Body Mass Index, Diabetes, and Neighborhood Characteristics:
Assessing the Role of Selection by Studying Movers and Stayers

Ken R. Smith
Heidi Hanson
Barbara B. Brown
Cathleen D. Zick
Lori Kowaleski-Jones
Jessie X. Fan

University of Utah

September 2010
**Introduction**

Obesity and overweight are major public health problems. An estimated 65% of US adults are overweight or obese with up to 280,000 annual deaths attributable to obesity. This has also been accompanied by an increase in the prevalence in type II diabetes (T2D) as well as gestational diabetes. In an effort to understand factors associated with adult obesity/overweight and T2D/gestational diabetes (T2D-GD), attention has recently focused on the potential effects of environmental influences. Yet, studies linking the physical environment to the risk of being unhealthy weight and diabetes are limited by the fact that residents are not randomly distributed by neighborhood. If significant associations are found between neighborhood characteristics and individuals’ body mass indices (BMI) or T2D-GD in observational studies, one cannot confidently draw conclusions about causality. Neighborhood features may cause people to be more physically active, or physically active individuals with low BMI’s may be more likely than overweight, sedentary individuals to choose neighborhoods that support their pre-existing healthy lifestyle. This may, in turn, affect the association between neighborhood qualities and T2D-GD.

If non-random selection into neighborhoods exists, then the observed association between individual BMI/T2D-GD and neighborhood features arises from two sources: (a) physical or socio-cultural features of the built environment give rise to variation in individual BMI/T2D-GD (i.e., a causal mechanism), and (b) unobserved characteristics that affect both residential location and individual BMI/T2D-GD (i.e., a non-random selection mechanism). If non-random selection occurs, estimates that do not correct for its effects will misstate the strength of the causal
relationship.

In this analysis, our aim is to characterize the relative contributions of the causal and selection explanations with non-experimental data by testing two hypotheses:

1. Walkable neighborhoods, as measured by population density, pedestrian-friendly design, land-use diversity (e.g., mixed use, proximity to open space, access to grocery stores), and lower levels of neighborhood-level BMI are associated with lower levels of individual BMI, healthier levels of pregnancy-related weight gain, and lower prevalence rates of T2D-GD, ceteris paribus.

2. Individuals moving from lower BMI neighborhoods to higher BMI neighborhoods will be a non-random subset (i.e., these movers will have higher BMI values, more unhealthy weight gain, and will be more likely to be diabetic [even before they make the move], all relative to stayers) of those living in the lower BMI neighborhoods; conversely, those moving from higher to lower BMI neighborhoods will have lower BMI values, less pregnancy-related weight gain and will be less likely to be diabetic [again, even before they make the move], in relation to stayers.

This study relies on a world-class population-based data source, the Utah Population Database (UPDB). The vast genealogical records in the UPDB are linked to state-wide vital records (pre-pregnancy weights on birth certificates) that contain longitudinal data on height and weight (used to construct BMI), pre-existing diabetes, gestational diabetes and residential location for individuals and their kin. The UPDB is also linked to U.S. Census, state, and county
information on neighborhood characteristics using Geographic Information Systems (GIS) databases. The UPDB represents a unique and comprehensive database with which to address neighborhood effects on BMI and T2D-GD; it encompasses the adult/adolescent population and represents the full range of neighborhood settings.

Literature Review

Over 65% of U.S. adults are considered overweight (BMI ≥ 25) and 31% are obese (BMI ≥ 30)\(^7\). Adult obesity is associated with shortened life expectancies and an excess risk of chronic diseases such as T2D\(^5\), heart disease\(^8,9\), osteoarthritis\(^10\), and some form of cancers\(^11\), as well as social stigmas and substantial economic costs\(^7\). Childhood overweight levels have tripled since the 1970’s\(^12\) and obese youth born in the year 2000 are estimated to face a 30-40% chance of becoming diabetic\(^12\).

Given rapid increases in obesity, researchers have begun to emphasize how obesogenic environments may account for this trend. In examining the relative influence of causal and selection forces on obesity patterns, our study addresses two of the four themes for research outlined in the Strategic Plan for NIH Obesity Research\(^7\). First, we propose to examine the role of the physical environment in supporting healthy physical activities and eating choices. Second, we address cross-cutting topics, such as the identification of at-risk groups using a multi-disciplinary approach.

Past research has found relationships among walkable neighborhood designs, support for physical activity and healthy eating, and obesity or overweight\(^13-15\). Walkable neighborhoods
are those designed to include the “3Ds”: population Density, Diversity of destinations, and pedestrian friendly Designs. High densities and diverse land uses together mean that many people are within walking distance of multiple desirable destinations. Well-connected streets, a measure of pedestrian friendly design, further support walking by allowing walking trips to be relatively short, direct, and convenient. Research has demonstrated that greater density neighborhoods have lower BMIs. Density, although not always associated with lower BMI, provides a critical mass of individuals that may encourage the development of walking destinations and may discourage exclusive reliance on cars. More pedestrian friendly street connectivity or accessible and high quality sidewalks have been associated with fewer weight problems (except see ). Indicators of diverse and walkable destinations in a neighborhood are associated with lower weight.

Many studies depend upon the Behavioral Risk Factor Surveillance System (BRFSS) or other national surveys. Although such surveys are useful, they do not provide extensive response rates in any one neighborhood. We propose to utilize birth certificate databases (using pre-pregnancy weights and reports of pre-existing diabetes/gestational diabetes) because they provide extensive coverage of neighborhoods (i.e., very large numbers per neighborhood). If birth certificate databases prove useful in the present study, as we anticipate given our preliminary analyses, the results could encourage researchers from other states to consider using similar databases. Extensive local databases on the obesity and diabetes problem may prove most relevant to policy makers and other local and state officials who will be needed as partners for neighborhood-based obesity prevention efforts.
All neighborhood studies of obesity/T2D-GD and the environment are vulnerable to the selection threat to internal validity. To date, most studies of obesity and neighborhood environments have assumed that no selection effects exist. Yet, it is very likely that neighborhood characteristics are not exogenous with respect to an individual’s BMI or risk of T2D-GD. Rather, unmeasured factors that influence an individual’s choice of residential location may also influence that individual’s energy balance (e.g., preferences for physical activity or types of eating establishments). Researchers must be able to answer the question: To what extent do residents with unhealthy (healthy) behavior patterns self select into unhealthy (healthy) neighborhoods? Efforts to redesign neighborhood environments will be misdirected if residents with unhealthy behaviors simply choose unhealthy neighborhoods. In order for neighborhood policy and design interventions to succeed, it is important to determine if neighborhood environments have an independent effect on the health behaviors of their residents.

Recent reviews from sociology, public health, epidemiology, and planning have summarized a variety of strategies to address selection threats in community studies\textsuperscript{30-33}; These include statistical, sampling, and research design techniques to deal with the selection problem.\textbf{Drawing from these approaches, we consider here the mover-stayer model as a way to examine overweight/obesity and risk of T2D-GD in relationship to neighborhood environments.}

Cross-sectional neighborhood research has identified many neighborhood correlates of obesity, such as collective efficacy\textsuperscript{34}, safety\textsuperscript{35}, and socioeconomic status\textsuperscript{36}. These studies are limited by the possibility that residents self-select into neighborhoods for reasons related to
obesity. Longitudinal studies may overcome these limitations by allowing researchers to assess how residents sort into neighborhoods. Past research shows that movers who relocate to more walkable neighborhoods reported fewer vehicle miles traveled and more walking than in their former neighborhoods; these studies did not examine obesity. However, higher BMI individuals tend to move to more sprawling, less pedestrian friendly neighborhoods and individuals who move to denser neighborhoods tend to lose weight.

Studies of migrants often have not considered BMI-related factors. Instead, life cycle factors, such as household size, resident ages, and socioeconomic status, have been identified as causes for moving. Recent CPS data indicated that residents move for reasons related to housing, family, and work. Few of these reasons directly relate to neighborhood food or physical activity qualities that might be related to BMI. In addition, only a few studies have assessed preferences for walkable neighborhoods directly; these studies showed that 33% to 49% of respondents might prefer more walkable neighborhoods.

Reasons for relocation also might indirectly relate to BMI. For example, moves for changes in family size might relate to less physical activity among new mothers. Moves for preferred schools may bring other physical activity amenities. Moves for job changes might involve changes in work and/or physical activity time allocations; job changes that require longer commutes may also be associated with higher BMI. Moves to higher density apartment living have been found to relate to a desire for accessibility and nearby recreational space, as well as work and accessibility reasons, which might predict lower BMI. To the extent possible, it is important to control for these changes that accompany residential relocation.
In this analysis, a mover-stayer analysis will be used to assess selection effects. Mover-stayer models have been used successfully\textsuperscript{47,48} to compare individuals and places between three groups: those who remain in the place of origin, those already living in the place of destination, and migrants moving between the two places. This permits an assessment of differential selection into and out of specific types of neighborhoods.

Data

\textit{Utah Birth Certificates}

A complete set of Utah birth certificates from 1947-2008 have been linked to the UPDB. Here we use a subset from 1994-2008 for reasons described below. The birth certificates contain information on complications, risks, abnormalities, method of delivery, birth weight, gestation, and number of previous live births and stillbirths to the mother. Using the parent-child information contained in these certificates, the UPDB links these records into parent-child dyads and sibships. This information allows the UPDB to be updated in terms of maternal and paternal reproductive histories and to identify a variety of kin in the UPDB. The PPR staff link these data with genealogy records. Birth certificates that do not merge into existing genealogy families have been linked together to create two and three generation families.

Starting in 1989, Utah birth certificates contain data on \textit{pre-pregnancy weight} as well as \textit{weight gain} associated with a given birth. They also contain data on the presence of maternal \textit{pre-existing diabetes} (which is unspecified on the birth certificate but is largely dominated [90-95\%] by type II diabetes\textsuperscript{49}) and \textit{gestational diabetes} associated with that specific birth. These
certificates also provide residence information at the time of each birth. Accordingly, for all women who bore children in 1989 or later, the UPDB contains longitudinal data on their residential location and pre-pregnancy weight. We find that there are 225,000 women who gave birth in Utah from 1989-2008. However, women with two or more children (N=135,300) born in Utah are the most informative because they provide longitudinal BMI and location data as well as additional information on T2D-GD.

The use of linked birth certificates in Utah to assess the association between BMI/ T2D-GD and neighborhood qualities (for causal or selection reasons) offers distinct advantages. First, Utah has a fertility rate that is higher than the national average and thus there are large numbers of women with two or more children. This provides data for initial BMI (at first birth) and neighborhood and subsequent changes for both variables. Second, birth certificates provide an extensive body of data captured by the UPDB including educational level of parents, race/ethnicity, and health conditions. Third, parents’ decisions about residential location are often driven by neighborhood, life-style and schooling considerations. The choices parents make because of child-based factors may be a driving force in affecting location decisions that in turn affect maternal BMI. Fourth, if these data are found to be useful in assessing the association between BMI and neighborhood quality, it is feasible to export the approach to other states. Finally, the analysis plan is particularly cost-effective because the record linkage that created a birth certificate history for each mother has already been done but unanalyzed in the way proposed here.
Use of birth certificate data has two potential limitations – both of which are addressable. First, and most obvious, the sample is restricted to mothers of reproductive age. Second, bearing children may create a weight gain profile that will obfuscate the association between neighborhood characteristics and maternal BMI. While we do not have the same depth of information on men nor women with no children, we are able to compare mothers with each other in terms of neighborhood and familial characteristics so that these potential biases shortcomings will be netted out. Moreover, weight gain during pregnancy itself is of interest since there may be patterning in unhealthy weight gain by neighborhood.

Environmental Data

To test our hypotheses, we have assembled an extensive data set on neighborhood environments in Salt Lake County, including measures of population Density, land-use Diversity, and pedestrian-friendly Design (i.e., the 3D’s) measured in the 2000 U.S. Census. During the grant period we will link this data set to U.S. Census data for 1990 and 2000 Census data from the rest of the state of Utah. For land use diversity we already have two census-based diversity measures (proportion of workers who walk to work and median age of housing in the neighborhood) that relate well to BMI. Other researchers have used mixed land uses from parcel based land use typologies; we have assembled parcel data from the Salt Lake County Surveyor’s Office that could be used to test other diversity measures. We also have street connectivity measures of pedestrian friendly designs. We have measures of pedestrian friendly design in the form of intersection density from road networks (data for 1985, 1997, and 2000
from Utah’s Automated Geographic Reference Center). Public transportation data including light rail transit and bus systems have been obtained from Utah Transit Authority (UTA); these data are not available for past decades.

Population health data stratified by socio-cultural position are critical for monitoring and analyzing health issues. In our past research we have found it useful to include census block group variables that tap aspects of the racial/ethnic composition of the neighborhood, specifically, the proportion of Hispanic, African-American, and Hawaiian/Pacific Islander populations in the block group. In addition, we have used Census data on the median family income and the median age of the block group 50.

Mover-Stayer Analysis

By moving between neighborhoods with a wide range of characteristics, migrants provide an excellent opportunity to examine the effects of selection on understanding variation in BMI. Numerous studies exist that assess how neighborhoods relate to diet, exercise, and obesity but they have generally excluded any examination of the neighborhoods from which people originated. For migrants, BMI levels may reflect their origin or destination neighborhood or they may reflect a select group with traits that increase or decrease the chances that they will be obese, irrespective of their origin or destination.

In this paper, we adopt the following strategy to compare movers and stayers from different neighborhoods. First, we rank all block groups in Salt Lake County in terms of their mean adult (18-64) BMI based on all driver license data. These data contain height and weight
information mothers in quartiles based on block group for the county (and state). These data are preferable for the purposes of characterizing block group BMI since they include all women (not just mothers) and men. We then rank these block groups from leanest to heaviest and categorize them into quartiles. Movers are those who migrate within their own quartile or move to another quartile between the two births. Accordingly, stayers are those who remain at the same address for both births. Here we consider only mothers where both births are in the county.

To assess selection, we assess whether individuals moving from lower BMI neighborhoods to higher BMI neighborhoods will have higher BMI values, more unhealthy weight gain, and will be more likely to be diabetic even before they make the move, all relative to stayers. In a similar fashion, we also examine whether those moving from higher to lower BMI neighborhoods will have lower BMI values, less pregnancy-related weight gain and will be less likely to be diabetic again, even before they make the move, in relation to stayers.

Results

Our preliminary analysis focuses on new mothers in Salt Lake County between 1989 and 2007. In order to include the largest portion of the data, we analyze for this analysis pre-pregnancy weights for the first and second births for approximately 30,000 women who satisfy the inclusion criteria.

Table 1 provides descriptive statistics for the sample considered here. The two broad groups, “mover” and “stayers” are defined as women who move or stay within the boundaries of Salt Lake County at the time of their two births. All block-groups in Salt Lake County were
categorized in terms of the average pre-pregnancy BMI for all births for all women for all the years covered in the data. At this stage we have placed each block group into one of four BMI quartiles: Q1 (lowest/leanest), Q2, Q3 and Q4 (highest/most overweight). If someone moves but remains within a BMI quartile then they are treated as movers but simply moving from, say, Q1 to Q1. Of course, women who remain at the same residence between the two births are stayers. Note that in Table 1 that movers (regardless of where they started and where they arrived) generally have lower BMI values, gain more weight as a percentage, and have longer birth intervals.

Figures 1 shows maps for block groups in Salt Lake County, for mean block-group pre-pregnancy BMI where each block-group is categorized into BMI quartiles. The top panel describes the distribution of pre-pregnancy BMI for first births while the lower panel show the comparable map for second births. Note that the western half of the county shows much higher BMI levels than the eastern half, partly reflecting a sociodemographic divide with the eastern portion having generally higher more walkable neighborhoods and higher SES values.

Figure 1 here

Figure 2 shows that women in the leanest quartile (Q1) who will move to the heaviest quartile (Q4) --- **but have not yet made the move** --- are among the heaviest residents in Q1 based on their pre-pregnancy weight before their first pregnancy. Similarly, women in the heaviest quartile for their first birth (Q4) who will move to the leanest quartile (Q1) --- **but have not yet made the move** --- are among the leanest members of the in Q4. *This suggests that there*
is non-random selection into different types of neighborhoods as detected by BMI differences in pre-pregnancy weights.

In Table 2 and 3 we again consider movers moving from Q1 to Q4 as well as movers from Q4 to Q1 except we now examine diabetes and gestational diabetes. With respect to diabetes (Table 2), we do not find evidence supporting the presence of non-random selection. For gestational diabetes (Table 3), there is suggestive evidence (p≈0.07) showing that persons starting in a lean neighborhood but who move to a heavier neighborhood have a higher chance of experiencing gestational diabetes. A comparable result holds for women moving from a heavy neighborhood to a lean neighborhood.

Conclusion

This paper has shown that non-random residential selection exists based on the patterning of BMI for women who will move (but have not yet) across a range of neighborhood types. We extend this analysis to consider how a well-demonstrated consequence of elevated BMI, diabetes, is also affected by these selective forces (at least for gestational diabetes). The full paper will supplement the diagnosis of diabetes from the birth certificates to also include medical records.
References


Table 1. Descriptive Statistics for Pre-Pregnancy Weights prior to First and Second Births Among Movers and Stayers in Salt Lake County, Utah, 1989-2007

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mover</th>
<th></th>
<th></th>
<th>Stayer</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>Std Dev</td>
<td>N</td>
<td>Mean</td>
<td>Std Dev</td>
</tr>
<tr>
<td>Pre-Pregnancy BMI: 1st Pregnancy</td>
<td>19299</td>
<td>22.860</td>
<td>4.521</td>
<td>10380</td>
<td>23.320</td>
<td>4.742</td>
</tr>
<tr>
<td>Pre-Pregnancy BMI: 2nd Pregnancy</td>
<td>19299</td>
<td>24.383</td>
<td>5.506</td>
<td>10380</td>
<td>24.316</td>
<td>5.391</td>
</tr>
<tr>
<td>Pre-Pregnancy BMI Difference Between 1st &amp; 2nd Pregnancy</td>
<td>19299</td>
<td>1.523</td>
<td>3.214</td>
<td>10380</td>
<td>0.996</td>
<td>2.699</td>
</tr>
<tr>
<td>Percent Change in Pre-Pregnancy BMI</td>
<td>19299</td>
<td>6.895</td>
<td>13.605</td>
<td>10380</td>
<td>4.453</td>
<td>11.074</td>
</tr>
<tr>
<td>Change in Pre-Pregnancy Weight (lbs) Between 1st &amp; 2nd Pregnancy</td>
<td>19299</td>
<td>9.130</td>
<td>18.622</td>
<td>10380</td>
<td>6.019</td>
<td>15.903</td>
</tr>
<tr>
<td>Percent change in Pre-Pregnancy Weight</td>
<td>19299</td>
<td>6.869</td>
<td>12.990</td>
<td>10380</td>
<td>4.420</td>
<td>10.623</td>
</tr>
<tr>
<td>Number of Months Between 1st &amp; 2nd Pregnancies</td>
<td>19299</td>
<td>36.956</td>
<td>20.572</td>
<td>10380</td>
<td>29.259</td>
<td>13.400</td>
</tr>
</tbody>
</table>
Figure 1. Spatial Variation in Pre-Pregnancy Weights for Salt Lake County Utah
Figure 2. Pre-pregnancy BMI by mover status for BMI before the first birth (these are differences in BMI BEFORE a woman makes the move that we know she will make)
Table 2. Test for Trend: Effects on Having Diabetes Before the Second Pregnancy based on Moving to a Heavier Neighborhood (if Starting in the leanest Neighborhood) or on Moving to a Leaner Neighborhood (if starting in the heaviest Neighborhood).

<table>
<thead>
<tr>
<th>Preexisting Diabetes</th>
<th>First Pregnancy Q1 (Leanest Neighborhood)</th>
<th>First Pregnancy Q4 (Heaviest Neighborhood)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>p</td>
<td>Point Estimate</td>
</tr>
<tr>
<td>Intercept</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Age 1st Birth</td>
<td>0.1638</td>
<td>1.08</td>
</tr>
<tr>
<td>Months Between 1st and 2nd Birth</td>
<td>0.1815</td>
<td>1.011</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.3796</td>
<td>1.833</td>
</tr>
<tr>
<td>Other Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>0.3811</td>
<td>2.497</td>
</tr>
<tr>
<td>Less than a High School Degree</td>
<td>0.8861</td>
<td>0.846</td>
</tr>
<tr>
<td>Some College</td>
<td>0.5685</td>
<td>1.418</td>
</tr>
<tr>
<td>College Graduate</td>
<td>0.3567</td>
<td>0.496</td>
</tr>
<tr>
<td>Test for Trend</td>
<td>0.1209</td>
<td>1.389</td>
</tr>
</tbody>
</table>
Table 3. Test for Trend: Effects on Having Gestational Diabetes During the Second Pregnancy based on Moving to a Heavier Neighborhood (if Starting in the leanest Neighborhood) or on Moving to a Leaner Neighborhood (if starting in the heaviest Neighborhood).

<table>
<thead>
<tr>
<th>Gestational Diabetes</th>
<th>First Pregnancy Q1 (Leanest Neighborhood)</th>
<th></th>
<th>First Pregnancy Q4 (Heaviest Neighborhood)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>p</td>
<td>Point Estimate</td>
<td>95% Wald Confidence Limits</td>
<td>p</td>
</tr>
<tr>
<td>Intercept</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Age 1st Birth</td>
<td>&lt;.0001</td>
<td>1.116</td>
<td>1.058</td>
<td>1.178</td>
</tr>
<tr>
<td>Months Between 1st and 2nd Birth</td>
<td>0.0006</td>
<td>1.015</td>
<td>1.006</td>
<td>1.023</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.1191</td>
<td>1.771</td>
<td>0.863</td>
<td>3.635</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>0.0587</td>
<td>2.744</td>
<td>0.964</td>
<td>7.812</td>
</tr>
<tr>
<td>Other Race</td>
<td></td>
<td>0.612</td>
<td>1.463</td>
<td>0.336</td>
</tr>
<tr>
<td>Less than a High School Degree</td>
<td>0.8326</td>
<td>0.899</td>
<td>0.335</td>
<td>2.411</td>
</tr>
<tr>
<td>Some College</td>
<td>0.0005</td>
<td>0.304</td>
<td>0.155</td>
<td>0.595</td>
</tr>
<tr>
<td>College Graduate</td>
<td>0.0023</td>
<td>0.363</td>
<td>0.189</td>
<td>0.697</td>
</tr>
<tr>
<td>Test for Trend</td>
<td>0.0772</td>
<td>1.227</td>
<td>0.978</td>
<td>1.539</td>
</tr>
</tbody>
</table>