# Differences in Obesity for Suburban and Central-City Environments[1]

Gregor A. Franz[2]

**November 12, 2007**

**Abstract**

This paper seeks to test the hypothesis that suburban living environments lead to a higher body weight for individuals compared to central city environments. I use the geographic information from the confidential NLSY79 files in conjunction with census tract population density data for the years 1981 to 2004. The effect of population density on the body weight of individuals who move is estimated conditioning on individual and area level characteristics. The paper concludes that the actual effect of population density on the weight of individuals is not as large as the public health literature would want people to believe, but it is neither as insignificant as the economic literature to date estimated.

[1] I am indebted to Jan Brueckner and Kitt Carpenter for guiding me in the development of this research project. I would also like to thank Francesca Mazzolari, David Neumark, and Marianne Bitler for their very helpful comments. All remaining errors are mine.

[2] Department of Economics, 3151 Social Science Plaza, University of California, Irvine; Irvine, CA 92697-5100. e-mail: gfranz@uci.edu

## 1. Introduction

The Centers for Disease Control (CDC) estimates that more than 30 percent of Americans are obese. But despite years of public awareness campaigns, people in the United States continue to gain weight. In search for the causes of this rise in obesity, the public health literature continues to extensively examine the links between a host of socioeconomic factors and weight gain. In recent years, this literature discovered that obese individuals appear to be more likely to live in less dense areas such as suburbs rather than in central city districts. Some articles in this literature conclude that living in suburban environments leads to weight gain.

Prima facie this claim seems convincing. In the stereotypical suburb, individuals primarily rely on automobiles for transportation, reducing the amount of physical activity. But is this relationship between suburban living and weight gain a causal one? Most of the health policy literature seems to argue this way. But apart from suburban living causing weight gain or loss, it is also possible that people who will gain weight sort into lower density neighborhoods. To test this hypothesis, I use over 20 years of panel data to examine the relationship between population density and body weight.

The longitudinal structure of my data allows me to follow individuals over time and control for unobserved characteristics that may bias cross-sectional estimates of the effect of population density on weight. I show that cross-sectional estimates are biased upwards, but there still remains a significant effect of population density on body weight. This difference is attributable to unobserved individual characteristics that cause individuals with a predisposition of gaining weight to also prefer lower density living environments.

Relative to the previous literature, this paper offers two significant improvements. First, I use a straightforward and disaggregated measure of sprawl, namely, population density at the census tract level, to examine the effect of suburban living on individual body weight. Second, no previous study has examined the effect of population density on body weight using such detailed and nationally representative data over such a long period of time. This allows me to rule out concerns about the amount of individual variation in body weight over time and gives me a very large data set when pooled for baseline analyses. While previous papers, that control for sorting of individuals found no effect of sprawl on body weight, I show that a significant portion of the weight gain remains, even after controlling for the spatial sorting of individuals.

If low population density in urban areas does in fact cause people to gain weight, without them being aware of this risk, this represents a possible market failure. The easily interpretable structure of my population density measure makes it possible to discuss if policy options designed to deal with such a market failure are desirable and if so, which policy instruments might yield the best results.

The following section summarizes the literature to date that examines the relationship between various measures of sprawl and body weight. In section three I describe the data used in this paper and outline the estimation methods applied in section four. I present results in section five, discuss policy implications in section six and conclude in section seven.

## 2. Literature review

The health policy and urban planning literature that examines the effects of urban sprawl on body weight can be roughly separated into two categories. The first group of studies uses nationally representative cross-sectional data, but employ a geographically dispersed measure of sprawl. In this category, Ewing et al. (2003) were among the first to note a relationship between urban sprawl and weight gain. They construct a unique sprawl index for U.S. counties and Metropolitan Statistical Areas (MSAs) based on residential density, land use mix, degree of centering, and street accessibility. Using cross-sectional data for the years 1998 and 2000, they conclude that a higher sprawl index is associated with higher BMI and obesity rates. Other studies that use nationally representative data subsequently develop various sprawl indices to measure their impact on the body weight of individuals (e.g. Lopez 2004, Nelson et al. 2006). These studies have two things in common. All of them use fairly imprecise geographic indices of sprawl at the county or MSA level and none of them think that the sorting of individuals would significantly bias their results. Ewing et al. (2006), even in light of results that seem to contradict this argument, think that sorting is a non-issue.

A second group of studies in the health policy and planning literature looks at a restricted set of geographic regions employing a finer measure of sprawl. While their sprawl measures allow for a better segmentation of geographic regions, they also fail to consider the possibility that individuals sort into certain neighborhoods based on unobserved individual characteristics. Additionally, it might be hard to extrapolate the findings from very specific neighborhoods to a national scale (see for example Saelens et al. 2003, Frank et al. 2004, Glass et al. 2006).

With spatial sorting, a problem ignored by all the above studies, heavier individuals might simply be more inclined to live in the suburbs for various reasons and thus may sort themselves into less dense living environments. For example, such individuals might have a distaste for walking and prefer larger amounts of living space. If these taste preferences are unobservable, simply controlling for observable characteristics will result in upward biased estimates of the effects of suburban living on body weight.

To date, two studies recognize this self-selection problem and try to solve it with different approaches. Plantinga and Bernell (2007) develop a two equation model in which they simultaneously determine built-environment and weight. They use the sprawl index developed by Ewing et al. (2003) and find only weak evidence for sprawl causing higher weight, but find greater evidence of higher weight leading individuals to select residential locations with a high index value.

In another attempt to control for the effects of sorting, Eid et al. (2007) use a continuous sprawl index developed from satellite imaging and restricted longitudinal data from the NLSY for the years 1988-1990 and 1992-1994. This approach allows them to follow individuals who move over the time frame of their analysis and to control for their individual characteristics. Using a first differencing approach to control for any unobserved individual characteristics that might be associated with the decision to move to a specific neighborhood, they find no evidence that living in an area with more sprawl leads to a higher body weight.

While the approach of Eid et al. (2007) mostly resolves the concern of imprecise locational measurement (e.g. at the county level), the time frame of their analysis is fairly short. As individual weight usually only changes gradually, their data might not contain

enough individual weight variation over time to estimate reliable results. Their measure for sprawl and mixed land use is time-invariant, with sample variation generated by individuals moving to different areas. As the time frame of their analysis is fairly short, this does not pose a serious problem for them. However, if a significantly longer time frame, such as in this paper, is considered, a time invariant sprawl index would fail to capture these trends, since city and suburban environments are subject to change over time and space.

As seen in the preceding discussion, many of the available studies that examine the relationship between suburban living environments and body weight construct specific sprawl indices. While the use of sprawl indices has certain advantages, it is unclear how results based on such indices can be used for a broader analysis other than just a comparison of regions within each study. It is difficult to make policy recommendations based on these indices, making their use and interpretation somewhat limited in scope.[3]

In this paper I use a straightforward measure, population density at the census tract level, to estimate the effect of suburban living on body weight. The detail and structure of my data allows me to distinguish movers from non-movers and follow them over time. Thus, I can observe the population density in the tract of residence in each year, while at the same time controlling for unobservable individual and region specific effects. Even though previous research considered the possibility of spatial sorting, the time frame of more than 20 years of individual level data from the National Longitudinal Survey of Youth 79 (NLSY79) allows me to significantly improve on this research.

[3] The indices usually incorporate various variables, such as population density, land use mix, and street accessibility.

Using this method, I can estimate the portion of weight gain that remains, even after controlling for the spatial sorting of individuals, while at the same time relying on an easily interpretable sprawl measure that can be applied straightforwardly for policy analysis.

## 3. Data

The main data come from the confidential National Longitudinal Survey of Youth 79 (NLSY79) zip code and census tract files. The NLSY79 is currently conducted on a biennial basis by the U.S. Department of Labor. During the first wave of interviews in 1979 the NLSY79 comprised a nationally representative sample of 12,686 men and women who ranged from 14 to 22 years of age. I use data from the NLSY79 for the years 1981 to 2004 as weight and height was not reported during the initial two waves of the survey. The confidential data files of the NLSY79 also report zip code and census tract identifiers not available in the public use or restricted Geocode version of the data. The NLSY79 is merged with population and area data at the tract level from the census Bureau for the years 1980 to 2004.

### 3a. Representativeness of sample

To make the survey nationally representative, the NLSY79 oversamples minorities and blacks. It is known that blacks have, on average, a higher weight compared to other racial groups. Additionally, the NLSY79 follows a specific age cohort and is therefore not representative for all individuals in the U.S. Many of the previous studies used the nationally representative Centers for Disease Control and Prevention's (CDC) *Behavioral*

*Risk Factor Surveillance System* (BRFSS) to assess the effect of sprawl on body weight. The BRFSS is available for the years 1984 to 2004. The survey is designed to be representative by state and is stratified by age, sex, and race (CDC 2003). To evaluate how the age cohorts of the NLSY79 compare against nationally representative data for all ages, I graph the BMI of the NLSY79 cohorts against the BRFSS data in figures 1 and 2.

Figure 1 shows how the weight of respondents behaves over the time frame of this analysis and how it compares to other data for same age cohorts in each year. I match the BMI of individuals in the NLSY79 and the corresponding age cohorts in the BRFSS against each other for males and females. The time trends of rising BMI appear to be approximately the same in both data sets. But the level difference in BMI within each year between the two data sets is especially large for females. In 2004 this difference means that the average female in the BRFSS is considered just overweight with a BMI of 26.5 (a BMI>25 is considered overweight) while the average female in the NLSY79 is at the threshold of becoming obese with a BMI slightly over 28 (a BMI>30 is considered obese). A possible explanation is the fact that the NLSY79 over samples African-Americans, who are known to have a higher BMI on average. To account for this oversampling, sample weights are employed in all estimations.

It is well documented that BMI increases with age, and the NLSY79 follows a specific age cohort over time. In 1979, when the survey was first given, individuals were between the ages of 14 and 21. In the last year of available data, in 2004, individuals were between the ages of 39 and 46. Figure 2 shows how the age cohorts of the NLSY79 compare to the average population using the BRFSS. We see that BMI increases much more rapidly for the NLSY79 age cohorts compared to the average population. While

exhibiting a much lower BMI in the beginning, NLSY79 men overtake the average U.S. male in 1995 when they range from 30 to 38 years of age. For women this overtaking does not occur until 2003, when the NLSY79 women are 38 to 46 years of age. While individuals of the same age groups in those two data sets seem to differ somewhat in the value of their BMI, the overall trend seems to be preserved, so that using the NLSY79 to analyze changes in body weight seems appropriate.

**3b. Sample restrictions and covariates**

As the definitions of metropolitan areas changes over the time period of my analysis I need to construct geographically consistent city designations. I construct time invariant city identifiers from the counties reported in the NLSY79 for all available years. These identifiers correspond to the designations of Metropolitan Statistical Area (MSA) in the 1994 County and City Data Book. I then calculate the population density at the tract level for every year from 1980 to 2004 using the census data and merge it with the NLSY79 tracts. Population density is measured in hundreds of thousands of people per square mile.

Table 1 reports the exclusion restrictions imposed on the full sample used in this analysis and the resulting sample size. The sample represents the pooled yearly survey data of the NLSY79 and the merged census tract information. Missing observations account for almost half of the full sample. This is due to the fact that weight is not recorded in all the available years and many individuals have missing data on their family income. Together, those two exclusions make up for almost 65,000 of the 79,588 missing observations.

Following missing observations, non-movers, with 18,827 observations, represent the largest and most important exclusion restriction. By focusing only on movers, for whom tract population usually changes as a result of their move, I can identify the effect of population density on body weight over time. I define a mover to be an individual who moves between counties in any given year of my sample period. The county level was chosen as it is sufficiently small to allow a movement from inner city area to suburban areas in most MSAs, but large enough so that I do not have to worry about individuals moving across the street into a different census tract. With such a move, everything for the individual essentially stays the same, but the tract change may yield a different population density. So everyone in the sample moves across counties, but about 80 percent of the sample stays within the same MSA. This means that county fixed effects might produce uninformative results, as county and population density would change in step in every individual case. As a result, MSA fixed effects appear to be appropriate in this context. While use of MSA fixed effects means that single-county MSAs will be excluded from the analysis, the sample contains enough observations from large multiple-county MSAs to make this an informative study.

Other exclusion restrictions eliminate individuals with farm income as well as active military personnel, with 657 and 367 observations respectively. As their type of work involves above average physical activity, it is unclear if population density changes would correctly capture weight changes in those individuals. As I want to focus on residents of metropolitan areas, individuals in rural locations are excluded (17,044 observations), as are individuals who do not live in an MSA or for whom I cannot identify the MSA of residence (4,726 observations). The MSA of an individual may not

be identifiable if the state or county of residence is not reported in a given year. My final sample size is 40,031 observations for the years 1981 to 2004. There are 5,562 unique respondents in this sample.[4]

Other variables included in the analysis are age, age squared, sex, marital status, race, family size, number of children, annual family income, educational attainment, work experience, and the individuals' fathers' and mothers' educational attainment. I assume that the height of individuals did not change since 1985, when it was last collected and respondents were aged 20 to 28. Table 2 reports the means and standard deviations of the variables for the sample used in this analysis. The average BMI of individuals seems somewhat low due to the fact that the NLSY79 age cohorts did not go through a whole life cycle and because sample sizes are larger for years in which individuals are younger. The same is true for the percentage of obese individuals in the sample.

The average population density in the census tracts of my sample is 6,900 people per square mile. Average density has a fairly large standard deviation of 14,900 people per square mile. While average density seems large, it should be noted that all census tracts in my sample are within MSAs and defined as "urban" by the respondent.

## 4. Methods

Three different dependent variables are used to estimate the effect of population density on weight gain. First, I use body weight controlling for height. The second dependent variable is an individuals' BMI, calculated as weight in kilograms over height in meters

[4] Although not reported, 174 pregnant individuals for the years 1981 to 2004 are excluded as well.

squared. As the final measure I use a dichotomous variable, *obese*, which indicates whether or not an individual has a BMI greater than 30.

For comparison to previous results and to show the difference due to sorting of individuals, I first estimate a set of linear regressions on the pooled cross sections of my data. Let $Y_{ijst}^{k}$ denote body weight, BMI, or an indicator whether or not the person is obese, for individual i in census tract j in region s at time t, and k being the respective dependent variable. The term $D_{ijrt}$ denotes the population density for individual i in census tract j in region s at time t, and $\mathbf{X}_{it}$ is a vector of individual characteristics for each person i at time t. $S_s$ represent *state* or *MSA* fixed effects, $T_t$ are year effects, and $\varepsilon_{ijrt}$ is the random disturbance term. Then, the model can be expressed as follows:

$$Y_{ijst}^{k} = \gamma D_{ijst} + \beta \boldsymbol{X}_{it} + S_s + T_t + \varepsilon_{ijst} \qquad (1)$$

If we think that there might be unobserved effects present at the individual level, the error term would be written

$$\varepsilon_{ijst} \equiv c_i + u_{ijst},$$

where $c_i$ denotes the unobserved effects and $u_{ijst}$ the idiosyncratic error. For this model to produce consistent estimates there should be no correlation between the right hand side variables of equation (1) and the unobserved effects $c_i$. Even if this assumption holds, we still need to estimate robust standard errors, given that $\varepsilon_{ijst}$ will be serially correlated, as the unobserved effects are present in each time period.

To test whether equation (1) introduces bias in estimates of γ, I proceed to estimate a second set of regressions using individual fixed effects. Using the above outlined notation and letting $c_i$ denote the individual effects, the estimating equation can be expressed in the following form:

$$Y_{ijst}^{k} = \gamma D_{ijst} + \beta \boldsymbol{X}_{it} + S_s + T_t + c_i + u_{ijst}, \qquad (2)$$

Any change in γ between equation (1) and (2) should effectively capture the weight change due to sorting of individuals into areas of different densities according to their unobserved individual characteristics. Such sorting, if it occurs, generates a correlation between density and the individual $c_i$ component of $\varepsilon_{ijst}$. The resulting bias in the estimate of γ is eliminated when individual fixed effects are employed.

When estimating a fixed effects model it is always a question whether there is enough within group variation in both the dependent variable and the independent variable of interest. As every individual in my sample moves between counties, this should introduce a sufficient amount of variation within individual tract measures of density. On average, body weight does not change drastically over 1 or 2 years, but the long time frame of my analysis should yield enough individual variation to satisfy this condition.

One concern is the possible endogeneity of moving decisions. If there are time-varying unobservable characteristics that cause an individual to move and if these characteristics are correlated with a person's weight, then my results will be biased. If, for example, certain health conditions require individuals to move to high density area because they are unable to use an automobile and at the same time cause individuals to lose weight, the estimated results will be biased downward. Any such bias, however, should be small compared to that produced by failing to control for time-invariant characteristics.

## 5. Results

I first estimate a set of pooled OLS models, as described in equation (1), to establish comparable results to previous cross-sectional studies. For the second set of estimates, I include individual fixed effects, as described in equation (2), to show the amount of bias in the OLS estimates that is due to sorting of individuals. Both models are estimated for the three dependent variables, *weight*, *BMI*, and *obese*, and using both *state* and *MSA* effects to show the effect of regional controls. Year dummies are included in all specifications. All models are estimated using robust standard errors and a sample weight supplied by the BLS to account for oversampling of certain population groups. Population density is measured as hundreds of thousands of people per square mile. It has a standard deviation of 0.149, meaning 0.15 times the population density is approximately a one standard deviation difference.

### 5a. Baseline results

Table 3 reports the results of estimating equation (1) using the dependent variable *Weight*, controlling for the height of each individual, *BMI*, and the dichotomous variable *Obese*. To have an additional regional comparison, I report the results of using state fixed effects in addition to MSA fixed effects. We might expect the results of the state effects model to be biased as it cannot capture the differing unobserved MSA regional effects. All estimations use the sample weight supplied by the BLS and report robust standard errors.

The first two columns of table 3 report the results for the dependent variable *Weight*, using state and MSA fixed effects. The coefficient on population density at the tract level is statistically significant and exhibits the expected negative sign in both

regional specifications. It ranges between -6.624 pounds per 100,000 additional people per square mile in the equation with state effects to -5.524 with MSA effects. Most other covariates exhibit the expected association. Other things being equal, blacks and hispanics weigh more than whites. The presence of more children in the household is associated with higher weight, while income and education are negatively associated with body weight. Higher parental education is also negatively correlated with weight. The coefficient for individuals that are divorced or widowed is negative, statistically significant, and fairly large. An interesting finding is that more work experience is associated with weight gain.

The next two columns of table 3 report the results of the OLS estimation using the dependent variable *BMI,* using state and MSA fixed effects. Again, all coefficients on population density show a negative and significant effect. With state effects, BMI decreases by -0.939 points per 100,000 additional people per square mile, while the drop is -0.812 BMI points using MSA effects. The control variables show similar signs and magnitudes as in the previous regressions, but the explanatory power of the model decreases from an R-squared of around 0.45 to around 0.20.

In the last two columns of table 3, I show the results of estimating a linear probability model that explains the probability of being obese[5] with state and MSA fixed effects. Population density has a negative and significant impact on the probability of being obese for both regional specifications. For the state effects model, an additional 100,000 people per square mile results in a 5 percent decrease in the probability of being

[5] With a linear probability model, it is possible that observations fall outside the region of support, taking on values greater than one or smaller than 0. To test the validity of this model, I also run probit models that verify the magnitude and direction of the reported coefficients. I do not report the results of those probit models but rather the coefficients of the linear probability model for ease of comparison.

obese. At the MSA level, the coefficient is -0.042, indicating about a 0.6 percent decrease in the probability of being obese for a one standard deviation increase in population density. Fewer of the other coefficients in the obesity regression on obesity are statistically significant compared to the previous two sets of estimations. An interesting result is that the mother's educational background is not statistically significant while the coefficient on the father's educational background is negative and significant at the 1 percent level.

**5b. Adding individual fixed effects**

As explained above, there is reason to believe that correlation between unobserved individual characteristics and the living environment is biasing the estimated effects of population density on weight outcomes. If individuals with unobservable characteristics that favor residence in low density tracts are at the same time more likely to gain weight, then the estimated coefficient of population density in model (1) will be biased upwards. The next sets of regressions estimate model (2) to see if such bias is present and to gauge its magnitude. In all specifications with individual fixed effects, time invariant covariates must be excluded. The affected covariates are height, race, and the educational attainment of each individual's mother and father.

The first two columns of table 4 show the results of the individual fixed effects regressions using body weight, again including state and MSA effects. The coefficients on population density are again negative and statistically significant, but much smaller than the OLS estimates. The effect of population density is reduced by 60 to 70 percent, indicating severe upward bias in the OLS estimates. Instead of the OLS weight loss of

0.75 to 1 pounds from a standard deviation increase in population density, the fixed effects estimates imply a weight loss of only about 0.3 pounds. Coefficients for other covariates also change somewhat. Most notably, there is a switch in significance and size in the coefficients of family size and the number of children in the household. In addition, the coefficient on work experience becomes negative and insignificant. A Hausman specification test confirms that the OLS model is indeed misspecified and that the fixed effects model is the appropriate one to use.[6]

The 3rd and 4th column of table 4 report the results of the individual fixed effects regression using BMI. Here also, the coefficients on population density are negative and statistically significant, but much smaller than the respective OLS estimates. The coefficients on population density are between 55 to 65 percent smaller than their OLS counterparts. In the original regressions, a standard deviation increase in density would reduce BMI between 0.10 and 0.15 points. In the fixed effects estimates, it will reduce BMI by only about 0.05 points. Use of individual fixed effects in the BMI regressions also results in most of the other covariates being statistically significant. Additionally, the explanatory power of the model doubles, and the R-squared of around 0.40 now matches that of the weight regressions. A Hausman specification test confirms that the OLS model is misspecified and the fixed effects model is the appropriate one to use.[7]

The last two columns of table 4 report the results of the individual fixed effects regression explaining the probability of being obese. The coefficients on population density are negative and statistically significant. Unlike in previous regressions, the

[6] The null hypothesis that the difference in coefficients is systematic is rejected with a Chi-squared value of 12,072 and 1,366 for the state and MSA effects model with 70 and 317 degrees of freedom respectively.
[7] The null hypothesis that the difference in coefficients is systematic is rejected with a Chi-squared value of 3,721 and 14,613 for the state and MSA effects model with 70 and 317 degrees of freedom respectively.

estimated effects of population density are not much smaller than the respective OLS estimates. They are only about 10 percent smaller than their OLS counterparts, meaning that an increase of one standard deviation in population density will still reduce the probability of being obese by about 0.6 percent. On the other hand, most of the other covariates are not statistically significant and the R-Squared is still fairly low at 0.10. Even though the difference is fairly small, a Hausman specification tests still shows the OLS model to be misspecified compared to the fixed effects model.[8]

**5c. Sensitivity Analysis**

This paper uses a time frame and estimation specification that has not been used in this context before. To identify the reason for the difference between my results and those of Eid et al. (2007), I adopt their time frame (a shorter period running from 1988 to 1994), while retaining my specification. If my results are just driven by my model specification, I should get similar results as in my original estimates when using different time periods. If my results are qualitatively similar to those of Eid et al. (2007), when using my model specification but their time frame, then the shorter time frame is the likely cause of the difference in the results.

Table 5 reports the population density coefficients and robust standard errors for OLS and individual fixed effects regressions, restricting the time period to the years 1988 to 1994. As in the original regressions, the OLS coefficients are all statistically significant at the 1 percent level. The magnitude of all population density coefficients is

[8] The null hypothesis that the difference in coefficients is systematic is rejected with a Chi-squared value of 182 and 692 for the state and MSA effects model with 70 and 317 degrees of freedom respectively.

somewhat larger than in the original estimates.[9] When including individual fixed effects to control for unobserved heterogeneity, all coefficients become statistically insignificant and their magnitudes decrease drastically. These results are qualitatively similar to those of Eid et al (2007), who found no sprawl effects on body weight when first differencing. This finding seems to indicate that it is the time period, rather than the model specification, that is driving the results of my main analysis. One explanation is that the shorter time period from 1988 to 1994 does not allow for enough time for variation in an individual's weight in response to changes in the living environment.

Most studies do not estimate their regressions separately for women and men. However, previous research observed that the incidence of obesity for men and women can be very different for some cities (Ezzati et al. 2006). To see if this is a concern in the current study, I estimate equations (1) and (2) separately for men and women.

Table 6 reports the estimated population density coefficients for women and men. First, for the OLS model and then including individual fixed effects. Focusing on the estimations with MSA effects, the OLS results do not look very different for men and women. Furthermore, they are also comparable in magnitude to the original estimates for both women and men. When individual fixed effects are included, the coefficients are again close in size and close to the original fixed effects estimates in table 4. However, the population density coefficient becomes statistically insignificant for the dependent variables *weight* and *BMI* for women and for the dependent variable *obese* for men. As the magnitude of the coefficients is similar to the original estimates in table 4, it would

[9] As the main interest is the comparison of the population density coefficients, I do not report the results for other covariates

seem prudent to use those results and not the ones estimated separately for women and men.

## 6. Policy Implications

Urban sprawl, defined as excessive suburbanization, has been the cause for much research in the urban planning and economics literatures. Specifically, the urban planning community includes some strong proponents of growth restrictions, who feel that cities in the U.S. have expanded too much. New discoveries of unfavorable relationships between low density living environments and various social factors are often used as arguments against sprawl. Subsequently, those relationships are often discovered to be non-causal by the economic literature. For example, Brueckner and Largey (2007) showed that low density living increases, rather than decreases, social interaction, contrary to the claims of some sprawl critics.

Previous findings that people in less dense urban areas display higher weight seemed to lend critics of sprawl additional justification for their calls to restrict the spatial growth of cities. However, this justification appeared to be weakened by the study of Eid et al. (2007), who found that the effect of density on weight disappeared after controlling for sorting. By contrast, this paper's results show that a portion of the negative effect of density on weight remains after controlling for unobserved individual characteristics. This remaining effect, which controls for the sorting of individuals across locations, must then be viewed as causal. Therefore, this paper's results buttress the claims of the earlier literature, which were unjustified at the time given the failure to control for sorting. The

results thus provide ammunition for the sprawl critics by showing a causal relationship between population density and body weight.

One important question that follows is whether the weight gain due to low density living environments brings about an externality. If an externality exists, it might provide a reason to implement policies that restrict suburban growth. The literature has identified some externalities that lead to such a prescription. For example, the benefits of open space on the fringe of cities are not taken into account when the land is developed for urban use, resulting in a negative externality for individuals who valued these benefits. Similarly, commuters travelling on congested highways generate a negative externality by slowing down their fellow drivers. With the social cost of commuting thus higher than the private cost, commuter trips are too long and cities too spread out. Policy instruments to restrict the excessive urban expansion caused by such externalities are readily available, as discussed by Brueckner (2000, 2001).

In contrast to these examples of externalities, there are no real externalities that result from the additional weight gain due to low density living. In other words, the negative effects of weight gain are mostly borne by the individual himself.[10] Hence, under perfect information, there is no basis for a policy intervention to restrict urban growth as a means of addressing the negative consequences of weight gain. If policy makers think that the general public is unaware of the possible weight gain due to low density living however, an information campaign to remedy this information asymmetry is a policy instrument that might be employed. If individuals are truly not aware of the

[10] It could be argued that increased health care cost for heavier individuals represents an externality, because health insurance costs will increase for all insured individuals, not just those who gained weight. However, the additional individual health insurance costs to people who did not gain weight are very likely to be negligible.

fact that moving to suburban areas might cause them to gain weight, an information campaign that educates the public to this possible risk might be enough to correct for the information asymmetry. Once the public is educated about the weight effects of moving to less dense areas, individuals can then either choose whether or not to move to suburbs, or they could increase their physical activity in response to a move. If individuals are actually aware of the possible weight gain and nevertheless choose to move to less dense areas without altering their physical activity, they alone will bear the cost of the additional weight gain they might experience.

## 7. Conclusion

The health policy literature originally discovered the correlation between suburban living and higher rates of obesity. Some papers go so far as to suggest that this relationship is causal, but given the data structure and estimation strategies employed in these studies, this conclusion cannot be drawn. By contrast, the urban economic literature suggests that most or all of the observed weight gain could be attributed to the sorting of individuals into certain neighborhoods (Eid et al. 2007, Platinga and Bernell 2007). The truth lies somewhere in between.

In this paper I estimate the amount of weight gain that can be attributed to the sorting of individuals due to unobserved characteristics. Compared to the previous urban economic literature, I show that a significant part of the estimated cross-sectional effect of suburban living on body weight remains after controlling for the spatial sorting of individuals. Using panel data over a 22 year time period, the analysis shows that cross-sectional estimates of the effect of population density on weight are upward biased. By

using individual and regional fixed effects, I estimate that sorting of individuals based on their unobserved characteristics accounts for up to 70 percent of the observed weight gain. The cross-sectional OLS estimates predict a weight gain of 11.7 pounds or an increase of 1.72 BMI points for a move from the densest census tract, located in the Northeast, to the least dense tract, located in the West. But when controlling for individual sorting, I estimate that the same move would cause an individual to only gain about 4.6 pounds or increase his BMI by 0.75 points[11].

This paper concludes that the actual effect of urban population density on the weight of individuals is not as large as the public health literature would want people to believe, but it is neither as insignificant as the urban economic literature to date estimated. But while a certain amount of weight gain can be attributed to differences in population density, policies to restrict urban spatial growth to address this effect are not warranted, given that no externality is involved. Informational policies that educate the public about this particular consequence of suburban living might be all that is needed to address the issue.

[11] Due to confidentiality concerns I am not able to identify the location of these census tracts beyond the region of the individuals' residence.

## References


Brueckner J.K., A.G. Largey. 2007. "Social interaction and urban sprawl." *Journal of Urban Economics*, forthcoming.

Brueckner, Jan K. 2000. "Urban Sprawl: Diagnosis and Remedies." *International Regional Science Review* 23(2):160-171.

Brueckner, Jan K. 2001. "Urban Sprawl: Lessons from Urban Economics." in William G. Gale and Janet Rothenberg Pack, eds., *Brookings-Wharton Papers on Urban Affairs*, Brookings Institution Press, pp. 65-89.

Centers for Disease Control and Prevention (CDC). 2002. *Behavioral Risk Factor Surveillance System Survey Data*. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Eid, J., H.G. Overman, Puga D., Turner M.A. 2007. "Fat City: Questioning the Relationship Between Urban Sprawl and Obesity." *Journal of Urban Economics*, forthcoming.

Ewing R., T. Schmid, R. Killingsworth, A. Zlot, S. Raudenbush. 2003. "Relationship between urban sprawl and physical activity, obesity, and morbidity." *American Journal of Health Promotion* 18(1):47–57.

Ewing, R., R.C. Brownson, D. Berrigan. 2006. "Relationship Between Urban Sprawl and Weight of United States Youth." *American Journal of Preventive Medicine* 31(6):464-474.

Ezzati, Majid, Hilarie Martin, Suzanne Skjold, Stephen Vander Hoorn, and Christopher J. L. Murray. 2006. "Trends in national and state level obesity in the USA after

correction for self report bias: analysis of health surveys." *Journal of the Royal Society of Medicine* 99(5):250–257.

Frank, L.D., M.A. Andresen, T.L. Schmid. 2004. "Obesity relationships with community design, physical activity, and time spent in cars." *American Journal of Preventive Medicine* 27(2):87–96.

Glass, T.A., M.D. Rasmussen, B.S. Schwartz. 2006. "Neighborhoods and Obesity in Older Adults: The Baltimore Memory Study." *American Journal of Preventive Medicine* 31(6):455-463.

Lopez, Russ. 2004. " Urban Sprawl and Risk for Being Overweight or Obese." *American Journal of Public Health* 94(9):1574-1579.

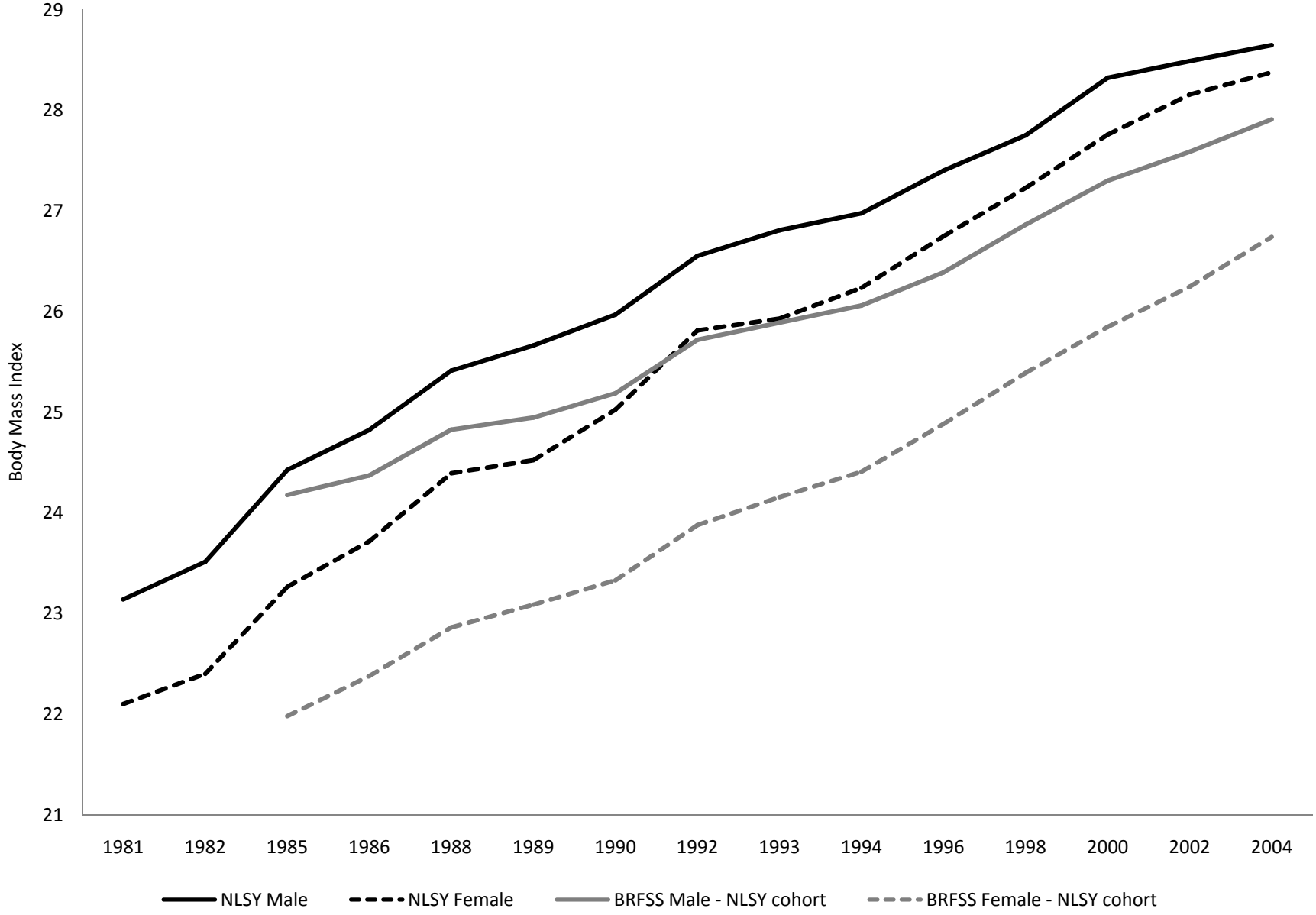
Nelson, M.C., P. Gordon-Larsen, Y. Song, B.M. Popkin. 2006. "Built and Social Environments: Associations with Adolescent Overweight and Activity." *American Journal of Preventive Medicine* 31(2):109-117.

Plantinga, A.J., S. Bernell. 2007. "The association between urban sprawl and obesity: Is it a two-way street?" *Journal of Regional Science* 47(5): 857–879.

Saelens, B.E., J.F. Sallis, J.B. Black, D. Chen. 2003. "Neighborhood based differences in physical activity: An environment scale evaluation." *American Journal of Public Health* 93(9):1552–1558.

Wooldridge, Jeffrey M. 2002. "Econometric Analysis of Cross Section and Panel Data." MIT Press. Cambridge, MA.

## Figure 1. Average Male and Female BMI
## NLSY 1981-2004 & BRFSS 1984-2004



## Figure 2. Male and Female BMI - U.S. Population and NLSY79 Cohort BRFSS 1984-2006

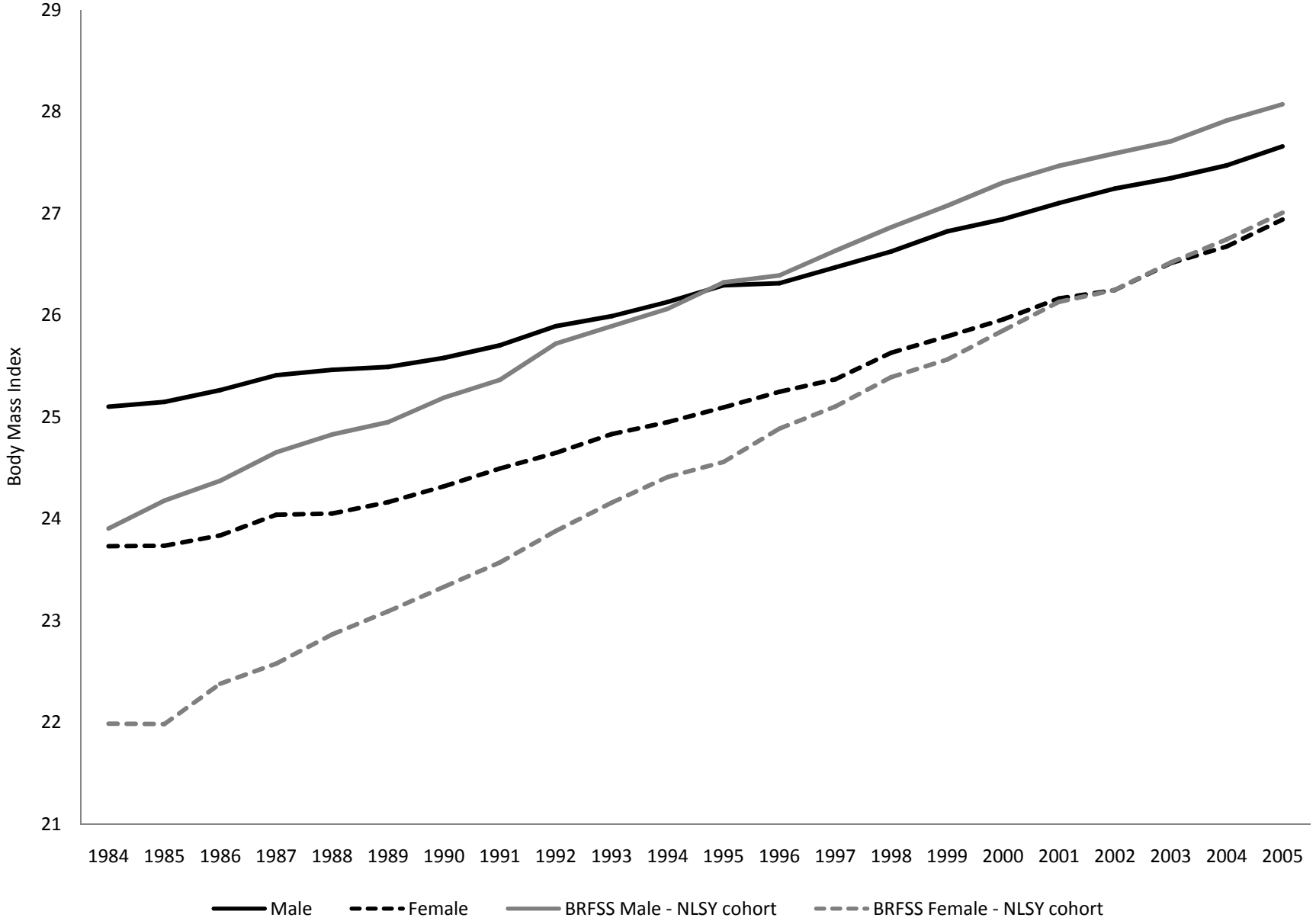


**Table 1. Restrictions on the full sample**
**NLSY 1981 to 2004**

| | Observations | Difference |
|---|---|---|
| Full Sample | 161240 | |
| (-) Missing observations | 81652 | 79588 |
| (-) Farm income | 80995 | 657 |
| (-) Active military | 80628 | 367 |
| (-) Rural population | 63584 | 17044 |
| (-) Not living in MSA or MSA not identifiable | 58858 | 4726 |
| (-) Non-movers | 40031 | 18827 |
| **Restricted Sample** | **40,031** | |

## Table 2. Sample Means and Standard Deviations

| | Mean | Standard Deviation |
|---|---|---|
| Population Density ('00000s per sqmi) | 0.069 | 0.149 |
| Weight | 162.30 | 37.62 |
| Height | 67.245 | 4.085 |
| Body Mass Index (BMI) | 25.103 | 4.821 |
| Underweight | 0.029 | 0.168 |
| Healthy Weight | 0.532 | 0.499 |
| Overweight | 0.302 | 0.459 |
| Obese | 0.138 | 0.344 |
| Age | 29.780 | 6.862 |
| Age Squared | 933.95 | 421.5 |
| Female | 0.502 | 0.500 |
| Single | 0.384 | 0.486 |
| Married | 0.463 | 0.499 |
| Other Marital Status | 0.153 | 0.360 |
| White | 0.600 | 0.490 |
| Black | 0.219 | 0.413 |
| Hispanic | 0.160 | 0.367 |
| Other Race | 0.022 | 0.145 |
| Family Size of HH | 2.982 | 1.687 |
| Number of Children living in HH | 0.889 | 1.152 |
| Annual Family Income in $'00000s | 0.624 | 1.118 |
| Highest Grade Completed | 13.453 | 2.509 |
| Real Work Experience | 8.096 | 5.132 |
| Mothers' Highest Grade Completed | 11.483 | 3.174 |
| Fathers' Highest Grade Completed | 11.658 | 4.029 |
| Sample Size | 40,031 | |

## Table 3. The Effect of Population Density on Weight, BMI, and Obesity - NLSY 1981 to 2004

| | Weight | | BMI | | Obese | |
|---|---|---|---|---|---|---|
| | State | MSA | State | MSA | State | MSA |
| **Population Density** | **-6.624*** ** | **-5.524*** ** | **-0.939*** ** | **-0.812*** ** | **-0.051*** ** | **-0.042*** ** |
| | (0.988) | (1.187) | (0.154) | (0.184) | (0.011) | (0.014) |
| Height | 4.475*** | 4.481*** | | | | |
| | (0.056) | (0.057) | | | | |
| Female | -11.087*** | -10.848*** | -1.397*** | -1.370*** | -0.005 | -0.003 |
| | (0.427) | (0.432) | (0.046) | (0.047) | (0.003) | (0.003) |
| Age | 2.765*** | 2.588*** | 0.446*** | 0.418*** | 0.012*** | 0.012*** |
| | (0.294) | (0.295) | (0.046) | (0.046) | (0.003) | (0.003) |
| Age Squared | -0.036*** | -0.033*** | -0.006*** | -0.005*** | -0.000* | -0.000* |
| | (0.005) | (0.005) | (0.001) | (0.001) | (0.000) | (0.000) |
| Married relative to single | 0.072 | 0.168 | -0.058 | -0.043 | -0.007 | -0.006 |
| | (0.401) | (0.402) | (0.063) | (0.064) | (0.005) | (0.005) |
| Other relative to single | -4.142*** | -4.175*** | -0.734*** | -0.738*** | -0.045*** | -0.044*** |
| | (0.502) | (0.505) | (0.080) | (0.080) | (0.006) | (0.006) |
| Black relative to White | 9.529*** | 9.785*** | 1.529*** | 1.575*** | 0.085*** | 0.091*** |
| | (0.411) | (0.425) | (0.065) | (0.067) | (0.005) | (0.005) |
| Hispanic relative to White | 5.872*** | 5.252*** | 0.957*** | 0.855*** | 0.043*** | 0.036*** |
| | (0.481) | (0.496) | (0.075) | (0.077) | (0.006) | (0.006) |
| Other relative to White | 0.025 | 0.041 | -0.062 | -0.062 | -0.003 | -0.004 |
| | (0.943) | (0.972) | (0.143) | (0.147) | (0.010) | (0.011) |
| Family Size | -0.074 | -0.164 | -0.008 | -0.021 | 0.001 | 0.001 |
| | (0.113) | (0.113) | (0.018) | (0.018) | (0.001) | (0.001) |
| # of Children in Household | 0.950*** | 1.008*** | 0.148*** | 0.157*** | 0.007*** | 0.008*** |
| | (0.199) | (0.199) | (0.031) | (0.031) | (0.002) | (0.002) |
| Family Income | -0.083*** | -0.080*** | -0.013*** | -0.013*** | -0.001*** | -0.001*** |
| | (0.013) | (0.013) | (0.002) | (0.002) | (0.000) | (0.000) |
| Highest Grade Completed | -0.471*** | -0.414*** | -0.081*** | -0.072*** | -0.005*** | -0.005*** |
| | (0.073) | (0.075) | (0.011) | (0.012) | (0.001) | (0.001) |
| Real Work Experience | 0.192*** | 0.215*** | 0.018** | 0.022** | -0.001 | -0.001 |
| | (0.057) | (0.057) | (0.009) | (0.009) | (0.001) | (0.001) |
| Highest Grade Completed (Mother) | -0.250*** | -0.275*** | -0.040*** | -0.044*** | -0.001 | -0.001 |
| | (0.064) | (0.064) | (0.010) | (0.010) | (0.001) | (0.001) |
| Highest Grade Completed (Father) | -0.248*** | -0.199*** | -0.044*** | -0.036*** | -0.003*** | -0.002*** |
| | (0.049) | (0.049) | (0.008) | (0.008) | (0.001) | (0.001) |
| Observations | 40031 | 40031 | 40031 | 40031 | 40031 | 40031 |
| R-squared | 0.45 | 0.47 | 0.18 | 0.21 | 0.08 | 0.10 |

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

All regressions include year effects although not reported

## Table 4. The Effect of Population Density on Weight, BMI, and Obesity
## Individual Fixed Effects - NLSY 1981 to 2004

| | Weight | | BMI | | Obese | |
|---|---|---|---|---|---|---|
| | State | MSA | State | MSA | State | MSA |
| **Population Density** | **-2.050*** ** | **-2.181** ** | **-0.332*** ** | **-0.351*** ** | **-0.043*** ** | **-0.039*** ** |
| | (0.782) | (0.862) | (0.122) | (0.134) | (0.014) | (0.015) |
| Age | 1.443*** | 1.449*** | 0.220*** | 0.217*** | 0.005 | 0.005 |
| | (0.292) | (0.294) | (0.046) | (0.046) | (0.005) | (0.005) |
| Age Squared | -0.026*** | -0.026*** | -0.004*** | -0.004*** | -0.000 | -0.000 |
| | (0.003) | (0.003) | (0.000) | (0.000) | (0.000) | (0.000) |
| Married relative to single | 2.896*** | 2.887*** | 0.436*** | 0.436*** | 0.013*** | 0.013*** |
| | (0.285) | (0.285) | (0.045) | (0.045) | (0.005) | (0.005) |
| Other relative to single | -1.290*** | -1.280*** | -0.233*** | -0.232*** | -0.016** | -0.015** |
| | (0.371) | (0.373) | (0.059) | (0.059) | (0.007) | (0.007) |
| Family Size | -0.282*** | -0.279*** | -0.047*** | -0.046*** | -0.001 | -0.001 |
| | (0.076) | (0.077) | (0.012) | (0.012) | (0.001) | (0.001) |
| # of Children in Household | 0.009 | -0.039 | 0.005 | -0.002 | 0.002 | 0.001 |
| | (0.141) | (0.142) | (0.022) | (0.022) | (0.003) | (0.003) |
| Family Income | -0.019*** | -0.018*** | -0.003*** | -0.003*** | -0.000*** | -0.000*** |
| | (0.006) | (0.006) | (0.001) | (0.001) | (0.000) | (0.000) |
| Highest Grade Completed | -0.251** | -0.180* | -0.048*** | -0.037** | -0.007*** | -0.007*** |
| | (0.106) | (0.107) | (0.016) | (0.017) | (0.002) | (0.002) |
| Real Work Experience | -0.015 | -0.016 | -0.030*** | -0.030*** | -0.001 | -0.002 |
| | (0.065) | (0.066) | (0.011) | (0.011) | (0.001) | (0.001) |
| Observations | 40031 | 40031 | 40031 | 40031 | 40031 | 40031 |
| Individual Observations | 5562 | 5562 | 5562 | 5562 | 5562 | 5562 |
| R-squared | 0.40 | 0.41 | 0.40 | 0.41 | 0.10 | 0.11 |

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p

All regressions include year effects although not reported

## Table 5. The Effect of Population Density on Weight, BMI, and Obesity - NLSY 1988 to 1994

| | Weight | | BMI | | Obese | |
|---|---|---|---|---|---|---|
| **OLS Regressions** | State | MSA | State | MSA | State | MSA |
| Population Density | -8.329*** | -7.361*** | -1.174*** | -1.050*** | -0.063*** | -0.083*** |
| | (1.682) | (2.026) | (0.265) | (0.319) | (0.020) | (0.024) |
| Observations | 16845 | 16845 | 40031 | 40031 | 40031 | 40031 |
| R-squared | 0.42 | 0.44 | 0.18 | 0.21 | 0.08 | 0.10 |
| | | | | | | |
| **Including Individual FE** | | | | | | |
| Population Density | -1.499 | -0.442 | -0.230 | -0.068 | -0.026 | -0.021 |
| | (1.220) | (1.313) | (0.194) | (0.209) | (0.024) | (0.026) |
| Observations | 16845 | 16845 | 16845 | 16845 | 16845 | 16845 |
| Individual Observations | 4499 | 4499 | 4499 | 4499 | 4499 | 4499 |
| R-squared | 0.15 | 0.18 | 0.14 | 0.17 | 0.03 | 0.06 |

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

All regressions include the same variables as reported in tables 3 and 4, including year effects although not reported

## Table 6. The Effect of Population Density on Weight, BMI, and Obesity - NLSY 1981 to 2004

| | MSA Effects | | |
|---|---|---|---|
| *FEMALES* | Weight | BMI | Obese |
| **OLS Regressions** | | | |
| Population Density | -4.131** | -0.698** | -0.042* |
| | (2.055) | (0.350) | (0.024) |
| Observations | 20106 | 20106 | 20106 |
| R-squared | 0.28 | 0.22 | 0.13 |
| | | | |
| **Including Individual FE** | | | |
| Population Density | -2.193 | -0.387 | -0.048** |
| | (1.433) | (0.243) | (0.024) |
| Observations | 20106 | 20106 | 20106 |
| Individual Observations | 2776 | 2776 | 2776 |
| R-squared | 0.37 | 0.38 | 0.12 |
| | | | |
| *MALES* | | | |
| **OLS Regressions** | | | |
| Population Density | -5.685*** | -0.789*** | -0.037* |
| | (1.696) | (0.240) | (0.021) |
| Observations | 19925 | 19925 | 19925 |
| R-squared | 0.36 | 0.23 | 0.12 |
| | | | |
| **Including Individual FE** | | | |
| Population Density | -2.085** | -0.295* | -0.032 |
| | (1.061) | (0.151) | (0.020) |
| Observations | 19925 | 19925 | 19925 |
| Individual Observations | 2786 | 2786 | 2786 |
| R-squared | 0.48 | 0.48 | 0.13 |

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

All regressions include the same variables as reported in tables 3 and 4, including year effects although not reported