in miles. This procedure rectified for all practical purposes the observed length disparities between 1990 and 2000. New street-length and segment statistics were then calculated from the theme table.