Effects of Urban Sprawl on Obesity

Zhenxiang Zhao*

Department of Economics

University of Illinois at Chicago

October 2007

* I thank my advisor Bob Kaestner. I also thank my committee members, Nathan Anderson, Evelyn Lehrer, Daniel McMillen and Lisa Powell. In addition, Tim Classen, Lorens Helmchen, Joshua Linn, Tony Lossaso, Surrey Walton, Xin Xu and participants in the UIC Health Economics and Health Policy seminars provided helpful comments and advice. A Pre-Doctoral Fellowship from the Chicago Center of Excellence in Health Promotion Economics and the Provost Award for Graduate Research from UIC provided financial support. All errors are mine.

Abstract

One often mentioned explanation of the significant rise in obesity in the U. S. is an increasingly sedentary lifestyle and an often cited cause of this lifestyle change is urban sprawl. However, whether urban sprawl is a cause of a sedentary life is unknown as the interpretation of this relationship is confounded by self selection of residents. This paper addresses the "selection" issue by using exogenous variation in urban sprawl caused by the original plan of the Interstate Highway System. By combining data from the National Health Interview Survey (1976-2001) and the U.S. Census (1970-2000), this paper assesses the extent to which urban sprawl has contributed to the increase in obesity. Empirical estimates show that a 10 percentage point decrease in the proportion of population living in dense areas leads to about a 1.5 percentage point increase in the obesity rate. If the average metropolitan area had not experienced the exogenous decline in population density, the estimates thus indicate that the rate of obesity would have been reduced by about 18%.

I. Introduction

Over the past thirty years, the prevalence of obesity in the US increased dramatically. Between 1960-1962 and 2004, the proportion of adults who were obese increased from 13.4% to 32.2% (Flegal at al. 2002; Ogden et al. 2006). This increase is worrisome because obesity has been found to be significantly associated with type II diabetes, high blood pressure, high cholesterol, asthma and poor health status. Obesity has been estimated to have caused 365,000 deaths in 2000 (Mokdad et al., 2005). Further, obesity-related morbidity has been estimated to account for 9.1% of total annual U.S. medical expenditures in 1998 ($92.6 billions in 2002 dollars).[1] Public financing of these costs is considerable since half of all health care is paid by government through Medicare and Medicaid (Finkelstein et al., 2003).

The spatial distribution of the population in the United States also has changed over the same period. Between 1950 and 2000, the share of the population living in metropolitan areas has grown from 56% to 80% (Transportation Research Board Special Report 282, 2005). While a greater proportion of the population is living in urban areas broadly defined, all of the growth in metropolitan areas occurred in suburban areas, as central cities actually declined in population (Baum-Snow, 2007). In 1950, the population of metropolitan areas was roughly evenly divided between the suburban fringe and the central city; currently, approximately two-thirds of the population of metropolitan areas resides in the suburbs, and this proportion has been rising (Pisarski, 2001). Table 1 shows that between 1970 and 2000, the population weighted population density for 53 major metropolitan areas has fallen over 19% with more dramatic declines observed for the densest parts of metropolitan areas.

[1] The Surgeon General's report has estimated the total cost of obesity to US was $117 billion in 2000 (USDHHS, 2001).

Urban sprawl, characterized by low-density development patterns, has been found to be related to higher rates of obesity (Ewing, 2003; Frank et al., 2004; Lopez, 2004; GilesCorti et al., 2003; Saelens et al., 2003).[2] Increased distance between home and destination, and poor accessibility to amenities that come with low-density development patterns increase the reliance on the automobile for many purposes and minimize walking. Furthermore, the lack of sidewalks or bicycle trails and the cul-de-sac street layouts of large residential developments may also affect residents' propensity to be physically active by walking less or bicycling less. In addition, greater availability of large chain stores (big box stores) in the suburbs would lead to lower costs of food. All of this is hypothesized to result in changes of obesity rates (Berrigan et al., 2002; Saelens et al., 2003; Cervero et al., 1995; Handy, 1996; Hess et al., 1999; Boarnet et al., 2000; Crane et al., 1998; Chung, 1999).

Results from previous studies has led the Centers for Disease Control and Prevention (2003) and the World Health Organization (2004) among others to advocate for using community (re)design as a tool to curb the rise in obesity.[3] However, a key policy question, which previous research has not adequately addressed, is to what extent urban sprawl causes people to weigh more. While there have been several studies that have examined the associations between urban sprawl and obesity, these studies have been limited in their ability to provide estimates of a causal relationship. The primary limitation of previous studies is that the relationship between urban sprawl and obesity is likely to be confounded by self selection of residents. For example, residents who dislike walking are more likely to be obese and to choose to live in places in which one can easily get around by car. There may be other confounding

[2] In the paper, urban sprawl, built environment and low population density are used interchangeably. With little agreement on a precise definition, urban sprawl is generally defined as the low density development pattern which changes the built environment in which individual resides.
[3] http://www.cdc.gov/nccdphp/dnpa/obesity/contributing_factors.htm

factors. Therefore, the observed relationship found in previous studies might reflect self-selection rather than supporting the hypothesis that urban sprawl and the built environment itself causes residents to be more or less obese.[4] If the cause of obesity is the built environment, then public policies can be devised to limit the harmful effects of urban sprawl by promoting types of development that increase physical activity and encourage healthy eating behaviors. If however, the self-selection is the explanation for the association between urban sprawl and obesity, then policies aimed at curbing urban sprawl would have little effects on obesity.

Part of the difficulty of obtaining causal estimates is due to the absence of exogenous variation in urban sprawl. The empirical strategy used in this paper exploits the plausibly exogenous variation in urban sprawl caused by the United States Interstate Highway System. The falling cost of intra-urban transport, which is largely due to the construction of the interstate highway system, is often cited as one of the causes of urban sprawl (Mieszkowski et al., 1993; Glazer et al., 2003). In this study, the original 1947 plan for the Interstate Highway System is used to identify the exogenous variation in urban sprawl, as the main purpose of the original Interstate Highway System was to connect major metropolitan areas, to serve national defense, and to connect the United States to Canada and Mexico (Federal-Aid Highway Act of 1944). It was not designed to facilitate local commuting. Therefore, using the original plan instead of the actual highway avoids the potential endogeneity of the highway construction. For example, the planned interstate highway might be altered or new interstate highways might be added during the construction period to meet the local commuting needs resulting from changed preferences for walking and use of the automobile.

[4] There are two exceptions: Eid et al. (2006) take advantage of the panel dimension of the NLSY79 and try to identify the causal effects of urban sprawl on obesity. However, their research can only eliminate time-invariant unobserved factors and may still subject to time-variant unobserved factors. Plantinga et al. (2007) use marital status and family size as instruments for urban sprawl. However, the validity of instrument is arguable as single people are less likely to be obese because they have the pressure from the marriage market.

Another part of the difficulty stems from the lack of the relevant data. To the best of my knowledge, there are no single data sets containing information on individual's weight and height, along with individual characteristics that may affect their weight status, and the built environment where individual resides for last three decades. In addition, even if I had the data with all the information described above, the possible endogeneity of the key urban sprawl variable would lead to biased estimates as individuals do not choose where they live randomly. Therefore, in the paper, I employ the two-step model (Murphy et al. 1985) to get around this data limitation. Specifically, in the first step I use the original 1947 highway plan to predict population density of metropolitan areas, a proxy for urban sprawl and the built environment, using the decennial Censuses of 1970, 1980, 1990 and 2000. In the second step, I use this estimated relationship to predict population density of metropolitan areas where individual lives using the National Health Interview Surveys 1976 to 2001, and then obtain estimates of the effect of (predicted) population density on obesity.

I find that there is a negative effect of population density on obesity and the results are robust across a wide range of specifications. Estimates in the paper indicate that if the average metropolitan area had not experienced the decline in the proportion of population living in dense areas in the last 30 years, the rate of obesity would have been reduced by about 18%. The rest of the paper proceeds as follows. Section II describes the empirical methodology adopted in the paper. Section III describes the data while section IV discusses the results. Finally, section V concludes.

II. Empirical Methodology

Figure I provides an overview of the empirical methodology adopted in the paper. Data on population density for 53 large metropolitan areas identified in the public-use National Health Interview Surveys data are obtained from the 1970, 1980, 1990 and 2000 Censuses. Population density is defined as the proportion of population living in dense areas. Different thresholds of density are used and an area is defined as a census tract. As noted, population density of an MSA may be correlated with factors that are likely unobserved and which may affect obesity. Therefore, I use the 1947 Interstate Highway plan to predict MSA population density in the first step. The second step is to use the predicted population density in a regression model of the determinants of obesity. Data on individuals' demographic information, their weight and height information are obtained from the National Health Interview Surveys 1976 to 2001. I now turn to a more detailed description.

The ultimate goal of the analysis is to obtain estimates of the effect of population density on individual's weight status. Equation (1) describes the model to be used.

$$\begin{aligned} &Obese_{ijt}(BMI_{ijt}) = \alpha_j + \gamma_t + \beta_1 X_{ijt} + \beta_2 PopDen_{jt} + \gamma Z_{jt} + \mu_{ijt} \\ &i = 1,...,N \quad \text{(persons)} \\ &\mathrm{j} = 1,...,\mathrm{J} \quad \text{(MSAs)} \\ &\mathrm{t} = 1976,...,2001 \quad \text{(years)} \end{aligned} \tag{1}$$

In equation (1), Obesity is measured by the body mass index (BMI) defined as weight in kilograms divided by height in meters squared (kg/m2). Persons with BMI ⩾30 kg/m2 are classified as obese. Probability of being obese (BMI) of person i in MSA j and year t depends on MSA fixed effects (α_j); year effects (γ_t); individual characteristics (*X*) such as age, race, sex, marital status and education; MSA population density (*PopDen*); and time-varying MSA characteristics (Z_{jt}) such as median family income and employment rates. As noted, population density is measured as the proportion of population living in dense areas where different thresholds are used to define dense areas to capture the degree to which population are centered in high density living areas.

In order to address the potential endogeneity of population density, I predict population density using the original 1947 plan of the Interstate Highway System.

$$PopDen_{jt} = \alpha'_j + \gamma'_t + \gamma'^2_t + \beta[(Year_t - 1947) * HWPLAN_j] + \gamma' Z_{jt} + \varepsilon_{jt}$$
$$j = 1,...,J \quad \text{(MSAs)} \qquad (2)$$
$$\text{t} = 1970,1980,1990,2000$$

In equation (2), population density of MSA j in year t depends on MSA fixed effects (α'_j), a quadratic time trend (γ'_t and γ'^2_t), and the number of planned highway rays in MSA j. The underlying logic of this approach is that the highway plan originated in 1947 was to be implemented over time and therefore, population density of an MSA would be affected in a diffused way over time. This diffusion process is captured by the specification of equation (2), which allows the effect of the highway plan to vary over time in a linear fashion. Other specifications, for example, quadratic were tried and rejected on the basis of statistical tests.

Models similar to equation (2) have been estimated before. Baum-Snow (2007) showed that metropolitan areas with a greater number of planned highways lost central city population over the last half of the 20th century. I find similar effects with respect to population density, which I present below. In sum, this evidence supports the approach I use because it demonstrates that the highway plan is a significant predictor of MSA population density. The interstate highway system changed the spatial distribution of the population and related development patterns, providing a quasi-experimental opportunity to examine how urban sprawl affects obesity.

As is discussed above, the original plan of 1947 was motivated by concerns related to national defense, to connect far away places and not to facilitate local commuting. The Federal-Aid Highway act of 1944 called for designation of a National System of Interstate Highways, to include up to 40,000 miles "…so located as to connect by routes, as direct as practicable, the principal metropolitan areas, cities, and industrial centers, to serve the national defense, and to connect at suitable border points with routes of continental importance in the Dominion of Canada and the Republic of Mexico". On August 2, 1947, Commissioner MacDonald and Federal Works Administrator Philip B. Fleming announced selection of the first 37,700 miles. The routes had been proposed by the State highway agencies and reviewed by the Department of Defense to meet the needs of national defense. A map of this 1947 plan is presented in figure II. Therefore, it is plausible that, conditional on controlling for MSA fixed effects, the original planned number of highways is exogenous—uncorrelated with unmeasured determinants of changes in obesity within an MSA.

III. Data

I use a variety of data sources to construct a data set for this analysis: individuals' demographic information, their weight and height status and also metropolitan area identifiers from the National Health Interview Survey (NHIS) from 1976 to 2001; the normalized census tract level data from the Neighborhood Changing Data Base for 1970, 1980, 1990 and 2000; the metropolitan area-county crosswalk file from historical metropolitan area definitions; the planned number of highway rays for metropolitan areas from the General Location of National System of Interstate Highways (1956); and the median family income and employment rates at metropolitan areas level from the Current Population Survey (CPS) March file from 1976 to 2001.

National Health Interview Survey

I used repeat cross sections from the National Health Interview Survey (NHIS) for the years 1976-2001 to obtain the information on individual's weight, height and their demographic characteristics including:

- age (18 and over);
- race (four race/ethnicity categories: non-Hispanic White, non-Hispanic Black, Hispanic and others);
- sex;
- education (five education categories describing the highest grade individual completed: elementary school, some high school, high school graduate, some college and college graduate);
- marital status (four marital status categories: single versus married, separated/divorced, or widowed);
- metropolitan area identifiers of where they live.[5]

Family income was not included in the primary specifications as the NHIS doesn't have a consistent measure of family income throughout the time period used in this study. The NHIS public use data present family income information in different categories. The cutoffs for the income categories are inconsistent throughout the time period. While more precise measures are available in each of the NHIS years, inconsistencies in the year-to-year coding make it less possible to construct a consistent income variable. Instead, I include metropolitan level median

[5] I used 1976 as the first year of the sample because 1976 is the first year NHIS has individual's weight and height information in the core data set. I used 2001 as the ending year because 1) the construction of the interstate highway was almost finished by 2001 and 2) it is the last year that NHIS public use data have MSA identifiers.

family income in some specifications to control for the income variation at the metropolitan level. The main results, however, remain unchanged.[6]

The metropolitan area identifier is the lowest geographic identifier available in the public-use NHIS data.[7] In addition, the NHIS only identifies large metropolitan areas and the number of metropolitan areas identified in the NHIS increased approximately every ten years. For the years 1976-1984, there are 31 largest metropolitan areas (SMSA/SCAs) identified in the public use file;[8] for the years 1985-1994 there are 33 largest metropolitan areas (MSA/CMSAs) and 16 selected sub-areas identified; for the year 1995-1996 there are 52 largest metropolitan areas (MSA/CMSAs) and 32 selected sub-areas identified; for the year 1997-2001 there are 45 largest metropolitan areas (MSA/CMSAs) and 19 selected sub-areas identified. In the final sample, I include metropolitan areas that are identified in the NHIS for at least one of the three sampling periods (1976-1984, 1985-1994, 1995-2001) throughout which the geographic definition of those areas did not change.[9] I also exclude metropolitan areas that have no central city.[10] Honolulu, HI is also excluded from the sample as no interstate highway connecting Honolulu with continental America. This leaves 53 unique metropolitan areas in the final sample.[11]

Metropolitan areas are defined by the U.S. Office of Management and Budget (OMB). The general concept of a metropolitan area is that of a core area containing a substantial

---

[6] In the alternative specifications not presented in the paper, I also included a coarse measure of family income as an additional covariate. I calculate the mid-point of each income category and create a continuous income measures. This income measures, of course, subject to measurement errors as cutoffs of income categories are different across years. The regression results, however, are not significantly different from those of the primary specification.

[7] Respondents who either live in non-MSA areas or not self-representing MSA areas are excluded from the estimation sample due to lack of their geographical location information.

8 SCA stands for Statistic Consolidated Area; SMSA stands for standard metropolitan statistical area; MSA stands for metropolitan statistical area; and CMSA stands for consolidated metropolitan statistical areas.

[9] Those metropolitan areas which are only identified for two years (1995, 1996) in NHIS are excluded from the sample: Dayton, OH MSA; Ventura, CA PMSA; Vineland, NJ PMSA; Wilmington, DE PMSA; Richmond, VA MSA; Santa Cruz, CA PMSA; Santa Rosa, CA PMSA; Albany, NY MSA; Bridgeport, CT PMSA; Stamford, CT PMSA; Danbury, CT PMSA; Jersey City, NJ PMSA; Dutchess County, NY PMSA; New Heaven, CT PMSA; New Burgh, NY PMSA; Trenton, NJ PMSA; Waterbury, CT PMSA; Birmingham, AL MSA; Vallejo-Fairfield-Napa, CA PMSA.

[10] Those metropolitan areas that have no central cities are also excluded from the sample: Bergen-Passaic, NJ PMSA; Middlesex, NJ PMSA; Nassau-Suffolk, NY PMSA.

population nucleus, together with adjacent communities having a high degree of economic and social integration with that core. The metropolitan areas standards are revised according to new information obtained from each decennial Census. Changes in the definitions of metropolitan areas since the 1950 census have consisted chiefly of: the recognition of new metropolitan areas when the requirement on population was reached; the addition of counties to existing metropolitan areas; transfer of counties from one area to another; and dropping of counties from an area due to changes in population, the economic or social tie to the central counties of metropolitan areas. The large majority of changes have taken place on the basis of decennial census data.

For data from the NHIS 1976-1984, metropolitan area identification codes are based on the 1971 metropolitan area definition. [12] From the years 1985-1994, metropolitan areas identification codes in the NHIS are based on the 1983 metropolitan areas definitions. Starting with the 1995 NHIS, metropolitan areas identification codes are based on the 1993 metropolitan areas definition.[13] Because of these historical changes in geographic definitions, caution must be taken in comparing data for these statistical areas from different dates. For example, most metropolitan areas encompass less territory during earlier years than in later years and those newly included areas are generally less dense compared to those already included areas. If, for some unobserved reasons, people living in less dense areas are more likely to be obese, changing definition alone would lead to the conclusion that decreased density leads to higher obesity rates without proper controls of MSA fixed effects.

This feature of the data implies a slight modification to equation (1):

---

[11] Out of these 53 metropolitan areas, 31 of them were available since 1976; 7 more metros were available since 1985; 15 more metros were available since 1995.

[12] Information about whether the household was located in an identified MSA was added by the National Center for Health Statistics (NCHS), not collected from respondents.

[13] Information was collected based on the personal communication with NHIS staff.

$$Obese_{ijt}(BMI_{ijt}) = \alpha_{j_{p1}} + \alpha_{j_{p2}} + \alpha_{j_{p3}} + \gamma_t + \beta_1 X_{ijt} + \beta_2 PopDen_{jt} + \gamma Z_{jt} + \mu_{ijt} \quad (1`)$$

P1=1976-1984 when MSA codes are based on the 1971 metropolitan area definition in NHIS;

P2=1985-1994 when MSA codes are based on the 1983 metropolitan area definition in NHIS;

P3=1995-2001 when MSA codes are based on the 1993 metropolitan area definition in NHIS;

In equation (1`), MSA fixed effects are period specific (p1, p2 and p3) which I refer to as unrestricted MSA fixed effects and there is an unbalanced panel of MSAs; some MSAs contribute observations in each of the three periods and others contribute observations only in the latest period. Definitions, means, and standard deviations of all variables employed in the NHIS are reported in Appendix Table 1.

U.S. Census Data

The US Census from 1970, 1980, 1990 and 2000 were used to calculate population density for each MSA. The constant-geography (boundary) census tract level data were obtained from the Neighborhood Changing Data Base (NCDB). Because the boundaries of many census tracts changed between the decennial censuses (either splitting into several tracts or merging with other tracts), the comparisons of tracts across years would be difficult without linking these tracts to standardized geographic boundaries.[14] Specifically, the NCDB has remapped 1970, 1980, and 1990 tract level census data to Census 2000 tract boundaries to form the normalized census tracts across years. Unfortunately, NCDB only normalize the Census data forwards, from 1970, 1980 and 1990 data to 2000 but not backwards. Therefore, in order to get the constant-geography MSA level data based on three-period MSA definitions (1971 definitions, 1983 definitions and 1993 definitions), I first aggregate the normalized census tract data into 2000-

[14] For example, 49 percent of all 2000 census tracts experienced boundary changes since the 1990 census.

definition county level data as changes in boundaries of counties happen rather rarely.[15] Then I create the county-MSA cross walk file based on historical metropolitan area definitions to link counties to the 1971-definition MSAs, the 1983-definition MSAs and the 1993-definition MSAs.[16] Few counties with changing boundaries were restored to form their original counties when matching counties to MSAs. By doing this, I am able to obtain information of, for example, the 1971-definition MSA for each census year between 1970 and 2000, though the 1971-definition MSAs would not exist any longer after 1980. In order to match with 53 MSAs identified in the public-use NHIS data, 122 constant-geography MSAs were created and each MSA has four observations for census years 1970, 1980, 1990 and 2000.[17]

Here too, this feature of the data has implications for model specification. I can illustrate this with a modified equation (2).

$$PopDen_{jt} = \alpha'_{j_1} + \alpha'_{j_2} + \alpha'_{j_3} + \gamma'_t + \gamma'^2_t + \beta[(Year_t - 1947) * HWPLAN_j] + \varepsilon_{jt} \qquad (2`)$$

Note that in equation (2`), I include up to three intercepts (fixed effects) for each unique MSA based on three-period MSA definitions and there are 488 constant-geography MSA-year observations.

Using these constant-geography MSA definitions, I calculate the population density of each census tract within each MSA for census years 1970, 1980, 1990 and 2000. I define a census tract as a "dense area" based on the calculated population density. Different threshold values have been adopted to define dense areas to capture the degree to which population are centered in high density living areas. The lowest threshold used is population density of 3,500 or more people per square mile. This threshold is the density at which people begin to use non-

[15] There were no county consolidated and only relatively few counties breaking off.
[16] County-level data can be aggregated to MSAs with complete accuracy outside the New England states. I assign one MSA identifier to those counties across MSA boundaries based on where the largest percentage of the population lived at the time of the 1980 census.

automobile modes of transportation (Lopez, 2004). The second threshold for dense areas is defined as population density of 5,000 or more people per square mile. About 50% of census tracts in the constant-geography MSAs have population density less than 5,000 people per square mile. The third threshold for dense areas is defined as population density of 9,000 or more people per square mile. About 75% of census tracts in the constant-geography MSAs have population density less than 9000 people per square mile. The highest threshold I use is the population density of 12,500 or more people per square mile, which is the lower limit of density needed to support mass transit (Ewing, 2003). By aggregating dense census tracts population within each MSA and dividing them with the total population living in the MSA, I obtain the proportion of population living in dense areas for each MSA. Further analysis of the data (Appendix Table 2) showed that all the decreases in the population density actually happened in the areas with density greater than 5000 people per square mile. Therefore, in the regressions, I will focus on the proportion of population living in dense areas with thresholds 5000, 9000 and 12500.

Highway Plan Data

The Federal-Aid Highway Act of 1944 authorized designation of a 40,000-mile Interstate Highway System but did not create a program to build it until 1947. On August 2, 1947, Public Roads Administration (PRA) designated approximately 37,700 miles of the nation's principal highways. The process was completed in September 1955 when the General Location of National System of Interstate Highways was published by Bureau of Public Roads.

The basic highway measure used in the paper is the number of planned highway rays emanating from the largest central city of an MSA. Therefore, I count 2 rays for a planned highway passing through the central city. For those planned highways ending in or near the

[17] 31 of them are 1971-definition MSAs, 38 of them are 1983-definition MSAs and 53 of them are 1993-definition MSAs.

central city, one ray is counted for each of them. Although the authorized interstate highway system by the Federal-Aid Highway Act of 1944 is still under construction, virtually the entire 1947 planned interstate highway system was built by 2000 (FHWA, 2001).

Current Population Survey

I used CPS March file to calculate (weighted) MSA level median family income and employment rates. Definitions of MSA identifiers in CPS also changed over time based on OMB's definitions. I merged the CPS data to the appropriate MSA definition.

IV. Results

To summarize, the objective is to obtain estimates of equations (1`) and (2`), which are reproduced here:

$$Obese_{ijt}(BMI_{ijt}) = \alpha_{j_{p1}} + \alpha_{j_{p2}} + \alpha_{j_{p3}} + \gamma_t + \beta_1 X_{ijt} + \beta_2 PopDen_{jt} + \gamma Z_{jt} + \mu_{ijt} \quad (1`)$$

$$PopDen_{jt} = \alpha'_{j_1} + \alpha'_{j_2} + \alpha'_{j_3} + \gamma'_t + \gamma'^2_t + \beta[(Year_t - 1947) * HWPLAN_j] + \varepsilon_{jt} \quad (2`)$$

The first step is to obtain estimates of equation (2`) using the Census data. Then using estimates from equation (2`), I predict population density and merge these predicted measures of population density to the NHIS data and estimate equation (1`). This two-step approach is similar to, but not the same as the two-sample instrumental variable approach of Angrist et al. (1992) as here equation (2`) is just an auxiliary regression to predict density, a proxy for the rate of urban sprawl, which is missing in the NHIS data. The assumption is that by using the 1947 Highway Plan, I have constructed an exogenous measure of population density. Later, I present a similar analysis using the conventional instrumental variables approach.

To obtain estimates of equations (1`) and (2`), I fit linear probability models rather than logit or probit models when the dependent variable is binary, given the large sample size. Standard errors are corrected using Murphy-Topel estimate of variance-covariance that recognizes the predicted nature of population density.

Table 2 presents the results from the first step estimates shown in the equation (2`). Estimated coefficients on planned highway rays have the expected negative sign and are statistically significant at better than the 1% level. Regression estimates indicate that, conditional on control variables, each additional planned highway ray will decrease the proportion of population living in dense areas by about 1.4-2.5 percentage points every 20 years. One might be concerned that the planned highway rays might be correlated with other MSA level time-varying factors that would affect individual weight status. Therefore, as a check on the exogeneity of the highway plan, MSA-level median family income and employment rates are included as an alternative specification in both steps. If the plan is exogenous, the inclusion of those MSA level time-varying variables should have little impacts on the estimated effects of urban sprawl. Column (3), (5) and (7) in Table 2 show that conditional on time-varying MSA level control variables, planned highway rays are still good predictors of population density with little change in the coefficients when these variables are added. Metropolitan median family income has a positive effect on population density which is consistent with previous literature (Glaeser et al., 2003).

By comparing the effects of planned highway rays across different dense area definitions, it implies that the largest decline of population shares occurs in areas with higher density, as one additional planned highway ray decreases about 9.7% of the population living in areas with density greater than 12, 500 people per square mile, while it only decreases about 5% of the population living in areas with density greater than 5,000 people per square mile. These findings are consistent with the prediction of the land use theory that central city population would decline with the construction of highway rays and population in metropolitan areas would spread out along highways (Baum-Snow, 2007). Figure 1, 2, 3 show the simulated effects of planned highway rays on the proportion of population living in dense areas with various thresholds for dense areas. The graphs illustrate nicely the relatively large effects of the highway plan on urban density. For example, two planned highway rays is associated with approximately 50% decline in the measures of urban density.

I now turn to estimates of the effect of population density on BMI and obesity. Table 3 presents these estimates for obesity and Table 4 presents estimates for BMI. Estimates in Table 3 have the hypothesized signs and are statistically significant, and indicate that population density is negatively associated with obesity. Estimates in Table 3 imply approximately 0.1 to 0.2 percentage points increases in the obesity rate for each additional percentage point decrease in the proportion of population living in dense areas. In addition, the effects are larger for higher density thresholds. For example, estimates associated with a density of 12500 are twice as large as estimates associated with a density of 5000. This pattern is consistent with the underlying hypothesis that high population density is associated with less automobile use and more walking. Notably, estimates are not sensitive to the inclusion of time-varying MSA controls for median family income and employment rates. Moreover, median family income has a statistically significant effect on BMI (shown below) and population density. This provides some evidence supporting the identification assumption, which assumes that changes in population density caused by the 1947 highway plan are uncorrelated with changes in other attributes of the MSA that are correlated with changes in weight status.

Table 4 presents estimates of the effect of population density on BMI. Estimates indicate that a one percentage point decrease in the proportion of population living in dense areas increases BMI by about 0.01 units and estimates are not significantly different from zero. Interestingly, comparing the effects on BMI and obesity suggests that instead of shifting the whole weight distribution to the right, the decline in population density has larger effects on the upper tail of the weight distribution.

Coefficients of the individual characteristics on obesity and BMI generally have the expected signs and are consistent with the previous literature (estimates are not presented). Age has an inverted U-shaped effect on the probability of being obese and BMI. Non-Hispanic, Black and Hispanic persons are more likely to be obese and have higher values of BMI, while persons of other races are less likely to be obese and have lower values of BMI compared to Whites. Males are less likely to be obese, but they have higher BMI levels compared to women. Married and divorced individuals are less likely to be obese and widowed persons are more likely to be obese compared to single (never married) individuals while married and widowed person have higher BMI levels.[18] Years of formal schooling completed have negative effects on the probability of being obese and the BMI.

Instrumental Variables Estimates

To obtain instrumental variables estimates, I use only Census year data from the NHIS: 1980, 1990, and 2000, where I can match the population density measures from the Census data directly to the individual-level data from the NHIS. Summary statistics for the "matched" NHIS sample are shown in Appendix Table 1 and indicate that the matched sample is very similar to the full sample. I perform three different estimations using this matched sample. First, OLS is used to estimate the effects of urban sprawl on obesity directly where the urban sprawl measure is the actual proportion of population living in dense areas in each metropolitan areas rather than predicted values. Thus, the estimated coefficients on the urban sprawl variables would be biased due to the possible endogeneity problem. I then estimate the effects using the conventional instrumental variable approach where I use planned highway rays as the excluded instrument. Specifically, I estimate the following models:

$$Obese_{ijt}(BMI_{ijt}) = \alpha_j + \gamma_t + \beta_1 X_{ijt} + \beta_2 PopDen_{ijt} + \mu_{ijt} \qquad (3)$$

$$PopDen_{ijt} = \alpha'_j + \gamma'_t + \beta_1 X_{ijt} + \beta[(Year_t - 1947) * HWPLAN_j] + \varepsilon_{jt} \qquad (4)$$

The important aspect of these models is that I include only one intercept (fixed effect) for each MSA. Recall that in the full sample, I control for unrestricted MSA fixed effects to account for the potential bias due to changes in MSA geographic definitions. However, in the matched sample, it is impossible to control unrestricted MSA fixed effects as they are perfectly correlated with the planned highway measures. Therefore, I treat each MSA as if it was the same in each Census year. In addition, Equation (4) now includes year fixed effects instead of a quadratic time trend. Equation (3) and (4) are estimated using a sample 80,580 individuals. In order to add weight to the validity of my previous two-step model results, I perform the two-step procedure on this matched sample and compare the results with the IV results. Table 5 reports the first stage results for the IV and two-step procedure and table 6 and table 7 show the OLS, IV and two-step procedure results.

Table 5 reports the first stage results for the IV and two-step model estimates. The first stage results for the IV and two-step procedure are very similar to each other with F values

[18] One explanation for this is previous findings of marital status may also capture the built environment effects on obesity as married individuals are more likely to live in suburban areas with less population density. Therefore, after controlling for density, the "true" effects of being married on obesity would be negative.

significant at 1% level, indicating that the 1947 highway plan is associated with a decrease in population density. The results of second stage from the two-step and IV model are also very similar to each other, both in terms of point estimates and standard errors. Thus, I conclude that the two-step model I adopted in the full sample produces estimates similar to what would be obtained from a conventional IV model.

One notable pattern in the results of matched sample is that instrumental variables estimates of the effect of population density on obesity exceed OLS estimates. If the only source of endogeneity of urban sprawl is that individuals who are obese are more likely to choose to live in less dense areas as they prefer driving to walking, the instrumental variables approach would produce smaller estimates than those of OLS in absolute value. The reverse ordering is consistent with the existence of an additional bias, for example, low density areas are the types of areas for which other unobservables would tend to produce thinner people. One potential explanation is the heterogeneous demand for locally provided public goods. Suppose, for example, thinner people who invested more on their health also put more emphasis on their children's human capital investment. Therefore they may choose to reside in suburban areas in search of better public schools (Mieszkowski et al., 1993; Wassmer et al, 2005).

Conclusion

Previous research has documented a positive association between obesity and urban sprawl. Whether this association represented a causal relationship, however, was not addressed. In this paper, I addressed the causality issue by using a plausibly exogenous source of urban sprawl—the decline in population density caused by the Interstate Highway plan. Estimates indicate that a one percentage point decrease in the share of population living in dense areas increased the prevalence of obesity by 0.1 to 0.2 percentage points depending on which threshold of dense area was used.

To place these results in context, I evaluate the importance of urban sprawl in explaining the rising trend in obesity by examining the counterfactual obesity trend did urban sprawl not exist. Obesity rates increased on average by 145% from about 8.3 percentage points to 20.3 percentage points between 1976 and 2001 in the NHIS sample used in the paper. I take the predicted values of population density from the first stage to determine the "exogenous urban sprawl" by differencing population density between 2001 and 1976 resulting from 4 rays of planned highways (4 rays are the mean of the planned highway rays). The resulting number is about 10 percentage points. I then multiply this number (10 percentage points) by 0.15 (the coefficient of the population density with the threshold of 9000) to derive the average percentage point difference in obesity that can be attributed to the exogenous change in population density. I divide the estimated difference in obesity rates due to exogenous urban sprawl (1.5 percentage points) with the observed difference in obesity rates between 2001 and 1976 (12 percentage points). The estimates thus indicate that about 13% increases in obesity can be attributed to urban sprawl. If the average metropolitan areas had not experienced the exogenous decline in the proportion of population living in dense areas, the increase of the obesity rates would have been reduced by about 18%.

Overall, my results suggest that urban sprawl did cause an increase in obesity, but its effect was relatively modest. Thus, policy makers may want to look elsewhere for solutions to the obesity problem, particularly if urban and community redesign are costly. While this study has identified a relationship between population density and obesity, it did not identify the underlying mechanisms that link urban sprawl to weight. Future research is warranted to better understand the mechanisms through which urban sprawl has caused the changes in obesity rates documented in this paper.

References

Anderson P., Butcher K., Levine P., 2003. Maternal employment and overweight children. Journal of Health Economics 22(3):477–504

Angrist, J., Krueger, A., 1992. The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples. Journal of the American Statistical Association 87(418):328-336.

Baum, C., 2007. The effects of race, ethnicity, and age on obesity. Journal of population economics 20:687-705.

Baum-Snow, N., 2007. Did highways cause suburbanization? Quarterly Journal of Economics 122(2): 775-805.

Berrigan D, Troiano R.P., 2002. The association between urban form and physical activity in U.S. adults. American Journal of Preventive Medicine 23(2S):74–79.

Boarnet, M., Greenwald, M., 2000. Land use, urban design, and nonwork travel: Reproducing other urban areas' empirical test results in Portland, Oregon. Transportation Research Record 1722: 27-37.

Census March CPS Manual for 1994, 1996, Census Bureau.

Census Users Note 1, 1996, Census Bureau.

Cervero, R., Roger G., 1995. Commuting in transit versus automobile neighborhoods. Journal of the American Planning Association 61: 210-25.

Centers for Disease Control and Prevention, Director, Gerberding, Julie L, 2003. cdc's role in promoting healthy lifestyles. Statement by Julie L. Gerberding before the Senate Committee on Appropriations, Subcommittee on Labor, hhs, Education and Related Agencies.

Chou, S.Y., Grossman, M., Saffer, H., 2004. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. Journal of Health Economics 23(3):565–587.

Chung, C., Myers Jr., S.L., 1999. Do the poor pay more for food? an analysis of grocery store availability and food price disparities. Journal of Consumer Affairs 33 (2): 276–

296.

Cutler, D., Glaeser, E., Shapiro, J., 2003. Why have Americans become more obese? Journal of Economic Perspectives 17(3):93–118.

Crane, R., Crepeau, R., 1998. Does neighborhood design influence travel? A behavioral analysis of travel diary and GIS data. Working paper no. 374, Transportation Center, University of California, Berkeley.

Eid, J., Overman, H.G., Puga, D., Turner, M.A., 2007. Fat city: questioning the relationship between urban sprawl and obesity. Manuscript.

Ewing, R., Schmid, T., Killingworth, R., Zlot, A., Raudenbush, S., 2003. Relationship Between Urban Sprawl and Physical Activity, Obesity and Morbidity. American Journal of Health Promotion 18: 47-57.
Federal-Aid Highway Act of 1944, U.S. Department of Transportation.

F. H. W. A. U.S. Department of Transportation. “Route Log and Finder List”. U.S. Government Printing Office, Washington, DC (2001).

Flegal, K.M., Carroll, M.D., Ogden, C.L., Johnson, C.L., 2002. Prevalence and trends in obesity among U.S. adults, 1999–2000. JAMA 288(14):1723–1727.

Frank, L., 2000. Land use and transportation interaction: Implications on public health and quality of life. Journal of Planning Education and Research 20: 6-22.

Frank, L., Andresen, M., Schmid, T., 2004. Obesity Relationships with Community Design, Physical Activity, and Time Spent in Cars. American Journal of Preventive Medicine 27: 87-96.

Finkelstein, E.A., Fiebelkorn, I.C., Wang, G., 2003. National medical spending attributable to overweight and obesity: How much, and who’s paying? Health Affairs W3: 219–226.

General Location of National System of Interstate Highways, 1956. Bureau of Public Roads

GilesCorti, B., S. Macintyre, J. P. Clarkson, T. Pikora, and R. J. Donovan. 2003. Environmental and lifestyle factors associated with overweight and obesity in Perth, Australia. American Journal of Health Promotion 18(1):93–102.

Glaeser, E., Kahn, M., 2003. Sprawl and urban growth. Working paper9733, National Bureau of Economic Research.

Handy, S., 1996. Understanding the link between urban form and nonwork travel behavior. Journal of Planning Education and Research 15: 183-98.

Hess, P., Moudon, A., Snyder, M., Stanilov, K., 1999. Site design and pedestrian travel. Transportation Research Record 1674:9-19.

Huber, P.J., 1967. The behavior of maximum likelihood estimates under nonstandard conditions. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, California, pp. 221–233.

Lakdawalla, D., Philipson, T., 2002. The growth of obesity and technological change: A theoretical and empirical analysis. Working Paper 8946, National Bureau of Economic Research.

Lakdawalla, D., Philipson, T., Bhattacharya, J., 2005. Welfare enhancing technological change and the growth of obesity. American Economic Review 95(2):253–257.

Lopez, R., 2004. Urban sprawl and risk for being overweight or obese. American Journal of Public Health 94:1574 –1579.

Mieszkowski, P., Mills, E.S., 1993. The causes of metropolitan suburbanization. Journal of Economic Perspectives 7(3):135—147.

Mokdad, A., Marks, J., Stroup, D., Gerberding, J. (2005). Correction: Actual Causes of Death in the United States, 2000. Journal of the American Medical Association, 293, 293-294

Murphy, K., Topel, R., 1985. Estimation and Inference in Two-Step Econometric Models. Journal of Business and Economic Statistics 4(3): 370-379.

Ogden, C.L., Carroll, M.D., Curtin, L.R., McDowell, M.A., Tabak, C.J., Flegal, K.M., 2006. Prevalence of overweight and obesity in the United States, 1999–2004. JAMA 295(13):1549–1555.

Olsho, L., Kim, M., 2005. Reexamining Trends in Obesity Prevalence in the United States: the Effects of Shifting Population Demographics," manuscript 2005.

Plantinga, A.J., Bernell, S., 2007. Can urban planning reduce obesity? The role of self-selection in explaining the link between weight and urban sprawl. Review of Agricultural Economics 29(3): 557-563.

Pisarski, A., 2001. US Roads in *Millennium Book*, International Road Federation.

Saelens B.E., Sallis J.F., Frank L.D., 2003. Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. Annals of Behavioral Medicine 25:80–91.

Transportation Research Board Special Report 282, 2005. Does the Built Environment Influence Physical Activity? Examining the Evidence. TRANSPORTATION RESEARCH BOARD, INSTITUTE OF MEDICINE OF THE NATIONAL ACADEMIES

U.S. Department of Health and Human Services. The Surgeon General's call to action to prevent and decrease overweight and obesity. [Rockville, MD]: U.S. Department of Health and Human

Services, Public Health Service, Office of the Surgeon General; [2001]. Available from: US GPO, Washington.

Wassmer R., Edwards, D., 2005. Causes of Urban Sprawl (Decentralization) in the United States: Natural Evolution, Flight from Blight, and the Fiscalization of Land Use, manuscript.

Table 1

Trends in Population Density for Large Constant-geography Metropolitan Areas in U.S., 1970-2000[a]

	1970	1980	1990	2000	Percent Change 1970-2000
Population weighted population density[b]	7025	5754	5640	5680	-19.15
Proportion of population living in dense areas with density >3500 people per square mile (%)	56.8116	50.2725	48.6137	47.8532	-15.77
Proportion of population living in dense areas with density >5000 people per square mile (%)	43.2718	35.8751	34.221	33.8134	-19.04
Proportion of population living in dense areas with density >9000 people per square mile (%)	20.0296	14.6985	13.6036	14.0508	-29.86
Proportion of population living in dense areas with density >12500 people per square mile (%)	12.1656	8.5012	8.1335	8.1261	-33.20

[a] The sample includes those 53 largest metropolitan areas identified in the public-use NHIS data based on 1993 metropolitan area definitions.

[b] Population weighted population density is calculated by $\sum_i^I \frac{Pop_i}{Pop_{total}} * \frac{Pop_i}{Land_i}$, where Pop_i refers to the census tract population, Pop_{total} refers to the total metropolitan area population of that census tract, $Land_i$ refers to the square mileage of the census tract.

Table 2
Estimates of the Effect of the 1947 Highway Plan on Population Density

	Density: (>5000 people per square mile)		Density: (>9000 people per square mile)		Density: (>12500 people per square mile)	
Number of Planned Highway Rays	-2.3894***	-2.2428***	-2.1418***	-2.0251***	-1.3813***	-1.2800***
	(0.3679)	(0.4002)	(0.3924)	(0.4332)	(0.3075)	(0.3264)
MSA Employment Rates		-0.0799		0.2241		0.4093
		(0.3324)		(0.3675)		(0.2660)
MSA Median Family Income		0.2322***		0.1635*		0.1259*
		(0.0822)		(0.0890)		(0.0664)
MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	488	488	488	488	488	488
R square	0.95	0.95	0.95	0.95	0.95	0.95
F-statistics for excluded variable	42.18	31.41	29.80	21.85	20.18	15.38
P-value for excluded variable F-test	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Mean of dependent variable in 1970	47.6094	47.6094	23.2524	23.2524	14.1957	14.1957
(standard deviation)	(17.8668)	(17.8668)	(17.6566)	(17.6566)	(15.6411)	(15.6411)

Notes: "Planned Highway Rays" is calculated by multiplying the number of rays in the 1947 plan with the linear time trend and then dividing the resulting number by 20. Standard errors are in parentheses. Standard errors are clustered by MSA/year. All specifications control for time trend and MSA fixed effects.
* indicates significant at 10% level, ** indicates significant at 5% level, *** indicated significant at 1% level.

Table 3
Two-step Estimates of the Effect of Predicted Population Density on Obesity

	Density (>5000 people per square mile)		Density (>9000 people per square mile)		Density (>12500 people per square mile)	
Predicted Population Density	-0.0010**	-0.0012**	-0.0013**	-0.0015**	-0.0020**	-0.0023**
	(0.0005)	(0.0006)	(0.0006)	(0.0007)	(0.0010)	(0.0012)
MSA Employment rates		0.00003		0.0004		0.0011
		(0.0005)		(0.0006)		(0.0008)
MSA Median Family Income		0.0004		0.0004		0.0005
		(0.0004)		(0.0004)		(0.0004)
Individual Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	703544	703544	703544	703544	703544	703544
R square	0.04	0.04	0.04	0.04	0.04	0.04
Mean of Predicted Population Density	49.1184	48.5728	27.4516	26.9914	19.3775	18.9587
(Standard Deviation)	(21.4718)	(20.8045)	(22.5587)	(22.1618)	(21.2823)	(21.0312)
Mean of Obesity	0.1096	0.1096	0.1096	0.1096	0.1096	0.1096
(Standard Deviation)	(0.3124)	(0.3124)	(0.3124)	(0.3124)	(0.3124)	(0.3124)

Notes: In all the specifications I control for individual's demographic information: dummy variables for age, race (non-Hispanic Black, Hispanic and others, non-Hispanic White is the reference group), sex, education (some high school, high school graduate, some college and college graduate, elementary school is the reference group) and marital status (married, separated/divorced, or widowed, single is the reference group). Table 2 shows the model used to create the predicted population density. Standard errors are corrected using Murphy-Topel estimate of variance-covariance. Standard errors are clustered at MSA/year. The regressions are weighted with NHIS sampling weights.
* indicates significant at 10% level, ** indicates significant at 5% level, *** indicated significant at 1% level.

Table 4

Two-step Estimates of the Effect of Population Density on BMI

	Density (>5000 people per square mile)		Density (>9000 people per square mile)		Density (>12500 people per square mile)	
Predicted Population Density	-0.0087	-0.0114	-0.0095	-0.0128	-0.0147	-0.0203
	(0.0077)	(0.0085)	(0.0092)	(0.0101)	(0.0145)	(0.0154)
MSA Employment rates		0.0009		0.0047		0.0101
		(0.0054)		(0.0065)		(0.0094)
MSA Median Family Income		0.0105**		0.0101**		0.0105**
		(0.0051)		(0.0051)		(0.0053)
Individual Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	703544	703544	703544	703544	703544	703544
R square	0.13	0.13	0.13	0.13	0.13	0.13
Mean of Predicted Population Density	49.1184	48.5728	27.4516	26.9914	19.3775	18.9587
(Standard Deviation)	(21.4718)	(20.8045)	(22.5587)	(22.1618)	(21.2823)	(21.0312)
Mean of BMI	24.6591	24.6591	24.9529	24.9529	24.9529	24.9529
(Standard Deviation)	(4.5574)	(4.5574)	(4.7303)	(4.7303)	(4.7303)	(4.7303)

Notes: In all the specifications I control for individual's demographic information: dummy variables for age, race (non-Hispanic Black, Hispanic and others, non-Hispanic White is the reference group), sex, education (some high school, high school graduate, some college and college graduate, elementary school is the reference group) and marital status (married, separated/divorced, or widowed, single is the reference group). Table 2 shows the model used to create the predicted population density. Standard errors are corrected using Murphy-Topel estimate of variance-covariance. Standard errors are clustered at MSA/year. The regressions are weighted with NHIS sampling weights.

* indicates significant at 10% level, ** indicates significant at 5% level, *** indicated significant at 1% level.

Table 5
Estimates of the Effect of 1947 Highway Plan on Population Density
"Matched" Sample

	Density: (>5000 people per square mile)		Density: (>9000 people per square mile)		Density: (>12500 people per square mile)	
Mean of Dependent Variables	49.4003		27.5347		19.4068	
(Standard Deviation)	(22.2776)		(23.0227)		(21.4060)	
	IV	Two-Step	IV	Two-step	IV	Two-step
Planned Highway Rays	-2.0713***	-2.0939***	-2.4649***	-2.4885***	-2.0212***	-2.0416***
	(0.6000)	(0.6060)	(0.5905)	(0.5984)	(0.4653)	(0.4718)
Individual Demographic controls	Yes	No	Yes	No	Yes	No
MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Sample Size	80580	80580	80580	80580	80580	80580
R square	0.98	0.98	0.98	0.98	0.99	0.99
F-statistics for excluded variable	11.91	11.94	17.43	17.30	18.87	18.73

Note: "Planned Highway Rays" is calculated by multiplying the number of rays in the 1947 plan by the linear time trend and then dividing the resulting number by 20. Standard errors are in parentheses. Standard errors are clustered by MSA/year. All specifications control for time trend and restricted MSA fixed effects. Individual's demographic controls in first stage of instrumental variable approach include: dummies for age, race (non-Hispanic Black, Hispanic and others, non-Hispanic White is the reference group), sex, education (some high school, high school graduate, some college and college graduate, elementary school is the reference group) and marital status (married, separated/divorced, or widowed, single is the reference group).
* indicates significant at 10% level, ** indicates significant at 5% level, *** indicated significant at 1% level.

Table 6
Estimated Effects of Population Density on Obesity
"Matched" Sample

Density Variables and Estimation Method	Coefficient (standard error)	R-square	Mean of Density Variable (standard Deviation)	Mean of Predicated Density Variable (Standard Deviation)
Density (>5000 people per square mile)				
OLS	-0.0001 (0.0006)	0.05	49.4003 (22.2776)	
IV	-0.0018* (0.0011)	0.05		
Two-Step	-0.0018* (0.0010)	0.05		49.4323 (22.1229)
Density (>9000 people per square mile)				
OLS	-0.0002 (0.0004)	0.05	27.5347 (23.0227)	
IV	-0.0015* (0.0009)	0.05		
Two-Step	-0.0015* (0.0008)	0.05		27.5787 (22.8785)
Density (>12500 people per square mile)				
OLS	-0.0002 (0.0004)	0.05	19.4068 (21.4060)	
IV	-0.0019* (0.0011)	0.05		
Two-Step	-0.0019* (0.0010)	0.05		19.4448 (21.3011)

Note: In IV and OLS estimates I control for individual's demographic information as table 3. MSA fixed effects and year fixed effects are controlled in all specifications. Table 5 shows the model used to create the predicted population density using the conventional instrumental variable approach and the two-step procedure. The standard errors for the two-step estimation are corrected using Murphy-Topel estimate of variance-covariance. The regressions are weighted with NHIS sampling weights.
* indicates significant at 10% level, ** indicates significant at 5% level, *** indicated significant at 1% level.

Table 7
Estimated Effects of Population Density on BMI
"Matched" Sample

Density Variables and Estimation Method	Coefficient (standard error)	R-square	Mean of Density Variable (standard Deviation)	Mean of Predicated Density Variable (Standard Deviation)
Density (>5000 people per square mile)				
OLS	0.0047 (0.0084)	0.13	49.4003 (22.2776)	
IV	-0.0123 (0.0135)	0.13		
Two-Step	-0.0122 (0.0128)	0.13		49.4323 (22.1229)
Density (>9000 people per square mile)				
OLS	0.0056 (0.0059)	0.13	27.5347 (23.0227)	
IV	-0.0103 (0.0113)	0.13		
Two-Step	-0.0102 (0.0106)	0.13		27.5787 (22.8785)
Density (>12500 people per square mile)				
OLS	0.0079 (0.0059)	0.13	19.4068 (21.4060)	
IV	-0.0126 (0.0140)	0.13		
Two-Step	-0.0125 (0.0129)	0.13		19.4448 (21.3011)

Note: In IV and OLS estimates, I control for individual's demographic information as table 4. MSA fixed effects and year fixed effects are controlled in all specifications. Table 5 shows the model used to create the predicted population density using the conventional instrumental variable approach and the two-step procedure. The standard errors for the two-step estimation are corrected using Murphy-Topel estimate of variance-covariance. The regressions are weighted with NHIS sampling weights.
* indicates significant at 10% level, ** indicates significant at 5% level, *** indicated significant at 1% level.

Figure I Overview of the Two-Step Estimation Model

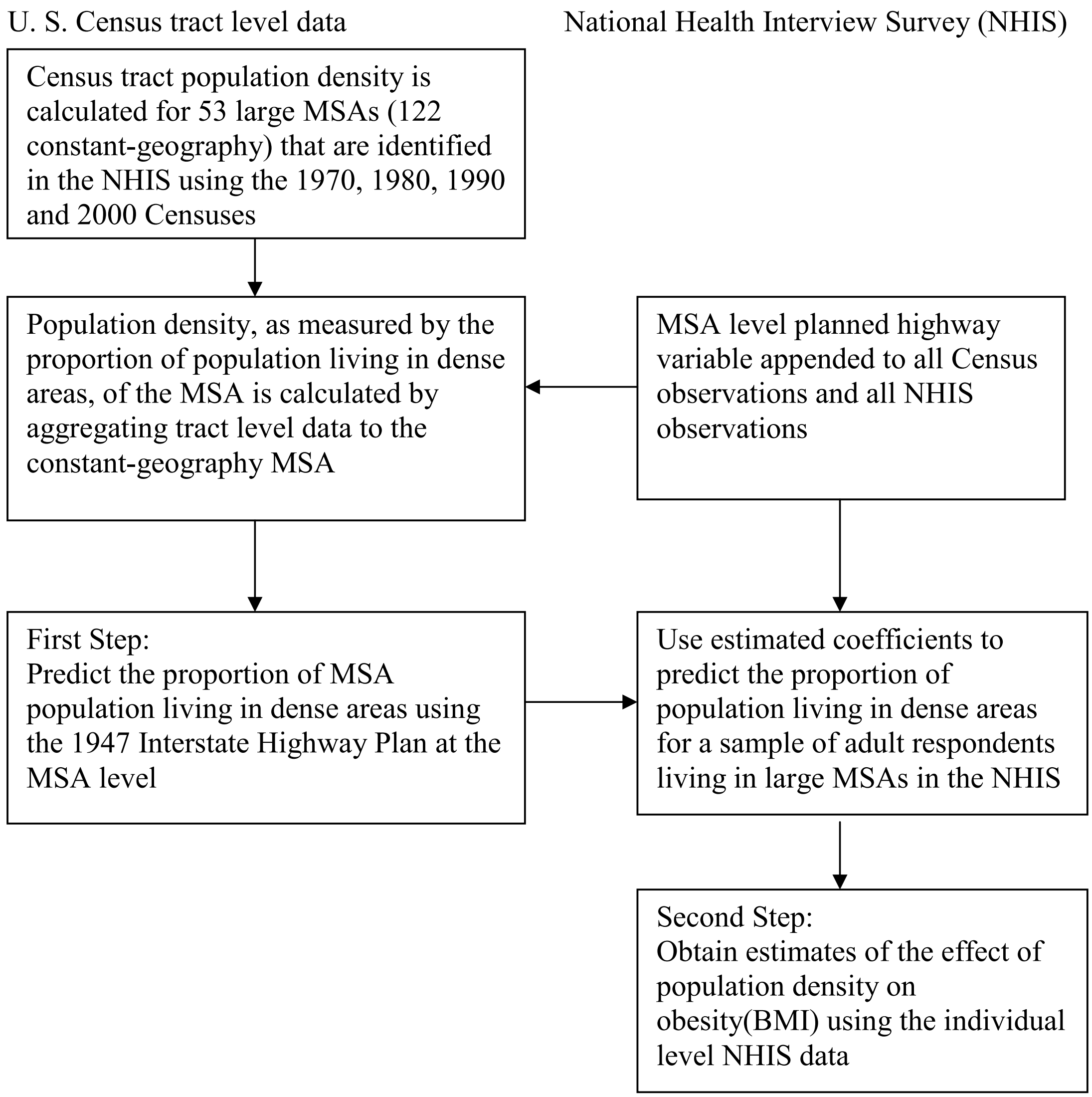


Figure II The Planned National System of Interstate Highways

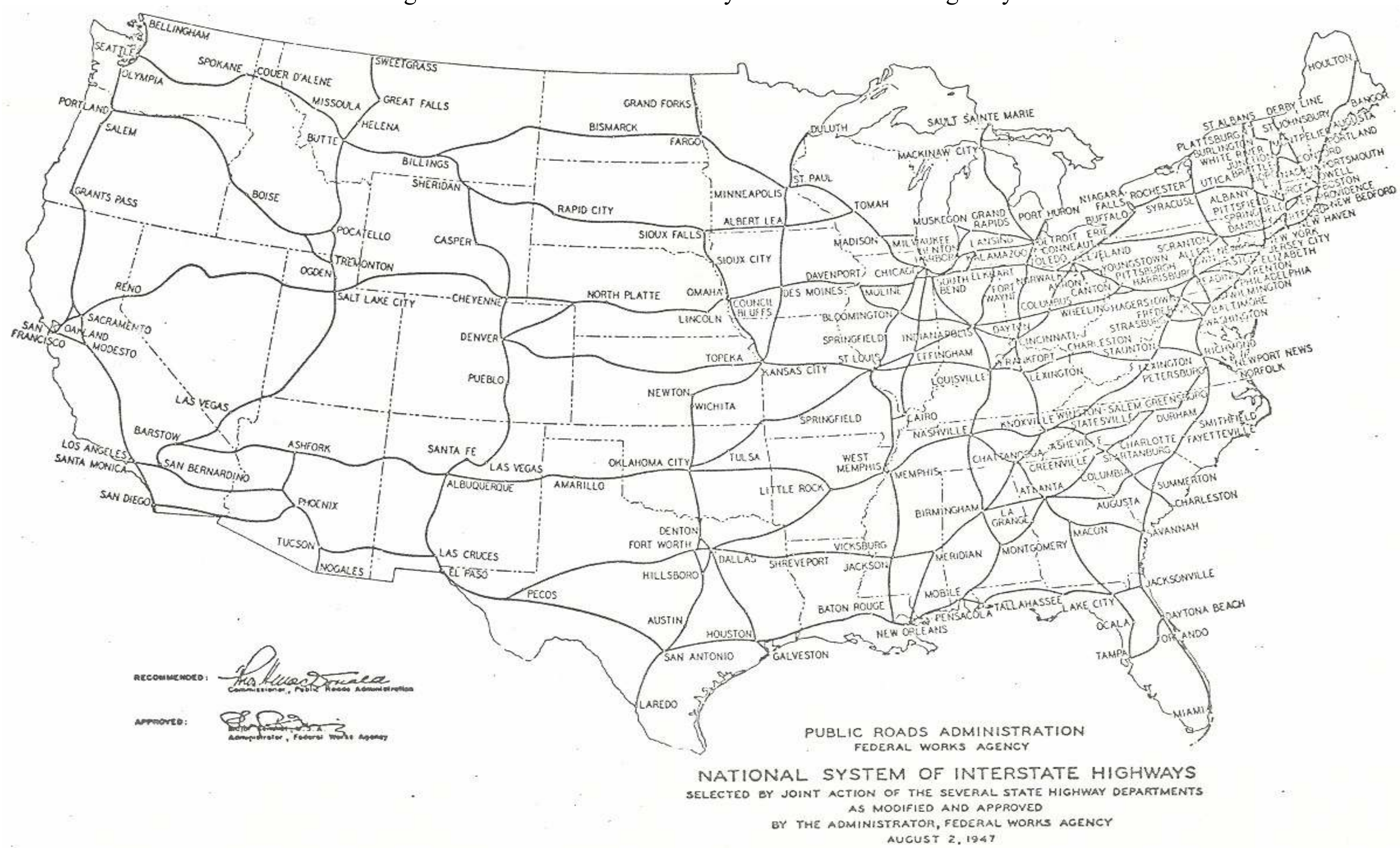


Figure 1 Predicted Proportion of Population living in Dense Areas (>5000 people per sq mile)

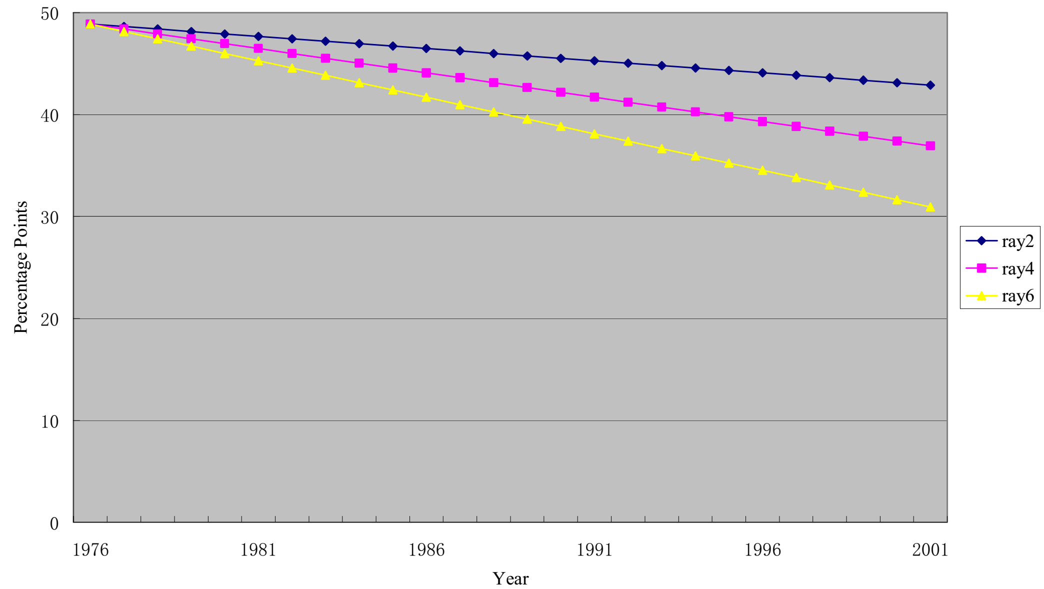


Figure 2 Predicted Proportion of Population Living in Dense Areas (>9000 people per sq mile)

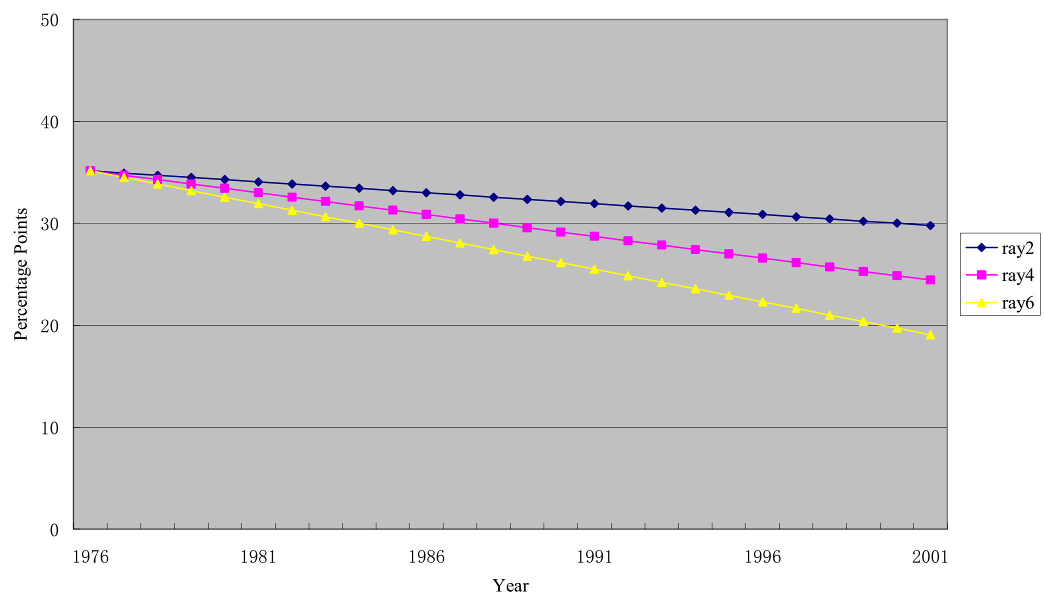
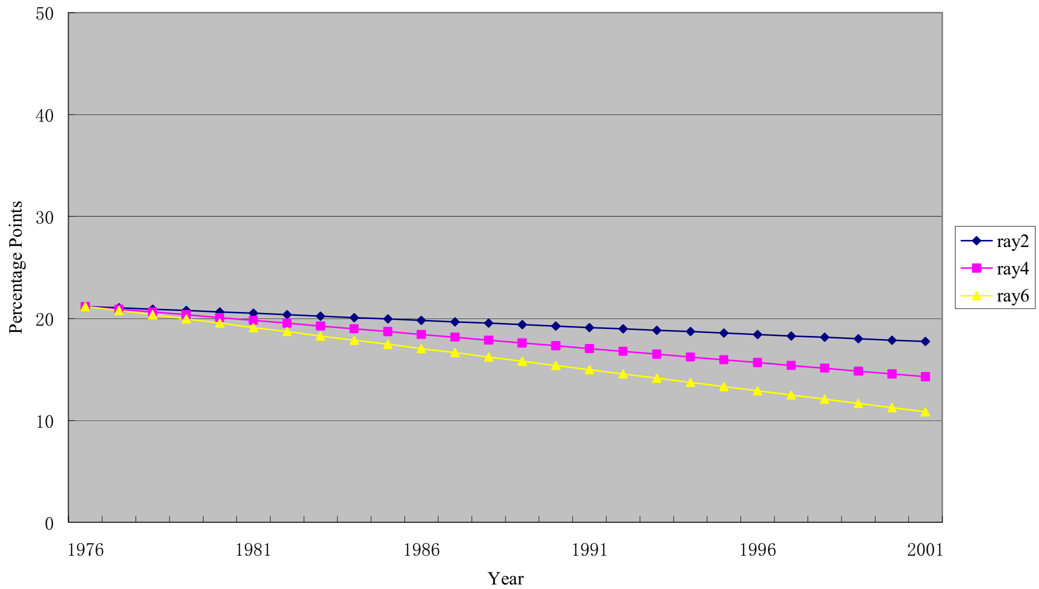


Figure 3 Predicted Proportion of Population Living in Dense Areas (>12500 prople per sq mile)



Appendix Table 1
Summary Statistics

	Full Sample		Matched Sample	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
BMI	24.6591	4.5574	24.6121	4.5565
Obese	0.1096	0.3124	0.1075	0.3098
Male	0.4723	0.4992	0.4711	0.4992
Age	43.0323	17.4858	43.0406	17.6489
Hispanic	0.0951	0.2933	0.0955	0.2939
Black	0.1323	0.3388	0.1327	0.3393
Other race	0.0327	0.1778	0.0312	0.1737
Some high school	0.1256	0.3313	0.1278	0.3339
High school graduate	0.3540	0.4782	0.3522	0.4777
Some college	0.1988	0.3991	0.1935	0.3951
College graduate	0.2183	0.4131	0.2196	0.4140
Married	0.6131	0.4870	0.5948	0.4909
Divsep	0.0995	0.2993	0.1057	0.3074
Window	0.0717	0.2579	0.0738	0.2615
Planned highway rays	4.1562	1.6903	4.1890	1.7003
MSA employment rates	93.5200	2.2604	93.7634	1.6373
MSA family income (in thousands 1982-1984 real dollars)	25.9106	3.4824	31.2155	4.2225
Observations	703544		80580	

Note: Mean and standard deviations are shown. NHIS sample weights are used in calculating the mean and standard deviation. See text for additional information about samples.

Appendix Table 2 Summary Statistics on Population Density Measures (Mutually Exclusive Categories)

	1970	1980	1990	2000	percent change 1970-2000
Proportion of population living in areas with density<3500 people per square mile	39.46776	45.20924	46.59261	47.50691	20.37
Proportion of population living in areas with density >3500 and <5000 people per square mile	12.92282	14.35135	14.33773	13.94781	7.93
Proportion of population living in areas with density >5000 and <9000 people per square mile	24.35704	23.01852	22.6554	21.60482	-11.30
Proportion of population living in dense areas with density >9000 and <12500 people per square mile	9.056692	7.203126	6.513102	7.003191	-22.67
Proportion of population living in dense areas with density >12500 people per square mile	14.19569	10.21775	9.901166	9.937274	-30.00
Number of Observations	122	122	122	122	