

Measuring Sprawl and Its Transportation Impacts

Reid Ewing, Rolf Pendall, and Don Chen

Across the United States, urban sprawl, its impacts, and appropriate containment policies have become the most hotly debated issues in urban planning. Today's debates have no anchoring definition of sprawl, which has contributed to their unfocused, dogmatic quality. Efforts to measure sprawl and test for relationships between sprawl and transportation outcomes are described. This is the first use of the newly minted Rutgers–Cornell sprawl indicators. Sprawl is operationalized by combining many variables into a few factors representing density, land use mix, degree of centering, and street accessibility. This consolidation of variables is accomplished with principal component analysis. These factors are then related to vehicle ownership, commute mode choice, commute time, vehicle miles traveled per capita, traffic delay per capita, traffic fatalities per capita, and 8-h ozone level. These associations are made with multiple regression analysis. For most travel and transportation outcomes, sprawling regions perform less well than compact ones. The exceptions are average commute time and annual traffic delay per capita, which do not clearly favor compactness over sprawl. The main limitation of this study has to do with the data it uses. By necessity, the study uses highly aggregate data from a variety of sources that are not always consistent as to the area under study and time period. They are simply the best data available from national sources with sufficient breadth to provide a panoramic view of sprawl in the United States. Results will have to be validated through follow-up work of a more focused nature.

Across the United States, urban sprawl, its impacts, and appropriate containment policies have become the most hotly debated issues in urban planning. Today's debates have no anchoring definition of sprawl, which has contributed to their unfocused, dogmatic quality. There is little consensus about how urban sprawl affects everything from housing affordability to open space preservation to racial segregation (1). The impacts of sprawl on travel and transportation outcomes are the focus of this paper.

There have been many previous attempts to measure sprawl (2–9). The most notable feature of past studies (with few exceptions) is the failure to define sprawl operationally in all its complexity. Previous attempts to operationalize the construct “sprawl” with a variable or two include, for example, *USA Today's* proportion of metropolitan population living outside the urbanized area (2), and Kahn's proportion of metropolitan employment more than 10 mi from the central business district (5). A variable or two cannot adequately represent the inherent complexity of sprawl.

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

A third notable feature of the literature is how little attention is paid to the impacts of sprawl. With the exception of a few studies focusing on individual impacts, the literature is nearly devoid of impact assessment. Sprawl is presumed to have negative consequences—not shown to have them.

In this study, sprawl is operationalized by combining many variables into a few factors representing residential density, land use mix, degree of centering, and street accessibility. This consolidation of variables is accomplished with principal component analysis. These factors are then related to vehicle ownership, commute mode choice, commute time, vehicle miles traveled (VMT) per capita, traffic delay per capita, traffic fatalities per capita, and 8-h ozone level. These associations are made with multiple regression analysis.

DATA AND MEASURES

Sample

The sample of U.S. metropolitan areas originally consisted of the largest 101 metropolitan statistical areas (MSAs), consolidated metropolitan statistical areas (CMSAs), and New England county metropolitan areas (NECMAs). These were the largest as of 1990, the year for which the most complete set of land use, road network, and sprawl impact variables are available.

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

Within the sample of 101 MSAs/CMSAs/NECMAs are 139 MSAs/PMSAs/NECMAs. Ultimately, the sample was further limited by size considerations and the availability of data sets. Smaller metropolitan areas appear to be fundamentally different from large ones when it comes to land use patterns. They are more likely to be monocentric, for example, while large metropolitan areas are likely to be polycentric.

R. Ewing, Alan M. Voorhees Transportation Center, Rutgers University, 33 Livingston Avenue, Suite 400, New Brunswick, NJ 08901. R. Pendall, Department of City and Regional Planning, Cornell University, 212 West Sibley Hall, Ithaca, NY 14853. D. Chen, Smart Growth America, 1100 17th Street, NW, 10th Floor, Washington, DC 20036.

A decision was made to opt for fewer metropolitan areas with complete data sets instead of more metropolitan areas with partial data sets. The availability of data drops off as metropolitan population declines, and sample sizes shrink for those land use and outcome measures based on samples. The final sample consists of 83 metropolitan areas with populations over 500,000 as of 2000. In that year, these metropolitan areas collectively were home to more than 150 million Americans, more than half of the entire U.S. population.

Metropolitan areas were defined by their 1990 boundaries for both 2000 and 1990. A consistent set of metropolitan boundaries was required to compare the degree of sprawl over time, the change in outcome measures 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 metropolitan area boundaries than to add then rural counties to 1990 metropolitan area boundaries. As a practical matter, the use of 1990 boundaries should have minor effects on sprawl statistics because the recently added rural counties will have populations too small to appreciably affect metropolitan averages.

Sprawl Measures

Consistent with the technical literature, this study characterizes sprawl in multiple dimensions. The operational variables that together make up each dimension of sprawl are defined in the following subsections, along with the data sources from which they came.

Residential Density

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 metropolitan population living 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 (fewer 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 metropolitan area boundaries.

1. *dens* = gross population density in persons per square mile.
2. *l1500p* = percentage of population living at densities less than 1,500 persons per square mile, a low suburban density.
3. *g125cp* = percentage of population living at densities greater than 12,500 persons per square mile, an urban density that begins to be transit supportive.
4. *dgcent* = estimated density at the center of the metropolitan area derived from a negative exponential density function.

One density variable was derived from the U.S. Department of Agriculture’s National 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 permits independent measures of urban area density (independent of the census, which uses population density and jurisdictional boundaries to establish thresholds between urban and nonurban areas). The 1997 densities from NRI were taken as representative of the end of the decade, while the 1987 densities

were used to represent the beginning. This gave a 10-year interval for calculating changes in densities, equivalent to the decennial census.

5. *urbdn* = gross population density of urban lands.

One density variable was derived from the national microdata sample of the American Housing Survey (AHS). The national survey is conducted every 2 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 weighted average value was used to adjust for different probabilities of sample selection in the original data set.

6. *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. Their spheres of influence may cross metropolitan area boundaries. For example, Jersey City, New Jersey, 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 City and Newark, New Jersey. A population center density variable was computed for each metropolitan area as the weighted average density for all population centers within the given area. The average densities were weighted by the resident populations in the sphere of influence of each population center.

7. *dncen* = weighted density of all population centers within a metropolitan area.

Principal components were extracted from this set of density-related variables, and the principal component that accounted for the greatest variance became the density factor. Factor loadings (that is, correlations of these variables with the density factor) are presented in Table 1. The 0 at the end of the variable name refers to 2000 (97 refers to 1997, etc.). Again, the factor loadings are for 2000, as this is the base year.

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: the percentage of population living at fewer 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 1 Variable Loadings on the Density Factor

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

Land Use Mix

Two 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, and those representing the diversity of land uses within subareas of a region (*IO*). Both types were estimated for metropolitan areas in the sample and became part of the mix factor.

The first three mixed-use variables were derived from the national microdata samples of the AHS. Samples were pooled and weighted as described previously.

1. *ecom* = percentage of residents with businesses or institutions within one-half block of their homes.
2. *shop* = percentage of residents with satisfactory neighborhood shopping within 1 mi.
3. *sch* = percentage of residents with a public elementary school within 1 mi.

Three additional mixed-use variables were derived from the Census Transportation Planning Package (CTPP) for 1990. CTPP is the only census product that summarizes data by place of work as well as by place of residence; it alone permits one to measure the degree of balance between employment and population (jobs and residents) for subareas of metropolitan areas as well as the degree of employment mixing for subareas. For most metropolitan areas 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 the 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.

Two balance variables were derived from the CTPP. One measures the degree of balance within TAZs between jobs and residents, where balance equals 1 for TAZs with the same jobs-to-residents ratio as the metropolitan area as a whole, 0 for TAZs with jobs or residents but not both, and intermediate values for intermediate cases. The expression used to calculate job–resident balance was as follows:

$$\sum_{i=0}^{i=n} \{1 - [ABS(J_i - JP \times P)] / (J_i + JP \times P)\} \times [(J_i + JP \times P) / (TJ + JP \times TP)]$$

where

- i* = TAZ number (usually a TAZ),
- n* = number of TAZs in the metropolitan 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, and
- TP* = total residents in the metropolitan area.

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.

A job mix variable was also derived. 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. The expression for this measure is as follows:

$$\sum_{i=1}^n \sum_j - \{ [P_j \times LN(P_j)] / LN(j) \} \times (RJ_i / TRJ)$$

where

- i* = TAZ number,
- n* = number of TAZs in the metropolitan 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, and
- TRJ* = total number of retail and total personal services jobs in the metropolitan area.

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.

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 values, absolute values were divided by the natural logarithm of jobs plus residents per TAZ in each metropolitan area. The resulting mixed-use variables were as follows:

1. *nbal* = job–resident balance.
2. *nrbal* = population-serving job–resident balance.
3. *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 the mix factor. Loadings of these variables on the mix factor are presented in Table 2. 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, the 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.

Degree of Centering

Metropolitan centers are concentrations of activity that provide agglomeration economies, support alternative modes and multi-

TABLE 2 Variable Loadings on the Mix Factor

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

purpose trip making, create a sense of place in the urban landscape, and otherwise differentiate compact metropolitan areas from sprawling ones. 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 around the mean, the more centering and subcentering exists within the metropolitan 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.

1. *coefvr* = coefficient of variation of population density across census tracts (standard deviation divided by mean density).
2. *dggrad* = density gradient (rate of decline of density with distance from the center of the metropolitan area).

The degree of centralization of employment within the metropolitan area was represented by two variables borrowed from the work of Glaeser et al. (7) (see preceding literature review). For the 100 largest U.S. metropolitan areas, they calculated the share of overall metropolitan area employment within a 3-mi ring of the central business district (CBD), the share of metropolitan area employment within a 10-mi ring of this spot, and the share beyond the 10-mi ring. All were measured for 1996. Two of the three variables became part of the centers factor:

3. *l3emp* = percentage of metropolitan employment less than 3 mi from the CBD.
4. *g10emp* = percentage of metropolitan employment more than 10 mi from the CBD.

The last two variables contributing to the centers factor were derived from Claritas databases. 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 metropolitan area, a population center in a different metropolitan area, a population center outside all metropolitan areas, or no population center at all. Most of the population of Akron, Ohio, 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, Ohio, part of the same CMSA; a portion on the east side relates to centers in Youngstown, Ohio; and a little bit does not relate to centers at all.

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

5. *popcen* = percentage of metropolitan population relating to centers or subcenters within the same MSA or PMSA.
6. *rdnap1c* = ratio of weighted density of population centers within the same MSA or PMSA to the highest density center to which a metropolitan area relates.

Factor loadings are presented in Table 3. 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 mi from the center

TABLE 3 Variable Loadings on the Centers Factor

<i>coefvr0</i>	0.21
<i>dggrad0</i>	-0.74
<i>l3emp96</i>	0.76
<i>g10emp96</i>	-0.76
<i>popcen0</i>	0.17
<i>rdnap1c0</i>	0.48
eigenvalue	2.01
% of variance	33.49

(*g10emp*), both of which assume higher values in decentralized metropolitan 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 of density.

Street Accessibility

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, approximate block length can be determined. And from the U.S. Census Summary files, block sizes are 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, street centerline miles and street segments were tallied for entire counties, and approximate block lengths were computed from these. The resulting network measure was inflated by large portions of many metropolitan counties that are undeveloped. Therefore, approximate block lengths were recalculated by using 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 this network measure is not entirely equivalent for the 2 years. When boundary files become available for 1990 urbanized areas using 2000 criteria, consistent area definitions will be applied.

Block sizes were tabulated and two measures derived for each metropolitan area. One was the average block size, and the other was the proportion of blocks 1/100th of a square mile or less in size (which is a traditional urban block, a little more than 500 ft on a side). It became obvious from a review of the data that huge rural tracts could distort averages and should be excluded from the calculation. A ceiling of 1 mi², the size of a standard section and large superblock, was established for this purpose. This resolved the same issue as before, 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:

1. *bklnu* = approximate average block length in the urbanized portion of the metropolitan area.

2. *bksz* = average block size in square miles (excluding blocks > 1 mi²).
3. *smbk* = percentage of small blocks (<0.01 mi² or smaller).

Factor loadings are presented in Table 4. 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. The one factor captures 76% of the variance in the original data set.

Rescaling Factors

Factor scores derived with principal component analysis had mean values of 0 and standard deviations of 1 for sampled metropolitan areas in 2000. Individual factor scores were converted to a scale with a mean value of 100 and a standard deviation of 25. The linear transformation performed in this step did not affect rankings of metropolitan areas or 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 provided the ability to test for nonlinear relationships between sprawl and outcomes.

The four rescaled sprawl factors are denoted as follows:

1. *denfac0* = density factor for 2000 (a weighted combination of seven density variables).
2. *mixfac0* = mix factor for 2000 (a weighted combination of six mixed-use variables).
3. *cenfac0* = centers factor for 2000 (a weighted combination of six center-related variables).
4. *strfac0* = streets factor for 2000 (a weighted combination of three 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. A few metropolitan areas are compact in all dimensions. Boston, New York, and San Francisco fall into this group. A few sprawl badly in all dimensions. These include Atlanta, Greensboro–Winston Salem–High Point, North Carolina, and Riverside–San Bernardino, California. But most metropolitan areas are mixed, which again emphasizes the independent nature of the different sprawl factors.

1990 Values and 10-Year Changes

To compare the degree of sprawl in 1990 and 2000 and to compare relationships between sprawl and outcomes for the 2 years, a consistent measure of sprawl was needed. The year 2000 became the base year. Complete data sets were also available for 1990, so both years were considered critical to an understanding of sprawl and its effects. The easiest way to achieve consistency in the measurement

of sprawl was to apply factor weights (coefficients reflecting the contribution of each operational variable to the overall factor score) for 2000 to 1990 data. Thus, for certain factors, it became possible to compare the extent of sprawl (or its components) in the 2 years and to judge whether sprawl increased or decreased over the decade. Sprawl ratings for both years are presented by Ewing et al. (11).

For one factor, it was not possible to measure changes over the decade. Of the six variables that make up the mix factor, three came from the 1990 CTPP and cannot be updated to 2000 until the new CTPP is released. The other three came from the AHS and are based on samples so small as to produce sizable sampling errors. Thus, there are two factors left for which changes can be reliably measured—the density factor and the centers factor—and one factor for which changes can be measured with a degree of consistency—the streets factor.

The sample of metropolitan areas divides into three groups, and a small number are becoming more sprawling with respect to all three factors; a larger number are becoming less sprawling in three dimensions; and most are becoming more sprawling in some respect but less in another. The “mores” include Akron, Ohio; Ft. Worth, Texas; and Tampa, Florida—a hard group to generalize across. The “lesses” are mostly fast growing metropolitan areas concentrated in the West and Florida, including Anaheim, California; Phoenix, Arizona; and Orlando, Florida. The mixed group spans the continent. Many are losing density and becoming less centered but are thrust into the mixed category by a rise in the streets factor accompanying subdivision of land.

The phenomenon of sprawl, when measured in multidimensional terms over a span of years, is far more complex than most of the technical literature, and all the popular literature, makes it out to be.

Travel and Transportation Outcomes

As already noted, the main purpose of this study was to see what outcomes are associated with sprawl, controlling for other influences. No development pattern is inherently good or bad. It is negative outcomes, such as high VMT or severe congestion, that make one development pattern superior to another. The authors were fortunate to have among their outcome measures recently released 2000 journey-to-work data from the U.S. Census and just-released 2000 congestion data from the Texas Transportation Institute (TTI).

Outcomes attributed to sprawl are summarized by Burchell et al. (12). Those related to travel and transportation became the dependent variables in the analyses. Several travel and transportation outcome measures were derived from the U.S. Censuses of 1990 and 2000:

1. *vehph* = average vehicles per household.
2. *pubtx* = percentage of commuters using public transportation (including taxi).
3. *walk* = percentage of commuters walking to work.
4. *mjwjk* = mean journey-to-work time in minutes.

From the TTI mobility database came traffic congestion data for 1990 and 2000. TTI data apply to urbanized areas instead of metropolitan areas. They are available for 60 urbanized areas corresponding to the 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

TABLE 4 Variable Loadings on the Streets Factor

<i>bkln0u</i>	-0.83
<i>bksz0</i>	-0.86
<i>psmbk0</i>	0.92
eigenvalue	2.28
% of variance	76.0

between urbanized area and metropolitan area is close enough to retain these cases for subsequent analysis (the urbanized areas dropped from the sample were New York–northeastern New Jersey; Chicago–northwestern Indiana; San Francisco–Oakland; Los Angeles; and Dallas–Ft. Worth). Mobility measures are computed instead of measured in the field, and so they are no better than the formulae upon which they are based.

5. *dlycap* = annual hours of delay per capita.

From the FHWA Highway Performance Monitoring System (HPMS) came VMT data for 1990 and 2000. Like TTI data, HPMS data apply to urbanized areas instead of metropolitan areas. It was necessary to piece together values for metropolitan areas from the urbanized areas that make them up. VMT and population were estimated for each urbanized area that has land within a given metropolitan area, with estimates based on the proportion of an urbanized area's total land area that falls within metropolitan boundaries. These estimates were summed over all urbanized areas in a given metropolitan area and a weighted average VMT per capita thereby derived. For example, the Dayton–Springfield, Ohio, 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 metropolitan area included all of the Springfield VMT and population and 96% of the Dayton VMT and population.

Pieced together this way, HPMS data were available for 77 metropolitan areas in the final sample of 83 metropolitan areas. The correspondence between urbanized areas and metropolitan areas was close enough to retain 72 of these urbanized areas for subsequent analysis. (For 1990, there were 75 urbanized areas to begin with and 70 after 5 were dropped. For both 1990 and 2000, the 5 dropped were the same for HPMS data as for TTI data.)

6. *vmtcap* = daily VMT per capita.

From the National Highway Traffic Safety Administration's Fatality Analysis Reporting System came highway fatality data. Because data are available for all counties in the United States, fatal accident rates can be computed for metropolitan areas exactly as defined in 1990, the reference year for the metropolitan area definitions. Rates are available for both 2000 and 1990.

7. *facap* = annual highway fatalities per 100,000 persons.

While not strictly transportation related, the final outcome measure relates to the maximum ozone level in the metropolitan area, a criteria pollutant closely linked to motor vehicle use. [The Environmental Protection Agency (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.] Ozone was selected for analysis over carbon monoxide because the former manifests itself regionally instead of 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.

8. *oz8h* = fourth highest daily maximum 8-h average ozone level.

ANALYSIS AND RESULTS

Given the aggregate nature of this analysis, the statistical method of choice used to test for significant relationships is multiple regression analysis. Significant relationships were tested for by running a series of ordinary least-squares regressions for travel and transportation outcomes in 2000 and 1990. In all regressions, an outcome was regressed on the full set of sprawl factors (all four) and a standard set of control variables (four of these too).

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 that 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 one to control for confounding influences, provided that they are measured and included in the regression equation.

Control Variables

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

1. *metpop* = metropolitan area population (MSA or PMSA).
2. *hhsz* = average household size.
3. *pwkage* = percentage of population of working age (20 to 64 years).
4. *pcinc* = per capita income.

Transportation outcomes are arguably influenced in part by conditions beyond metropolitan area boundaries, especially for PMSAs that are parts of CMSAs. The analysis reported here therefore should be considered preliminary; future analyses may attempt to account for various spillover effects from nearby metropolitan areas.

Outliers

Figure 1 presents a plot of one outcome, transit mode share on the journey to work, versus the density factor. In this plot, there are obvious outliers, having much higher transit mode shares and much higher densities than the other metropolitan areas. 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.

At the same time, the position of the outlying data points in two-dimensional space appears as a continuation of a trend line established by the other cases. This is true not only for transit mode share but for other outcomes as well. 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, New York City and Jersey City, were dropped.

Results for 2000

Results for 2000 are presented in Table 5 and discussed in the following subsection. Regression coefficients and *t*-statistics appear

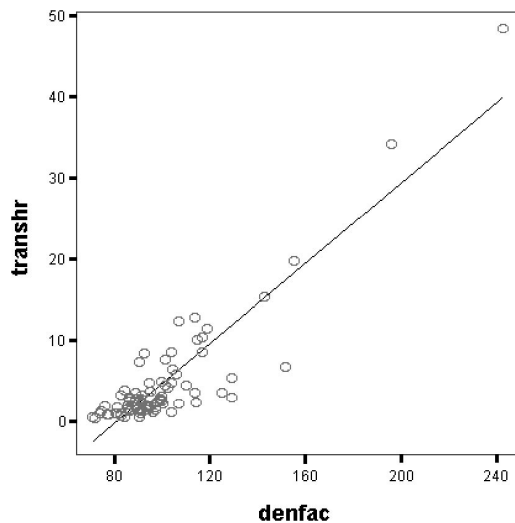


FIGURE 1 Plot of transit commute share versus the density factor (2000).

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 as well as 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 (1 standard deviation on the density scale) is associated with a 0.13 drop (25×-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 car. With a range on the density factor of 3.4 standard deviations (excluding the two outlying metropolitan areas, 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 (25×0.118) 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

TABLE 5 Outcomes Versus Sprawl (2000 Cross-Sectional)

	Transportation Outcomes							
	<i>vehph</i>	<i>transhr</i>	<i>walkshr</i>	<i>mtime</i>	<i>dlycap</i>	<i>vmtcap</i>	<i>facap</i>	<i>oz8h</i>
<i>constant</i>	-0.382	-14.11	0.566	4.77	-119.4	2.24	20.16	0.112
<i>denfac</i>	-0.00534 (-4.7) ***	0.118 (3.9)***	0.0315 (2.6)**	-0.0245 (-0.9)	-0.110 (-0.9)	-0.215 (-3.0)**	-0.105 (-2.5)*	-0.0006 (-3.8) ***
<i>mixfac</i>	0.000659 (1.5)	-0.00924 (-0.8)	0.00046 (0.1)	-0.0242 (-2.2)*	0.00728 (0.2)	0.00023 (0.0)	-0.041 (-2.5)*	0.00012 (2.0)*
<i>cenfac</i>	-0.00117 (-2.7)**	0.0351 (3.0)**	0.0199 (4.3)***	-0.0181 (-1.6)	-0.110 (-2.2)*	-0.0462 (-2.0)	-0.037 (-2.3)*	-0.00012 (-1.9)
<i>strfac</i>	0.000492 (0.9)	0.00347 (0.2)	-0.00272 (-0.5)	0.0424 (3.2)**	0.130 (3.0)**	0.0128 (0.5)	0.0149 (0.8)	-0.00014 (-2.0)
<i>metpop</i>	-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)	-9.4E-08 (-0.3)	4.27E-09 (3.5)***
<i>hhsz</i>	0.412 (7.0)***	-1.68 (-1.1)	-0.678 (-1.1)	4.32 (3.0)**	14.77 (2.9)**	1.76 (0.6)	0.667 (0.3)	0.00305 (0.4)
<i>pwkage</i>	0.0246 (4.5)***	-0.0207 (-0.1)	-0.0268 (-0.5)	0.0576 (0.4)	1.47 (3.0)**	0.667 (2.4)*	0.226 (1.1)	0.00047 (0.6)
<i>pcinc</i>	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)	-0.00032 (-2.6)*	6.22E-08 (0.1)
adjusted R²	0.56	0.67	0.36	0.61	0.63	0.28	0.44	0.40

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

efficiency of automobile use, which in turn affect vehicle ownership. But a 25-unit increase in the centers factor (1 standard deviation on the centers scale) is associated with only a 0.03 drop (25×-0.00117) in average vehicles per household, less than one-fourth the change associated with the density factor.

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 (1 standard deviation on the centers scale) is associated with only a 0.88 percentage point rise (25×0.035) in public transportation mode share, just over one-third the change associated with the density factor.

The mix factor is significant in only three cases, as a mitigating influence on travel time to work and fatal accidents and as an aggravating influence on the maximum ozone level. The last of these relationships 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, which are related to one another. Perhaps land use mix has not been successfully operationalized because of problems with the underlying data sets from the AHS or CTPP. Problems include the small sample sizes for some metropolitan areas included in the AHS and the imperfect correspondence between the metropolitan 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 TAZs, 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 1 mi 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 metropolitan areas with dense street grids translate into more total delay, most delay being occasioned at intersections instead of on the stretches between them. Conventional traffic engineers have always argued as much. Another possibility is that the TTI delay measure, which is computed instead of measured, has sufficient error attached to it 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 the authors' results, it apparently does. For 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 these relationships make intuitive sense.

Results for 1990

The same regressions were run for 1990. Results are presented by Ewing et al. (11). Here some interesting differences were discovered. 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, transit share of commute trips, and maximum ozone level.

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

The mix factor proves significantly related to average commute time and fatal accident rate in 1990, as in 2000. Indeed, it is the only important environmental correlate of 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.

CONCLUSION

This study measures sprawl in multiple dimensions and investigates its impact on an array of transportation-related outcomes. For most outcomes, sprawling regions perform less well than compact ones. This is true of everything from transit use to traffic fatalities. The exceptions are average commute time and annual traffic delay per capita, which do not clearly favor compactness over sprawl.

The main limitation of this study has to do with the data it uses. By necessity, the study uses highly aggregate data from a variety of sources that are not always consistent as to the area under study and time period. They are simply the best data available from national sources with sufficient breadth to provide a panoramic view of sprawl in the United States. Results will have to be validated through follow-up work of a more focused nature.

ACKNOWLEDGMENTS

The authors acknowledge numerous individuals who generated or processed data upon which the 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 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 data sets such as the AHS and CTPP: Robert Diogo, Danny Knee, Rachael Kennedy, Kurt Paulsen, Jee Shin, and Yue Wu.

REFERENCES

1. Ewing, R. Is Los Angeles-Style Sprawl Desirable? *Journal of the American Planning Association*, Winter, 1997, pp. 107–126.
2. *USA Today*. Feb. 22, 2001.
3. *The Dark Side of the American Dream: The Costs and Consequences of Suburban Sprawl*. Challenge to Sprawl Campaign. Sierra Club, College Park, Md., undated.
4. Galster, G., R. Hanson, M. Ratcliffe, H. Wolman, S. Coleman, and J. Freihage. Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept. *Housing Policy Debate*, Vol. 12, No. 4, pp. 681–717.
5. Kahn, M. Does Sprawl Reduce the Black/White Housing Consumption Gap? *Housing Policy Debate*, Vol. 12, No. 1, 2001, pp. 77–86.
6. Downs, A. Some Realities About Sprawl and Urban Decline. *Housing Policy Debate*, Vol. 4, No. 4, 1999, pp. 955–974.
7. Glaeser, E., M. Kahn, and C. Chu. *Job Sprawl: Employment Location in U.S. Metropolitan Areas*. Center for Urban and Metropolitan Policy, The Brookings Institution, Washington, D.C., July 2001.
8. Pendall, R. Do Land-Use Controls Cause Sprawl? *Environment and Planning B*, Vol. 26, No. 4, 1999, pp. 555–571.
9. Fulton, W., R. Pendall, M. Nguyen, and A. Harrison. *Who Sprawls Most? How Growth Patterns Differ Across the U.S.* Center for Urban and Metropolitan Policy, The Brookings Institution, Washington, D.C., July 2001.
10. Ewing, R., and R. Cervero. Travel and the Built Environment: A Synthesis. In *Transportation Research Record: Journal of the Transportation Board*, No. 1780, TRB, National Research Council, Washington, D.C., 2001, pp. 87–114.
11. Ewing, R., R. Pendall, and D. Chen. *Measuring Sprawl and Its Impact*. Smart Growth America, Washington, D.C., 2002. www.smartgrowthamerica.org.
12. Burchell, R. W., N. A. Shad, D. Listokin, H. Phillips, A. Downs, S. Seskin, J. S. Davis, T. Moore, D. Helton, and M. Gall. *TCRP Report 39: The Costs of Sprawl—Revisited*. TRB, National Research Council, Washington, D.C., 1998.

Publication of this paper sponsored by Committee on Transportation and Land Development.