Using Residential Sorting Behavior to Better Understand the Relationship Between Urban Greenspace and Health

Austin Williams

Department of Agricultural and Applied Economics, University of Wisconsin, Madison

December 30, 2016

Preliminary version. Please do not cite without permission.

Using Residential Sorting Behavior to Better Understand

the Relationship Between Urban Greenspace and Health

Abstract

A recent wave of research in public health finds associations between the proximity of an individual to greenspace and various health outcomes, including obesity, cardiovascular disease rates, depression, and anxiety. Based on these associations, it is sometimes assumed that better access to parks will lead to improved health outcomes. My project addresses this assumption by focusing on an issue sometimes ignored in this literature: people sort themselves into neighborhoods based on the characteristics of those neighborhoods and their personal preferences. Using observed neighborhood location decisions by young adults from the National Longitudinal Study of Adolescent to Adult Health (Add Health), I find evidence that accounting for time varying unobservables is crucial when estimating the relationship between greenspace and health. My study complements the current literature by yielding new evidence on how greenspace amenities impact health outcomes, and how heterogeneous amenity valuations may account for observed greenspace/health associations.

1) Introduction

The high percentage of Americans who classify as overweight or obese continues to raise public health concerns in the United States. Obesity rates have steadily increased over the past two decades, with recent estimates indicating that 69 percent of adults in the U.S. are overweight, and 35 percent are obese (Expert Panel Report, 2014). Though the trend has started to level off in recent years, the high rates may have long term health consequences. Clinical evidence suggests that obesity increases an individual's risk of morbidity from a wide range of conditions including hypertension, heart disease, and diabetes. Finkelstein et al. (2009) estimate that obese individuals incur $1,429 more in medical expenses annually than normal weight individuals. This translates to $147 billion dollars in additional expenditures in the United States, or 9.1 percent of total spending. Further, costs of obesity may include decreased worker productivity and other morbidity costs beyond those captured by medical expenditures.

Due to the high costs of obesity, understanding its underlying causes is important for designing policies to address it. While it is true that genetic differences explain why some individuals are more likely to gain weight than others, these differences do not explain the recent trends in obesity rates. At its core, obesity is a product of too much caloric intake relative to caloric expenditure. Potential underlying causes of this caloric gap include changes in relative food prices that encourage more food consumption, occupational changes from labor intensive jobs in manufacturing to jobs in the service or technology sectors, and built environment changes that discourage physical activity (Rosin 2008).

Even though decreasing caloric intake is the best way to manage weight (Expert Panel Report, 2014), a recent wave of research in public health finds associations between the proximity of an individual to built environment characteristics related to physical activity, which I will refer to broadly as greenspace, and obesity. Since physical activity has health benefits beyond weight loss, this literature relates greenspace to a broad range of health outcomes. Access to public parks is often the environmental characteristic of interest, but greenspace can also include other measures such as local tree coverage. Studies have found correlations between greenspace and obesity rates, cardiovascular illness, stress, depression, anxiety, and self-reported health (Lee and Maheswaran 2010; Beyer et al. 2014). Though the evidence specific to park-obesity associations has generally been weak or mixed (Coombes et al. 2010, Potestio et al. 2009), a stronger relationship is often found between parks and physical activity measures.

There are several hypothesized causal pathways that link greenspace to health outcomes. Close proximity to a park reduces the cost of utilizing it and may encourage individuals to engage in more healthy activities, such as walking, running, or playing sports. Even if the park is not used for vigorous exercise, walking to and from the park alone may constitute an activity increase. Over time, this increase in physical activity can lead to health improvements. Further, greenspace may be negatively correlated with other environmental "bads", such as localized air pollution (McPherson et al. 1994). Finally, it is possible that natural scenery is intrinsically good for mental health and offers a refuge from otherwise stressful urban environments.

My study addresses one of the biggest empirical issues prevalent in this literature: people sort themselves into neighborhoods based on the characteristics of those neighborhoods and their personal preferences. Much of the association between parks and health outcomes or physical activity measures likely comes from the fact that people who are physically active will choose to live in neighborhoods with amenities that cater to active recreation. There is strong evidence that housing prices implicitly include the values of local amenities, such as air pollution and open space (Klaiber and Phaneuf 2010; Bayer, Keohane, and Timmins 2009). These studies yield insight into how much people are willing to pay for different levels of an amenity. In the current setting, greenspace valuations may vary by individual preferences, which will impact residential location choices. This is the key mechanism through which heterogeneous preferences may lead to observed associations between health characteristics and local greenspace. For instance, a physically fit person may have preferences that induce them to seek out amenities that support their healthy lifestyle. This would lead to estimates which overstate the effect of greenspace on health. Alternatively, if unfit individuals seek out healthy amenities in order to improve their own health, then sorting will bias results in the opposite way.

To address this concern, some authors have implemented within person estimators to control for individual preferences that are constant over time. If residential sorting is based on unchanging unobserved characteristics, then a within person estimator resolves the issue. Using a first difference estimator, Eid et al. (2008) find no significant relationship between obesity status and urban sprawl. Boone-Heinonen et al. (2010) apply an individual fixed effects estimator to data on built environment characteristics and physical activity rates. They find a small but positive and significant impact of private recreation facilities (e.g. private gym or athletic club), but no effect from other characteristics including public park facilities, street connectivity, and landscape diversity. In addition, Baum and Chou (2015) implement a fixed effects estimator and find that urbanization impacts obesity. These studies provide more reliable results of health/environment relationships, but they do not account for sorting based on time-varying unobservables.

Other studies use instrumental variables (IV) to address the endogeneity issue. Courtemanche and Carden (2011) use an instrument based on Walmart store location decisions to explore how the spread of Walmart stores impacts obesity rates. They find that an additional Walmart in a community increases an individual's probability of obesity by 2.3 percent, an effect that operates through gaining access to cheaper food. Similarly, Dunn (2010), explores how access to fast food impacts obesity. Using the number of interstate exits to instrument for fast food restaurants, Dunn finds evidence of a positive effect in counties with medium population density. Also using an IV strategy, Zhao and Kaestner (2010) find that urban sprawl contributed to 13 percent of the recent increase in obesity rates in the U.S. Finally, multiple papers have used IV estimators when studying the impact of the built environment on travel behaviors (Boarnet and Sarmiento 1998, Khattak and Rodriguez 2005).

I use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) to estimate the relationship between parks and obesity in a more comprehensive way. My study utilizes both fixed effects and IV to identify the impact of parks on obesity, which allows me to compare the impact of both strategies in an empirical setting. Add Health follows students from grades 7-12 into young adulthood and combines data on both person-level characteristics and detailed neighborhood features, including the number of parks in proximity to a person's home. For a given residential address, park counts within 1 km do not change much over time, so my estimates rely on spatial variability in built environment characteristics. Because I use individual fixed effects to control for unobserved heterogeneity, I restrict my sample to individuals that move between survey waves to ensure sufficient within-person variability in neighborhood attributes. While this sample restriction invokes selection concerns, this is less of a problem in my setting, because most children move away from home in the years following high school graduation.

To address time-varying unobserved variables, I use logic from the residential sorting literature to select valid instruments in my setting. In particular, I use neighborhood characteristics that are likely to be correlated with parks, but unrelated to health outcomes, conditional on other covariates. Including person fixed effects decreases the magnitude of the estimates when compared to pooled OLS, but the IV strategy more than compensates for this reduction. In my preferred specification, I find that one additional park within 1 km reduces residents' BMI by 1.25 percent.

2) Data

My data comes from Add Health[1], which combines data on both person-level characteristics with detailed neighborhood features. Add Health follows a cohort from grades 7-12 into early adulthood. Adolescents undergo a number of significant life transitions during this period, so time-varying unobservables are of particular concern for this population. I focus on Waves 1 (1994-95) and 3 (2001-02) of the survey, because they include comprehensive neighborhood amenity characteristics. These characteristics come from a number of sources, including government and proprietary data sets, and were merged with person-level survey data before being released for use. Geographic data is not explicit, but masked identifiers down to the Census block group level are observed by the researcher. Because of this, I do not know where the survey participants live, but I do know if two participants live in the same census block group, tract, county, or state. Sampling in Wave 1 is clustered on school systems, but participants are relatively dispersed by Wave 3. Specifically, the subsample of participants that move between waves live in 972 distinct census tracts (in 138 counties) in Wave 1, and 5062 census tracts (in 876 counties) by Wave 3. Approximately 79 percent of participants change residence between these two interview periods, and 70 percent live independently (no longer with parents) by Wave 3. Participants who move away from home by Wave 3 are of particular interest for two

[1] Due to the confidential nature of the data, access required the completion of a restricted use data contract through the Carolina Population Center at University of North Carolina-Chapel Hill.

reasons. First, moving ensures variation in environmental characteristics, which tend not to change much over time. Second, participants who do not live with their parents are more likely to make independent housing choices. Concerns about movers being a selected sample are in part alleviated by the fact that most individuals move away from home during this period of their life, and are therefore fairly representative of this age group.

A measure of Body Mass Index (BMI) will be the main health outcome utilized in my empirical analysis. BMI is a function of height and weight, and serves as a proxy for obesity. This is not a perfect health measure, as it does not take body fat percentage into account, but it is widely used in the literature and the best measure available in the data. A number of individual level demographic variables are available, in addition to the health measures. Table 1 presents summary statistics for these variables, including education, marriage status, and if the respondent has children. By Wave 3, participants are aged between 18 and 27. Around 21 percent are obese (classified as having a BMI of 30 or greater), over 20 percent are married, and 27 percent are enrolled as full time students.

Most residential locations are identified by geocoded address or GPS measurement, so contextual variables in the data are precisely measured. Neighborhood characteristics come from outside sources and do not rely on respondent recall or estimates, and they are merged at the census block group level or the exact residential address. For example, information on median housing value and housing unit density are known by block group. Other variables, such as number of parks, are measured as counts within a 1 kilometer street network of an individual's residence. The network distance best represents how far an individual must travel to access a park, so it offers the best measure of access cost. Previous work has found the strongest association between built environment features and physical activity within 1-3 km buffers

(Boone-Heinonen 2010), so I follow this in my preferred specification. These measures are also available at 5 and 8 km distances, and using alternative measures does not qualitatively impact my results. Additional neighborhood variables are summarized in Table 2. The alpha index measures street connectivity, which proxies for the walkability of a neighborhood. Higher values indicate higher connectivity. The mean fractal dimension index (MFDI) serves as a second land use control. Values near one are indicative of an urban environment, while higher values are associated with more natural settings. These two variables control for urban sprawl/urbanization, which is known to be associated with obesity. Further, I control for economic, weather, and health variables that may impact BMI and/or how people interact with parks.

If parks are heterogeneous in terms of quality, a simple park count may not be the best measure of the amenities offered. To address this, I use additional park measures that may be more indicative of quality or accessibility. These include categories of parks or facilities associated with physical activity that are grouped based on primary Standard Industrial Classification (SIC) codes, which implies they have some sort of commercial operations. For example, I compare the effect of parks that require a membership, like a private country club, to public parks without fees or other entry restrictions. Definitions and examples of these alternative park definitions can be found in Table 3.

Table 4 presents summary statistics by obesity status and income indicators. Columns 1 and 2 compare park counts for individuals living in block groups with housing values below and above the median in my sample. Low housing value neighborhoods have fewer overall, membership, outdoor, and public parks. However, there are significantly more YMCA facilities in these neighborhoods, and the difference in means for the membership category is not significantly different from zero. The lack of significance for this category may indicate that private membership facilities, which are not public goods, are not reflected in nearby housing prices. Still, this generally supports the idea that parks are positive amenities whose value is reflected in housing prices. An obesity comparison yields the expected results, in that obese individuals live near fewer parks across all types.

3) Empirical Strategy

To be clear about the empirical challenges faced, consider estimating the health equation

$$Health_i = \beta_0 + \beta_1 Parks_i + \beta_2 N_i + \beta_3 X_i + \varepsilon_i \,, \tag{1}$$

where *N* includes other neighborhood characteristics and *X* is a vector of person-level controls. I am interested in the effect of an additional park on health, but estimates will be biased if the error term contains unobservables that impact both the level of parks near one's residence and health. More precisely, bias occurs if the error term can be written as

$$\varepsilon_i = c_i + \eta_i \tag{2}$$

and $E[c_i | parks_i, N_i, X_i] \neq 0$. This will be the case if unobservable preferences for healthy lifestyles determine both the number of parks near one's chosen residence and level of fitness.

Within-Person Estimator

With panel data, the issue outlined above can be resolved. The estimating equation becomes

$$Health_{it} = \beta_1 Parks_{it} + \beta_2 N_{it} + \beta_3 X_{it} + c_i + \eta_{it}, \tag{3}$$

and the unobserved heterogeneity c_i can be removed through a first difference transformation or by including individual fixed effects. However, if the error term in (3) contains a time varying component d_{it} such that $E[d_{it} | parks_{it},\ N_{it},\ X_{it}, c_i] \neq 0$, a within-person estimator will not yield consistent results.

Instrumental Variables Estimator

If an appropriate instrument is available, then an IV strategy will allow for time-varying unobserved variables that are correlated with parks. Some examples of using this approach to address endogeneity caused by residential sorting come from the transportation literature. Boarnet and Sarmiento (1998) estimate how land use influences transportation behavior. They use neighborhood racial composition and age of the housing stock as instruments for land-use characteristics, such as population density and street connectivity. Validity of the IV strategy requires unobserved preferences for the instruments, which are also taken into account by the individual when choosing a residential location, to be unrelated to preferences for the land use variables. For instance, in the aforementioned study, preferences for racial composition must be unrelated to preferences for population density.

I propose a novel instrumental variables strategy using information from housing markets. The general idea is that greenspace will be correlated with other neighborhood amenities and housing characteristics that are plausibly unrelated to health. For instance, houses located near parks may also have more square footage or be located in better school districts. The data set includes a number of characteristics that may work well as instruments. First, I use median housing value in the respondent's block group. An increase in neighborhood quality will raise both demand for that location and the price of housing units there. In equilibrium, the value of parks, as well as unobserved neighborhood amenities, will be incorporated in median housing price. Through this mechanism, housing prices will be correlated with parks and also contain information on the unobserved quality of a neighborhood. Next, I use number of schools within 1 km. Schools may be associated with positive amenities such as safety or social cohesion, and they are often located adjacent to parks. Finally, I use number of housing units per square mile as an instrument. This serves as a measure of lot size, a positive characteristic. This variable may be higher in more suburban block groups, which may attract higher BMI residents, but controlling for land use characteristics in *N* should help alleviate these concerns. The Add Health data contains many health-related variables that will help absorb variation in my instruments that may be related to health outcomes.

Despite my careful controls, there may be channels through which an instrument may impact health apart from its relationship with parks. To address this concern, I use attributes from other block groups in the same county, as an alternative set of instruments. The reasoning for these new instruments is similar to that presented in Bayer and Timmons (2007). When making a residential location decision, individuals consider all available alternatives in the market and choose a location that maximizes their utility conditional on their preferences and budget constraint. Demand for a location may also influence the neighborhood attributes themselves. For example, if many people have preferences for living near city centers, then housing density and prices will be higher in these areas. More schools may be built in these neighborhoods to serve demand. They will also tend to have higher population densities, which could be seen as a negative amenity. Bayer and Timmons (2007) focus on this congestion effect in their paper. Housing market equilibriums result from this complex interplay between supply and demand, and it is therefore reasonable to assume that certain neighborhood characteristics will be a function of available amenities in other neighborhoods in the same market. However, the attributes of other neighborhoods are unlikely to have a direct impact on health.

I propose using instruments *Z* for greenspace in the following specification:

<u>First Stage:</u>

$$Parks_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 N_i + \alpha_3 X_i + \zeta_i \tag{4}$$

Second Stage:

$$Health_i = \beta_0 + \beta_1 \widehat{Parks}_i + \beta_2 N_i + \beta_3 X_i + \epsilon_i \tag{5}$$

The success of this IV strategy depends on two main requirements: (a) that *Z* is correlated with parks, and (b) that *Z* does not describe health-related amenities beyond what I control for in *N*. The first condition should be satisfied due to the nature of housing market equilibriums, and is verified in my analysis. The second requirement relies on having sufficiently controlled for health-related characteristics.

4) Results

Preliminary Evidence

With my rich data, I examine the relationship between greenspace and health using multiple estimation strategies. Table 5 offers some preliminary evidence of this relationship. Although I observe each individual twice in the data, I first treat all observations as independent and estimate a pooled OLS regression, the results of which are presented in the first column of Table 5. As hypothesized, an increase in number of parks within one kilometers an individual's residence is found to have a small but significant effect in reducing BMI. My log-linear specification implies that one additional park decreases BMI by 0.36 percent. Though it should not be interpreted as causal, finding a significant relationship here is interesting in itself, given the mixed results from previous studies that have attempted to find direct associations between parks and BMI. Being married, having children, and higher education levels are also positively correlated with BMI, while being a full time student is associated with a significantly lower BMI. Column 2 in this table regresses number of parks in wave 3 of the survey on number of parks in wave 1 of the survey. Even though I restrict my sample to individuals who move between these two survey ways, there is still a strong positive association between the numbers of parks near an individual over time. This suggests that individuals move to neighborhoods with characteristics similar to those where they lived previously, a reminder that these characteristics are not randomly distributed. Further, higher BMIs in Wave 1 are associated with fewer parks in Wave 3, justifying the concern that sorting is driving observed park/health relationships. Somewhat surprisingly, participants with children in the third wave are less likely to live near parks. This could be due to budget constraints, as raising children is expensive and parks are positive amenities that are reflected in higher housing values. In column 3 I regress log-transformed BMI on Wave 3 parks and Wave 1 BMI. Unsurprisingly, I find that BMI is highly persistent over time. However, even after controlling for previous BMI, the number of parks is still contemporaneously correlated with BMI. This initial evidence could suggest that, although sorting on health characteristics does occur, it may not fully explain correlations between parks and BMI.

Main Results

Comparing how estimates vary across specifications gives a more nuanced understanding of the factors that contribute to observed associations between greenspace and health. In column 1 of Table 6, I estimate a model that includes individual fixed effects. This approach accounts for individual level unobserved characteristics that are not changing over time. To the extent that sorting behavior is explained by these characteristics, fixed effects estimates are more representative of the causal relationship between parks and BMI. Similar to previous findings, the significance of the parks coefficients disappears in this setting. This result presents more evidence that sorting is driving the observed associations, and it implies that living near a park has no discernible impact on obesity.

Though fixed effects estimates control for unobservables that do not change over time, time-varying variables that are left unaccounted for may still bias results. This is of particular concern in my setting, as subjects move from adolescence into young adulthood. This is a period of immense change for many individuals, and it is reasonable that health attitudes or health-related behaviors may be changing during this transitional period. To address this concern, I implement the instrumental variables strategy described above. Columns 2 and 3 of Table 6 instrument for *parks* using the following instruments: median house value, housing unit density, and number of schools within 1 km. In all of the IV models, *parks* has a significant and negative impact on BMI. Consistent with the difference observed between the pooled OLS and Fixed Effects estimates, inclusion of individual fixed effects in column 3 leads to a smaller estimated coefficient than in column 2. This pattern is consistent with negative bias from unobserved time-constant variables, but the opposite-signed bias from time-varying unobservables. The latter may result from higher BMI individuals seeking out physical activity amenities in order to improve their fitness. Similar to Eid et al. (2008), I find no significant effect from land-use controls when fixed effects are included.

Robustness

Even after including health controls, there may be hypothesized channels through which each individual instrument impacts health. However, since my three instruments measure dissimilar amenities, there is not likely to be a common mechanism that influences BMI. Appendix Table A3, columns 3-5, shows results from the Fixed Effects IV model when each instrument is used separately. The striking similarity of results across instruments helps alleviate concerns that the instruments fail the required exclusion restriction. As a formal test, I use multiple instruments to run tests of overidentifying restrictions. As reported in Table 6, in each IV specification I fail to reject the null hypothesis that the instruments are valid. First stage F-statistics are also reported, and full first and second stage are presented in the appendix.

The final column in Table 6 uses my alternative set of instruments. Instead of using own location attributes to instrument for parks, I use the mean level of attributes in other block groups in the same county. That is, I use leave-one-out averages of block group median home price, housing unit density, and school counts. The coefficient estimate in column 4 is negative and significant, though somewhat smaller in magnitude than in column 3, which uses the same Fixed Effects IV specification. This is strong evidence that my results are not being driven by invalid instruments, since it is unlikely that amenities in other neighborhoods will have a direct impact on BMI.

The results from Table 6 use a count of all parks as the key independent variable. However, parks may vary widely in quality, accessibility, and amenities offered, so being more specific about park type gives a more nuanced understanding of the mechanisms through which parks impact health. Table 7 show results using alternative park definitions with the Fixed Effects IV strategy. All categories have negative and significant coefficients on the park measure. The coefficient on Membership facilities has the lowest magnitude. Since these types of parks have the highest barriers to entry, this result is consistent with the idea that increasing park access will increase use and therefore maximize health benefits. This may also be indicative of how these parks are used, since the impact of a private golf course on BMI likely differs significantly from a neighborhood park with a walking path. The magnitude of the estimates for the public and outdoor park categories are very similar to the estimates in column 3 of Table 6. Having a YMCA within 1 km of your residence is found to decrease BMI by over 4 percent. This large effect may relate to the types of physical activity amenities, such as swimming pools and gym equipment, often offered at a YMCA. These results suggest that the intensity of physical activities associated with a park facility matters for weight loss. Column 5 uses number of parks between 1 and 3 km from the participant's residence as the dependent variable. Parks in this outer perimeter are still found to have a significant impact on BMI, but the coefficient magnitude is substantially smaller. This supports the idea that proximity to a park influences its use and resulting health benefits.

Next, I investigate how climate influences park use. In places with relatively harsh winters, outdoor parks can only be utilized during part of the year. Theoretically, this should dampen the impact the park has on health. Columns 1 and 2 compare the effect of outdoor facilities in locations with January temperatures below and above the sample median. I find significant and negative effects on both subsamples, but, consistent with my hypothesis, the magnitude of the coefficient is almost 8 times larger in the warmer climates. As a further robustness check, I carry out the same comparison for the YMCA category. These facilities are more likely to have indoor physical activity amenities that can be used year round regardless of climate. The warmer climate coefficient in column 4 is large, but not statistically different from zero. I do find a significant effect for the colder climate subgroup. It is possible that people living in these climates must rely more heavily on indoor facilities, thus explaining why I only find a statistically significant effect for this group.

5) Conclusion

Using panel data with rich information on residential location choices, I estimate the relationship between greenspace amenities and health. I rely on an understanding of residential sorting behavior to find valid instruments for neighborhood amenity levels in order to address a standard endogeneity problem. My findings show that time-variant unobserved variables bias downward the estimated effect of access to greenspace on health. This implies that simply adding fixed effects in a panel setting may not be sufficient for identification.

OLS regressions on the pooled sample show a negative association between parks and BMI. Although this cannot be interpreted as a causal, finding a significant relationship in the cross section is promising, as previous cross-section studies have struggled to find direct associations between parks and BMI. Including individual fixed effects reduces the magnitude and significance of this finding. However, when jointly utilizing a fixed effects and instrumental variable estimation strategy, my preferred specification, I find that the addition of a generic park to a neighborhood reduces BMI by over 1 percent.

The policy implications of this finding depend on the costs of constructing and maintaining a park relative to the monetized health benefits of BMI reductions. A back of the envelope benefits calculation illuminates this trade off. A 3 km radius of a residential address is close to the geographic size of the average block in the sample, and I find that one additional park within 3 km decreases obesity rates by one third of a percent. On average, just over 2000 people live in the block groups in my sample, so this reduction equates to about 6.5 fewer obese individuals. If each obese individual incurs $1,429 in additional healthcare spending annually, then the monetized benefit of adding a park is $9,377 annually. Although this is a crude estimate, it demonstrates that health benefits alone likely cannot justify the costs of constructing and operating a park. However, my results suggest a much larger effect of health clubs, such as a YMCA, on BMI. Subsidizing gym memberships may be a more effective strategy if obesity reductions are the central policy objective.

6) References

Baum, C. L., & chou, S.-Y. (2016). Why Has the Prevalence of Obesity Doubled. *Review of Economics of the Household, 14*, 251-267.

Bayer, P., Keohane, N., & Timmons, C. (2009). Migration and Hedonic Valuation: The case of Air Quality. *journal of Environmental Economics and Management, 58*(1), 1-14.

Beyer, K. M., Kaltenbach, A., Szabo, A., Bogar, S., Nieto, F. J., & Malecki, K. M. (2014). Exposure to Neighborhoog Green Space and Mental Health: Evidence from the Survey of the Health of Wisconsin. *International Journal of Environmental Research and Public Health, 11*, 3453-3472.

Boarnet, M. G., & Sarmiento, S. (1998). Can land-Use Policy Really Affect Travel Behaviour? A Study of the Link between Non-Work Travel and Land-Use Characteristics. *Urban Studies, 35*(7), 1155-1169.

Boone-Heinonen, J., Guilkey, D. K., Evenson, K. R., & Gordon-Larsen, P. (2010). Residential Self-Selection Bias in the Estimation of Built Environment Effects on Physical Activity between Adolescence and Young Adulthood. *International Journal of Behavioral Nutrition and Physical Activity, 7*(70).

Coombes, E., Jones, A. P., & Hillsdon, M. (2010). The Relationship of Physical Activity and Overweight to Objectively Measured Green Space Accessibility and Use. *Social Science & Medicine, 70*, 816-822.

Courtemanche, C., & Carden, A. (2011). Supersizing Supercenters? The Impact of Walmart Supercenters on Body Mass Index and Obesity. *Journal of Urban Economics, 69*, 165-181.

Dunn, R. A. (2010). The Effect of Fast-Food Availability on Obesity: An Analysis by Gender, Race, and Residential Location. *American Journal of Agricultural Economics, 92*(4), 1149-1164.

Eid, J., Overman, H. G., Puga, D., & Turner, M. A. (2008). Fat City: Questioning the Relationship between Urban Sprawl and Obesity. *Jourbal of Urban Economics, 63*, 385-404.

Expert Panel Report: Guidlines (2013) for the Management of Overweight and Obesity in Adults. (2014, July). *Obesity, 22*(S2), S41-S409.

finkelstein, E. A., Trogdon, J. G., Cohen, J. W., & Dietz, W. (2009). Annual Medical Spending Attributable to Obesity: Payer- And Service-Specific Estimates. *Health Affairs, 28*(5), w822-w831.

Khattak, A. J., & Rodriguez, D. (2005). Travel Behavior in Noe-Traditional Neighborhood Developments: A Case Study in USA. *Transportation Research Part A, 39*, 481-500.

Klaiber, H. A., & Phaneuf, D. J. (2010). Valuing Open Space in a Residential Sorting Model of the Twin Cities. *Journal of Environmental Economics and Management, 60*, 57-77.

Lee, A., & Maheswaran, R. (2010). The Health Benefits of Urban Green Spaces: A Review of the Evidence. *Journal of Public Health, 33*(2), 212-222.

McPherson, E. G., Nowak, D. J., & Rowntree, R. A. (1994). *Chicago's Urban Forest Ecosystem: Results of the Chicago Urban Forest Climate Project.* U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station.

Potestio, M. L., Patel, A. B., Powell, C. D., McNeil, D. A., Jacobson, R. D., & McLaren, L. (2009). Is There an Association between Spatial Access to Parks/Green Space and Childhood Overweight/Obesity in Calgary, Canada? *International Journal of Behavioral Nutrition and Physical Activity, 6*(77).

Rosin, O. (2008). The Economic Causes of Obesity: A Survey. *Journal of Economic Surveys, 22*(4), 617-647.

Zhao, Z., & Kaestner, R. (2010). Effects of Urban Sprawl on Obesity. *Journal of Health Economics, 29*, 779-787.

Table 1: Wave 3 Person-Level Characteristics

Variable	Individuals	Mean	Std. Dev.	Min	Max
BMI	9093	26.167	6.151	12.293	66.130
Obese	9093	0.216	0.412	0	1
Education	9093	13.199	2.033	6	22
Married	9093	0.209	0.406	0	1
Children	9093	0.431	0.761	0	9
Full Time	9093				
Student		0.274	0.446	0	1
Age	9093	22.070	1.762	18	27

Statistics based on subsample of movers. Obese defined by having a BMI of 30 or greater.

Table 2: Summary Statistics of Neighborhood Characteristics

Variable	Definition	Obs	Mean	Std. Dev.	Min	Max
Park Measures						
Park Count 1k	Within 1 km of residence[a]	17239	0.911	1.425	0.000	16.000
Park Count 1-3k	Between 1-3 km of residence[a]	17239	4.812	6.428	0.000	54.000
Membership	Within 1 km[b]	17239	0.294	1.066	0.000	34.000
Outdoor	Within 1 km[b]	17239	0.145	0.512	0.000	15.000
Public	Within 1 km[b]	17239	0.142	0.484	0.000	11.000
YMCA	Within 1 km[b]	17239	0.050	0.293	0.000	7.000
Instruments						
Median Value	Median House Value[c] (BG)	14667	113126.100	89964.930	0.000	1000001.000
Unit Density	Housing Units per sq. km[c] (BG)	17232	598.038	2003.490	0.000	56482.500
Schools	Within 1 km[b]	17239	6.966	12.763	0.000	200.400
Geographic/Economic						
Area	Sq. km[c] (BG)	17239	26.501	107.052	0.012	4816.599
Alpha	Street Connectivity[a] (1 km)	17239	0.314	0.638	-8.000	10.000
MFDI	Landscape Diversity[d]	17239	1.071	0.028	1.004	1.182
MHI	Median Household Income[b] (BG)	17239	39094.950	20161.530	0.000	200001.000

Unemployment	Rate for >16 population[e] (BG)	17239	0.075	0.070	0.000	0.955
COLI	Cost of Living Index[f]	17239	1.083	0.219	0.850	2.370
<u>Health</u>						
Birthweight	Low birth weight proportion[a] (C)	17239	0.076	0.017	0.036	0.144
Medicaid	Spending per beneficiary[a] (S)	17239	3650.383	1281.579	441.714	7725.138
Mortality	Per 1,000[a] (C)	17239	8.336	2.015	1.488	18.338
Infant Mortality	White, Per 10,000[a] (C)	17239	55.668	57.618	0.000	240.000
Adult Arrests	Per 100,000[g] (C)	17239	714.260	394.695	0.000	9403.547
Juvinile Arrests	Per 100,000[g] (C)	17239	271.920	162.807	0.000	2310.586
<u>Weather</u>						
Precipitation	Mean total rainfall, July[h]	17239	3.371	2.057	0.000	9.110
Sun	Mean sunshine total hours, Annual[h]	17239	2785.888	412.815	1488.000	4015.000
Summer Temp	Mean daily max temp, July[h]	17239	86.996	6.415	63.800	108.700
Winter Temp	Mean daily min temp, January[h]	17239	28.119	12.406	-9.000	65.000
Snowfall	Mean total snowfall, Annual[h]	17239	17.167	20.812	0.000	86.900

BG indicates measure at Block Group level, C at the County level, and S at the State level. Weather norms come from the nearest weather station with non-missing data. Statistics based on subsample of movers. Data origin: a) ESRI StreetMap Pro; b) Dun and Bradstreet; c) U.S. Census; d) National land cover dataset; e) U.S. Bureau of Labor Statistics; f) American Chamber of Commerce Research Association; g) Uniform Crime Reporting data; h) Climate Atlas of the United States

Table 3 : Additional Park Measures

Park Type	Definition	Example
Membership	Require a membership	Country club, boating club, health club
Outdoor	Are "outdoor" in nature.	Campgrounds, ski slope, golf course, riding stable
Public	Free, public access	Tennis courts, community center, recreation center, public beach
YMCA	Non-profit aimed at improving community health and well-being	YMCA, YWCA

Categories based on Dun & Bradstreet primary Standard Industrial Classification. There is potential for overlap between categories. For example, a private golf course would be included in both the Membership and Outdoor categories.

Table 4: Wave 3 Parks by Income and Obesity Status

	(1) Low Value	(2) High Value	(3) Non-Obese	(4) Obese
Park Count	0.686	1.227	1.007	0.772
Membership	0.387	0.409	0.434	0.267
Outdoor	0.179	0.229	0.220	0.149
Public	0.168	0.196	0.192	0.147
YMCA	0.075	0.0566	0.070	0.050
Observations	4,546	4,540	7,122	1,964

Movers subsample. Park measures are counts within 1 km of residence in Wave 3. High value indicates greater than the median housing value ($96,300, block group level) in the sample. All value comparison means statistically different at the 5 percent level except for the membership category. For Obese comparison, YMCA means statistically different at 5 percent level, all other categories significant at 1 percent level.

Table 5: Preliminary Evidence

	(1) Log(BMI)	(2) Parks Wave 3	(3) Log(BMI) Wave 3
Parks	-0.00359**		
	(0.00161)		
Parks Wave 1		0.227***	
		(0.0195)	
Parks Wave 3			-0.00306***
			(0.000999)
BMI Wave 1		-0.00971***	0.0366***
		(0.00361)	(0.000448)
Full Time Student	-0.0551***	0.00343	0.00128
	(0.00681)	(0.0451)	(0.00352)
Children	0.0187***	-0.0391**	0.00499*
	(0.00543)	(0.0152)	(0.00263)
Education	0.00889***	0.0513***	-0.00245***
	(0.00120)	(0.00952)	(0.000802)
Married	0.0301***	-0.153***	0.0385***
	(0.00795)	(0.0340)	(0.00422)
Constant	2.899***	12.72***	2.375***
	(0.183)	(1.348)	(0.125)
Weather Controls	X	X	X
Health Controls	X	X	X
Time F.E.	X		
Observations	16,191	6,797	8,867
R-squared	0.148	0.229	0.565

Movers sample. Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Comparison of Fixed Effects and Instrumental Variable Approaches

	(1) FE	(2) Pooled IV	(3) FE IV	(4) FE IV
Park Count 1 km	-0.000184	-0.0239***	-0.0125**	-0.00784*
	(0.00161)	(0.00536)	(0.00529)	(0.00451)
F-Stat		50.78	14.79	33.37
Hansen J Stat (p-val)		0.5854	0.9159	0.9094
Observations	16,191	13,768	10,016	13,548
R-squared	0.511	0.133	0.525	0.502
Number of ID	8,949		5,008	6,774

Movers sample. Log-transformed BMI is the dependent variable. Median housing value, housing unit density, and school count of own location used as instruments for columns 2 and 3. Average median housing value, housing unit density, and school count of other locations in same county used as instruments for columns 4. Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Alternative Park Definitions

	(1) Membership	(2) Outdoor	(3) Public	(4) YMCA	(5) Parks ∈ (1,3] km
Park Count	-0.00466**	-0.0161**	-0.0173**	-0.0414**	-0.00208**
	(0.00217)	(0.00754)	(0.00729)	(0.0200)	(0.000911)
F-Stat	77.19	23.26	48.54	7.17	28.14
Hansen J Stat (p-val)	0.4123	0.3974	0.6956	0.5461	0.9696
Observations	10,016	10,016	10,016	10,016	10,016
Number of ID	5,008	5,008	5,008	5,008	5,008

Movers sample. Membership, Outdoor, Public, and YMCA counts within 1 km of residence. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 8: Climate's Influence on Park Utilization

	(1) Cold Outdoor	(2) Hot Outdoor	(3) Cold YMCA	(4) Hot YMCA
Park Count 1 km	-0.0113*	-0.0862**	-0.0303*	-0.167
	(0.00672)	(0.0397)	(0.0179)	(0.109)
Observations	4,498	4,874	4,498	4,874
R-squared	0.553	0.481	0.552	0.495
Number of ID	2,249	2,437	2,249	2,437

Log-transformed BMI is the dependent variable. Movers sample. Cold defined as having January minimum temperature below sample median (29 degrees Fahrenheit). Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Appendix

Table A1: Full Results

	(1) FE	(2) Pooled IV	(3) FE IV
Park Count 1 km	-0.000184	-0.0239***	-0.0125**
	(0.00121)	(0.00536)	(0.00529)
Full Time Student	0.00257	-0.0522***	0.00487
	(0.00549)	(0.00693)	(0.00642)
Children	0.00915***	0.0170***	0.0106***
	(0.00332)	(0.00535)	(0.00394)
Precipitation	0.00218	-0.00421**	0.000487
	(0.00222)	(0.00192)	(0.00281)
Sun	4.06e-06	-5.98e-07	4.88e-06
	(9.46e-06)	(7.68e-06)	(1.14e-05)
Summer Temp	0.000632	0.000798	0.000163
	(0.000437)	(0.000493)	(0.000521)
Winter Temp	-0.00177***	0.000198	-0.00149**
	(0.000466)	(0.000417)	(0.000592)
Snowfall	-0.000351	0.000110	-0.000336
	(0.000229)	(0.000211)	(0.000285)
Alpha	0.00102	0.00367	0.00149
	(0.00212)	(0.00380)	(0.00314)
Median Income	4.57e-08	-5.78e-07***	-2.75e-08
	(1.00e-07)	(1.27e-07)	(1.23e-07)
Education	0.00383***	0.00743***	0.00537***
	(0.00126)	(0.00132)	(0.00151)
Married	0.0371***	0.0259***	0.0288***
	(0.00593)	(0.00813)	(0.00712)
MFDI	-0.0436	-0.368*	-0.00676
	(0.132)	(0.193)	(0.160)
COLI	0.0151	0.00453	-0.00569
	(0.00973)	(0.0175)	(0.0181)
Birthweight	0.158	0.195	0.190
	(0.184)	(0.191)	(0.232)
Unemployment	0.0146	0.131***	0.0241
	(0.0272)	(0.0370)	(0.0363)
Medicaid	-2.97e-06	2.23e-06	-6.61e-06*
	(3.07e-06)	(3.35e-06)	(3.74e-06)
Mortality	0.00112	0.00165	0.00115
	(0.00127)	(0.00139)	(0.00161)
Infant mortality	-6.05e-05	-3.10e-05	-7.32e-05
	(7.93e-05)	(0.000140)	(0.000107)
Adult Arrests	6.07e-06	2.26e-05**	8.95e-06

	(7.68e-06)	(1.04e-05)	(8.98e-06)
Juvenile Arrests	-1.43e-05	-1.81e-05	-4.90e-06
	(1.74e-05)	(2.38e-05)	(2.09e-05)
Time	0.114***	0.154***	0.116***
	(0.0118)	(0.0182)	(0.0150)
Constant	3.053***	3.310***	
	(0.146)	(0.214)	
Observations	16,191	13,768	10,016
R-squared	0.511	0.134	0.523
Number of ID	8,949		5,008

Movers sample. Log-transformed BMI is the dependent variable. Median housing value, housing unit density, and school count of own location used as instruments for columns 2 and 3. Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2: First Stage Results

	(1) First Stage
Median Housing value	1.08e-06*
	(6.54e-07)
Housing Unit Density	4.05e-05
	(4.36e-05)
Schools	0.0231***
	(0.00485)
Full time student	-0.129*
	(0.0782)
Children	0.0389
	(0.0413)
Precipitation	-0.150***
	(0.0313)
Sun	4.10e-05
	(0.000152)
Summer Temp	0.000265
	(0.00727)
Winter Temp	0.0285***
	(0.00652)
Snowfall	0.0128***
	(0.00301)
Alpha 1 km	0.0408**
	(0.0162)

Median Income	-7.40e-06***
	(2.02e-06)
Education	0.0427**
	(0.0167)
Married	-0.0676
	(0.0632)
MFDI	-5.276***
	(1.399)
COLI	0.272
	(0.210)
Birthweight	7.087***
	(2.194)
Unemployment	-0.312
	(0.606)
Medicaid	-1.76e-05
	(4.44e-05)
Mortality	-0.0353**
	(0.0174)
Infant mortality	0.000537
	(0.000943)
Adult Arrests	0.000257**
	(0.000102)
Juvenile Arrests	-0.000214
	(0.000212)
Time	0.302**
	(0.148)
Constant	4.645***
	(1.770)
Observations	13,985
Number of ID	8,813
R-squared	0.140

First stage results for the specification in Column 3 of Table 6. Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A3: Alternate IV Specifications

	(1) Obesity	(2) lnBMI	(3) Value	(4) Density	(5) Schools
Park Count 1k	-0.0215*	-0.0125**	-0.0143	-0.0141***	-0.0112**
	(0.0119)	(0.00529)	(0.0215)	(0.00463)	(0.00449)
F-stat	13.79	14.79	6.06	24.50	87.83
Observations	10,016	10,016	10,022	14,478	14,484
R-squared	0.143	0.525	0.520	0.499	0.503
Number of ID	5,008	5,008	5,011	7,239	7,242

Dependent variable is obesity status in column 1 and log-transformed BMI in columns 2-5. Median housing value used as instrument in column 3, housing unit density used as instrument in column 4, schools used as instrument in column 5, and all three are used in columns 1 and 2. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1