THE EFFECTS OF SOCIO-ECONOMIC AND TRANSPORTATION ACCESSIBILITY ON AREA-LEVEL DIABETES COUNTS: A LATENT-VARIABLE STRUCTURAL EQUATION MODEL APPROACH

Yiyi Wang, Ph.D.
214 Cobleigh Hall
Civil Engineering Department-Western Transportation Institute
Montana State University
Bozeman, MT 59717
Email: yiyi.wang@ce.montana.edu
Voice: 406-994-6051
Fax: 406-994-6105

To submit for presentation at the Transportation Research Board Annual Meeting and for publication at the Transportation Research Record

(2 Figures + 3 Tables) × 250 words each + 5374 words = 6624 word counts

Submit date: August 1, 2014
ABSTRACT
Land use and transportation can affect public health via the pathways of neighborhood accessibility and physical activity levels. Health issues, like diabetes, represent a complex interplay of nutrition, genetics, and environment. This work aimed to test if and how greater transportation accessibility (measured by higher density, mixed land-use, and proximity to bike lanes) is associated with diabetes risk. This work also assessed if residential self-selection plays a significant role in deciphering this relationship. This work uses the diabetes count data at the traffic analysis zone (TAZ) level and a handful of accessibility and socio-economic variables (e.g., proximity to bike lanes, land-use balance, population density, land value, etc.) in Travis County, Texas.

Results suggest that accessibility and socio-economic factors play an important role in understanding diabetes prevalence. On average, forty-eight percent of multi-family dwelling units are within a half mile distance from bike lanes in the county. Increasing housing that is within a half mile of bike lanes at this average level could reduce diabetes rates by 0.7 percent. This would then translate into a reduction of 70 diabetic adults for a TAZ with a 1,000 population. For a one percent rise in land-use balance, diabetes risk is estimated to decrease by 0.28 percent. Outputs also suggest that, to curb diabetes prevalence, emphasis should be placed not only on enhancing accessibility, but also on narrowing the health disparities perhaps by improving access to quality health care and raising awareness of healthful lifestyles for people with lower socio-economic status.

Keywords: latent variables, structural equation model, accessibility, active transportation, diabetes, socio-economic
**INTRODUCTION**

Active transportation (e.g., biking and walking) plays a pivotal role in achieving sustainable cities and improving health. A sedentary lifestyle was estimated to cause as many as 300,000 deaths annually in the United States (1) and cost $51.6 billion per year in direct medical expenses, of which $32.4 billion was spent treating type 2 diabetes—an illness strongly associated with obesity (and physical inactivity) (2). Ample evidence confirmed the effects of physical activity on the reduction of diabetes, cardiovascular disease, and certain types of cancer (3). This public health benefit even outweighs the negative externalities associated with traffic crashes involving non-motorized travelers like pedestrians and cyclists (3, 4, 5).

Nonetheless, 60 to 80 percent of the world’s population does not meet the physical activity levels recommended by the World Health Organization, and even people in industrialized countries lead lifestyles with activity levels that do not promote good health (6). A contributing factor to this inactivity trend is rampant urban sprawl and a historical emphasis on motor vehicle traffic in making engineering and planning decisions. But even as more attention is placed on active transportation, important questions remain: if and how does the built environment (BE) affect health through the pathway of physical activity, and how can the health component be operationalized in the transportation decision-making processes to build more healthful cities?

The linkage between health and BE has been recently studied in several Health Impact Assessment (HIA) projects. A comprehensive review of HIA studies is found in a report by the Florida Department of Health (7). HIAs have been used to evaluate the health impact of the Duwamish River Cleanup on local residents in the state of Washington (8), the health impact of oil and gas exploration in Alaska (9), and the health impact on vulnerable populations of transportation projects in Massachusetts (10). HIAs have also been widely used by public health professionals in various promotional studies (e.g., tobacco regulations and HIV/AIDS counseling programs). Insightful and translational in their own right, these HIA studies focused on narrative descriptions of the health status and qualitative analysis without identifying tangible ties, if any, between transportation-land use and public health.

A research (11) investigated the impacts of urban sprawl on the body mass index (BMI) of youth in the United States using cross-sectional and longitudinal models. Urban sprawl was measured as a composite of residential density and street accessibility at the county level through principal component analysis. They concluded that urban sprawl is a significant predictor of obesity in U.S. youth in the cross-section analysis but found no such relationship in the longitudinal analysis. They also acknowledged sample sizes, unobserved effects (like self-selection), and missing variables may explain the lack of significant relationship in the longitudinal study. A study (12) concluded that high- and low-income individuals differ more in their motivation to eat healthy food than in their ability to access and afford healthy food. Another study (13) identified low-income women as the priority group to receive more information about H1N1 vaccination because “these individuals tend to have the least knowledge and the least acceptance of vaccinations”.

Residential self-selection is a situation where people choose to live in a neighborhood based on their travel “abilities and preference” (14). This type of self-selection has played an important
role in understanding the relationship between the built environment and travel behaviors including weekly miles driven (15), rail commuting (16), and vehicle ownership (17).

As is the case with these behavioral studies, self-selection may play a similar role in gauging the true effects of BE (e.g., neighborhood accessibility) on health outcomes. For example, people with higher income (and presumably better access to quality health care) may choose to live in more accessible neighborhoods that could be characterized by higher density, mixed land use, and proximity to biking/walking opportunities. As a result, the spurious correlation between neighborhood attributes and health outcomes may merely be a result of a common factor (income). Without accounting for self-selection, statistical results would yield biased estimates. A paper (18) used a structural equations modeling approach to estimate the effects of residential neighborhood type on miles traveled by transportation mode while controlling for variables related with socio-demographic and attitudes. Recognizing the sizable impacts of attitudes and lifestyle on travel behavior (18), another research study (19) introduced and investigated the concept of residential neighborhood dissonance—mismatch between a person’s attitude/lifestyle and the characteristics of the neighborhood she resides in using a standard Tobit model. They found that neighborhood dissonance is a significant predictor of vehicle miles traveled and miles traveled via walking and biking.

The goal of this paper is to test if and how greater transportation accessibility affects diabetes rates and if residential self-selection plays a significant role in deciphering this relationship. The analysis is cast in a structural equation model (SEM) paradigm with latent variables and uses diabetes estimates, built environment, and socio-economic data for Travis County, Texas. The rest of the paper is structured into Data Sets—summarizing the data used and variables computed, Methodology—presenting the SEM specification with latent variables, Analysis and Results—discussing estimation and inference of model results, and Conclusions.

DATA SETS
Diabetes prevalence data were obtained from the Behavioral Risk Factor Surveillance System\footnote{The BRFSS does a telephone survey that randomly draws adults (18 years and older) owning a landline or a cell phone and collects data on lifestyle risk factors for chronic diseases and associated mortality (Wikipedia). The weighted data provides information about the number of adults who are diagnosed by a doctor as having diabetes in Travis County, Texas.} survey (BRFSS) for Travis County, Texas. To ensure reliable sub-group analysis, BRFSS data are aggregated over 10 grouped zip code tabulation areas (ZCTAs) and over two years, 2010 (n =1570) and 2012 (n =1524). Grouping helps to ensure the relative standard error (RSE) is no greater than the 30 percent and a minimum of 50 respondents from the unweighted sample. Fine-grained and reliable diabetes counts are crucial to investigating if and how neighborhood attributes affect diabetes prevalence. Respondents’ home locations are unavailable in the interest of confidentiality. Consequently, the grouped-ZCTA level diabetes rates, as shown in Figure 1, were used to impute diabetes counts across neighborhoods, as defined by Traffic Analysis Zones (TAZs) for the county. Imputation assumes an even spatial distribution of diabetes rates within each grouped ZCTA and proportionality to population counts. ZCTA-level population counts were obtained from the Census 2012 American Community Survey, while the TAZ-level population counts come from the City of Austin Metropolitan Planning Organization’s

1 The BRFSS does a telephone survey that randomly draws adults (18 years and older) owning a landline or a cell phone and collects data on lifestyle risk factors for chronic diseases and associated mortality (Wikipedia). The weighted data provides information about the number of adults who are diagnosed by a doctor as having diabetes in Travis County, Texas.
Wang

(CAMPO’s) 2010 demographic data. The proportionality assumption implies that diabetes rates among populations within the same ZCTA remain homogenous spatially.

FIGURE 1 Mean Estimates and Coefficient of Variation (CV) of Diabetes Rates for Grouped ZCTAs (overlaid onto street networks) in Austin, Texas
Notes: CV= coefficient of variation, and measures the standard deviation normalized by the mean.

As seen from Figure 1, the eastern part of the county bore a higher diabetes share during the 2010–2012 period at 8.51 percent (darker blue), whereas the western part enjoyed less than half the prevalence rate at 3.77 percent (lighter blue) — a rate that is, in fact, 61 percent lower than the state average. On the surface, a linkage may exist between road network and land use and diabetes, as the western part of the county, which enjoys lower diabetes rates, is populated with a denser network of local streets (green lines). Coefficient of variation (CV) is computed as the ratio of the standard error to the mean of diabetes rates and measures the variability of the (estimated) rates in relation to their mean for each grouped ZCTA. The relative variability of diabetes rates is greater (larger yellow dots) in the western part than in the eastern part.

The diabetes counts combine two types of diabetes—type I, or juvenile diabetes, constituting approximately 5 percent of the diabetic population, and type II accounting for 90 to 95 percent of all diabetes cases. A general consensus is that type I diabetes is mostly genetic and type II diabetes is strongly affected by environmental factors. Since the majority of diabetes is type II, an illness highly related with physical activity (and hence built environment), it is legitimate to use the combined diabetes counts.
The distinction of this paper from epidemiological studies lies in its focus on establishing linkage, if any, between the built environment (specifically transportation accessibility) and diabetes risk after teasing out self-selection bias and confounding variables, rather than investigating individual-level risk factors (e.g., family history and nutrition). This work defines transportation accessibility as a composite measure for land-use balance, proximity to walking/biking facilities, and density of road networks. Such a study can help transportation practitioners better understand the impacts of planning decisions on public health and provide guidance for building a more healthful urban form.

Table 1 presents the summary statistics of the response variable (number of diabetic people at the TAZ level) and covariates related with transportation accessibility and socio-economic factors. Accessibility can be measured by population density; land-use entropy; job-population ratio; density of sidewalk (sidewalk lengths per square mile of land); vehicle-miles traveled (VMT) density; road lane-miles density; shares of single- and multi-family parcels within a half mile of bike lanes; shares of single- and multi-family parcels within a half mile of parks. Land entropy is a dimensionless indicator of land use balance, with a maximum value of 1 representing a balanced situation where land is equally distributed among all land use types and a minimum value of 0 indicating a situation where a single land use dominates the area. Socio-economic status is measured through median household income and average market value of land parcels at the TAZ level. Data processing is discussed in the following texts.

Vehicle-miles traveled (VMT) are computed as the product of segment length (in miles) and total capacity (in vehicle/hour) for each roadway link based on CAMPO’s 2010 network data. The VMT is then divided by the land area to obtain VMT density. Access roads are jointly determined by the CAMPO map, which contains collectors and a fraction of local streets, and Census Tiger/Line®’s extensive road networks, which were used to identify local streets that are missed by the CAMPO map. High-speed roads are defined, in this paper, as those that primarily serve fast vehicular flows with no or infrequent access points for pedestrians and cyclists, including interstates, freeways, expressways, principal arterials, minor arterials, and associated facilities like frontage roads and ramps. To account for size effect, all these variables are normalized by the land area of each TAZ in square miles.

Land-use entropy is computed on the basis of four land use categories: single-family residential; multi-family residential; commercial/service/office/industrial/government; and parks (and public space). Commercial, service, governmental, office, and industrial were collapsed into a single category to represent commercial and employment destinations. Parks and public space represent opportunities for exercise and physical activities. Single-family use was distinguished from multi-family use to capture the effects of different residential density. Open space is not considered because vacant parcels generally do not attract work, recreation, or shopping trips and hence interact little with other land-use types. Land-use entropy is formulated as: $\sum_{j=1}^{4} p_j \ln p_j$ (20), where $p_j$ denotes the percentage of land areas for category $j$. Job population balance is proxied by the ratio between number of jobs (from CAMPO demographic estimates for 2010) and the number of resident workers (from American Community Survey 2010). Opportunities to biking are captured by the percentage of homes (i.e., single-family and multi-family) within a half mile of bike lanes, using ArcGIS’s spatial join and buffer functions.
Median household income is obtained from the 2012 release of American Community Survey’s 5-year estimates. Market land value is weighted by area for each TAZ, and the parcel-level land value data come from the Travis County Appraisal District’s appraisal archive. Income and land value variables are used to proxy for the unobserved socio-economic factor.

### TABLE 1 Summary Statistics of Response and Predictors (No. of Obs. = 606)

<table>
<thead>
<tr>
<th>Population Density (persons per sq. mile)</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land-Use Entropy</td>
<td>0.490</td>
<td>0.250</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Job-Population Ratio</td>
<td>30.79</td>
<td>198.2</td>
<td>0.00</td>
<td>3132.00</td>
</tr>
<tr>
<td>Sidewalk Density</td>
<td>18.38</td>
<td>15.93</td>
<td>0.00</td>
<td>58.54</td>
</tr>
<tr>
<td>VMT Density</td>
<td>1.26E+05</td>
<td>1.59E+05</td>
<td>0.00</td>
<td>1.01E+06</td>
</tr>
<tr>
<td>Access Road Density</td>
<td>5.99E+04</td>
<td>4.29E+04</td>
<td>0.00</td>
<td>2.06E+05</td>
</tr>
<tr>
<td>Multi-Family w/in ½ mile Bike%</td>
<td>0.48</td>
<td>0.44</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Single-Family w/in ½ mile Bike%</td>
<td>0.44</td>
<td>0.48</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Multi-Family w/in ½ mile Parks%</td>
<td>0.02</td>
<td>0.04</td>
<td>0.00</td>
<td>0.36</td>
</tr>
<tr>
<td>Single-Family w/in ½ mile Parks%</td>
<td>0.18</td>
<td>0.31</td>
<td>0.00</td>
<td>2.84</td>
</tr>
<tr>
<td>Land Market Value ($1,000 per sq. mile)</td>
<td>2,297</td>
<td>4,885</td>
<td>0.00</td>
<td>5.14E+04</td>
</tr>
<tr>
<td>Median Household Income ($1,000)</td>
<td>70.23</td>
<td>42.80</td>
<td>0.00</td>
<td>233.13</td>
</tr>
<tr>
<td>Diabetes Counts (Response Variable)</td>
<td>119.95</td>
<td>182.51</td>
<td>0.00</td>
<td>1291.00</td>
</tr>
</tbody>
</table>

Model input data was checked to ensure that multicollinearity did not have a detrimental effect on model estimation and prediction. Multicollinearity refers to the severe linear correlation between two or more predictors (i.e., covariates). It means one covariate can be explained by another covariate in a linear function with exact accuracy (21). Multicollinearity can result in unstable estimates of the coefficients such as erratic changes in the estimate when a different initial value is used and exceedingly large standard error of the estimates. It has detrimental impact on the estimation and interpretation of individual coefficients, but not on the overall goodness-of-fit and predictive power of the model at least within the sample. A correlation diagram (omitted here) is used to detect any severe collinearity among the predictors. High-speed road density and vehicle-miles-traveled density are highly correlated, so are sidewalk density and access-road density. Hence, only one covariate (from each pair) is chosen to enter the model.

### METHODOLOGY

The main aims of the paper are to test if and how greater transportation accessibility (measured by higher density, mixed-use, and proximity to bike lanes) is associated with lower diabetes risk; and if residential self-selection plays a significant role in deciphering the impacts of neighborhood accessibility on health outcomes.

A challenge in quantifying the effect of a neighborhood’s built environment on health pertains to bias due to residential self-selection—a situation where people choose their home locations based on their socio-economic abilities and preference (or attitude). Often these abilities and
preference are inadequately observed. For example, preference or attitude-related attributes are
generally unavailable in many empirical data sets, while socio-economic factors are roughly
represented by income, ethnicity, and education variables. It is impossible to capture all these
variables. Hence, a seemingly positive effect of accessible neighborhood on good health may, in
fact, arise due to a common driving factor like household income, which determines the quality
of neighborhoods and the quality of health care people can afford.

A thorough review (22) discusses generic methods to counteract self-selection for studying travel
behavior, such as direct questioning of attitude, instrument variable (although the selection of
instrument is infamously tricky and only offers a partial answer [23]), sample selection models,
longitudinal studies, and joint models (e.g., structural equation models [SEMs]).

All methods can be applied to better understanding the effects of transportation accessibility on
health. However, the choice of methods depends on the circumstances of each study, including
data availability (e.g., availability of direct attitude information), the observation unit (e.g.,
knowing residential choice at the individual level facilitates the use of sample selection models),
and temporal scale of the data (e.g., longitudinal design offers advantages in purging attitude and
other unobserved factors that tend to remain stable over the years). The diabetes counts (i.e.,
number of diabetic residents) aggregated across TAZs, coupled with a host of TAZ-level
attributes of built environment and socio-economic information, determined that the SEM is the
most appropriate approach for purging self-selection bias in this study.

Latent variables (24) are used within the SEM specification to capture a multitude of factors
underlying diabetes prevalence, including socio-economic status (SES) and neighborhood
accessibility. This is motivated by the fact that accessibility and SES variables are often
imperfectly measured, through proxies that are available to the analysts and only describe some,
but not all, facets of the latent variables of interest. For example, land-use balance, proximity to
parks and bike lanes, and local street density indeed measure neighborhood accessibility.
However, they miss some specific aspects of accessibility such as topography (soil slope), bike-
lane quality, and safety (if the person feels safe to travel on foot and by bike). SES is proxied by
average market value of land parcels and median household income, on the premise that people
with similar socio-economic status may afford a similar quality of health care, have similar
knowledge of health, and lifestyles (12, 13). The error terms associated with the latent factors
measuring accessibility and SES, respectively, represent missing variables (e.g., nutrition,
lifestyle, and ethnicity for SES, and slope and safety for accessibility) and hence lend more
flexibility and behavioral realism to the model.

This work has three phases. The first phase describes diabetes risk via a Poisson-lognormal
process and assumes that diabetes counts result from a latent diabetes risk, \( f_1 \), and a
heterogeneity error term, \( \nu \):

\[
D \sim \text{Poisson} \left( Z \right) \text{ and } Z = \text{POP} \cdot \exp(f_3 + \nu) \tag{1}
\]

where:

- \( D \) is the observed diabetes count at each TAZ,
- \( Z \) denotes diabetes rates (after controlling for population [POP] as the exposure
term)
- $\exp(\cdot)$ denotes the exponential function and ensures non-negative diabetes rates, $f_3$
denotes the underlying (latent) diabetes risk factor
- $\nu$ captures a random term unique to each TAZ and is assumed to follow a normal
distribution with unknown variance terms: $N(0, \sigma^2)$

The second phase employs a SEM framework to (1) establish linkage between the latent diabetes
risk ($f_3$) and latent covariates including neighborhood accessibility ($f_1$) and SES ($f_2$), and (2) to
capture the relationship between SES and the accessibility of the neighborhood:

$$\begin{align*}
\begin{cases}
f_3 = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \epsilon \\
f_1 = \beta_4 f_2 + \delta
\end{cases}
\end{align*}$$

where unknown coefficients ($\beta_1$ and $\beta_2$) measure the signs and magnitude of the effects of
accessibility and SES on area-level diabetes risk, with $\beta_0$ as an intercept term. The coefficient $\beta_4$
explains, to some extent, the degree of residential self-selection.

The last phase contains a measurement model that describes the latent covariates through
observed variables (24). Expressed in thousands of dollars, average market value of land parcels
and median household income are used to measure SES.

Accessibility is proxied by a handful of BE factors including land entropy, job-population ratio,
vehicle-miles traveled (normalized by land area), density of local streets, percentage of single-
and multi-family units within a half mile of bike lanes, and population density. Sidewalk density
and shares of single- and multi-family units within a half mile of parks were initially included in
the measurement model for accessibility, but were removed due to statistically and practically
insignificant estimates regardless of models used (i.e., a negative binomial model, a spatial
conditional autoregressive model [CAR], the latent variable SEM model). To improve model
performance, these covariates are standardized so that the resulting distributions center on zero
with one standard deviation. Their histograms suggest that these distributions resemble a normal
distribution.

Model validation is achieved by comparing the results and performance among three different
modeling tools: the first is the SEM model with latent SES and accessibility factors, the second
is a SEM model without these latent factors, and the third is a spatial model (namely a
conditional autoregressive model [CAR]) that is widely used to study count responses (e.g.,
number of diabetic people, number of traffic crashes, etc.). Model No.1 has been detailed
previously. Model No.2 directly links diabetes rates to the observed income and neighborhood
accessibility variables. This model is still able to capture self-selection thanks to a structural
equation model that permits interactions between residential location (represented by population
density) and diabetes rates. Model No.3 captures spatial pattern in diabetes rates across TAZs.
This type of model is commonly used to study count data with geographic information, such as
crash counts along road networks (25) and cancer deaths across census tracts (24).

**ANALYSIS AND RESULTS**
The SEM Latent-Variable model is estimated using Bayesian Markov chain Monte Carlo
(MCMC) methods via WinBugs (26), with results summarized in Table 2. In this table, the t
statistics is used to evaluate statistical significance of a variable, and a value greater than +2.0 or
less than -2.0 indicates statistical significance at a confidence level of 95 percent. Elasticity is
Wang defined as the percentage change in the response variable for one percentage increase in a predictor. It is a dimensionless measure for practical significance. For example, the elasticity of land-use balance is -0.28, which means a one percent increase in land-use balance would be associated with a 0.28 percent decrease in diabetes risk.

### TABLE 2 Estimation Results of the Latent Variable SEM Model

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Node</th>
<th>Mean (t stats)</th>
<th>Elasticity</th>
<th>Variance Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SEM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant $\beta_0$</td>
<td>$\beta_0$</td>
<td>-3.124 ( -35.08)</td>
<td>—</td>
<td>Node $\tau_v$</td>
</tr>
<tr>
<td>Accessibility $\beta_1$</td>
<td>$\beta_1$</td>
<td>-1.233 ( -7.190)</td>
<td>—</td>
<td>$\tau_e2$</td>
</tr>
<tr>
<td>Attitude $\beta_2$</td>
<td>$\beta_2$</td>
<td>-0.838 ( -7.490)</td>
<td>—</td>
<td>$\tau_e1$</td>
</tr>
<tr>
<td>Self-selection $\beta_3$</td>
<td>$\beta_3$</td>
<td>0.642 ( 6.528)</td>
<td>—</td>
<td>$\tau_e3$</td>
</tr>
<tr>
<td><strong>Measurement Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land-use entropy $\lambda_1$</td>
<td>$\lambda_1$</td>
<td>0.608 ( 3.565)</td>
<td>-0.280</td>
<td>$\tau_1$</td>
</tr>
<tr>
<td>job pop ratio $\lambda_2$</td>
<td>$\lambda_2$</td>
<td>1.071 ( 1.890)</td>
<td>0.012</td>
<td>$\tau_2$</td>
</tr>
<tr>
<td>VMT Den $\lambda_3$</td>
<td>$\lambda_3$</td>
<td>1.104 ( 0.348)</td>
<td>0.034</td>
<td>$\tau_3$</td>
</tr>
<tr>
<td>AccessRd Den $\lambda_4$</td>
<td>$\lambda_4$</td>
<td>0.948 ( 0.297)</td>
<td>-0.045</td>
<td>$\tau_4$</td>
</tr>
<tr>
<td>BikeSingFam% $\lambda_5$</td>
<td>$\lambda_5$</td>
<td>0.390 ( 3.548)</td>
<td>-1.03E-15</td>
<td>$\tau_5$</td>
</tr>
<tr>
<td>BikeMulFam% $\lambda_6$</td>
<td>$\lambda_6$</td>
<td>1.194 ( 3.572)</td>
<td>-0.049</td>
<td>$\tau_6$</td>
</tr>
<tr>
<td>Pop. Density $\lambda_7$</td>
<td>$\lambda_7$</td>
<td>1.000 ( -- )</td>
<td>—</td>
<td>$\tau_7$</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Market Value $\lambda_8$</td>
<td>$\lambda_8$</td>
<td>7.424 ( 3.729)</td>
<td>-0.366</td>
<td>$\tau_8$</td>
</tr>
<tr>
<td>Med. HH. Income $\lambda_9$</td>
<td>$\lambda_9$</td>
<td>1.000 ( -- )</td>
<td>—</td>
<td>$\tau_9$</td>
</tr>
<tr>
<td>“—” not applicable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The measurement model defines the SES and accessibility factors, both being latent, via observed income and built environment variables. The latent SES factor is measured indirectly by median household income and average market value of land parcels for each TAZ. Results suggest that market land value substitutes for the SES factor in a positive and statistically significant way. An additional $1,000 per sq. mile in average market land value is estimated to correlate with an increase of 7.424 units in the SES score ($f_2$) with a significant t-stats of 3.729.
Balanced land use and proximity to bike lanes contribute to greater accessibility in statistically significant ways. Based on a statistical test\(^2\), providing biking facilities near multi-family developments (i.e., high density) improves the accessibility factor in a more powerful way than what would result from providing these opportunities near single-family development (i.e., low density). This differential effect between single- and multi-family developments is consistent with previous conclusions on the complementary effects of residential density on the propensity to walk and bike (27) and on the reduction of vehicle ownership (17). The job-population ratio appears to be merely a mild positive component in explaining accessibility. Perhaps more robust measures should be used to describe job-housing balance. For example, one could use commuting distance or the shares of intra-zonal commute trips (28). However, these variables are unavailable at the TAZ level, but could be included for future work conducted at other geographic units where recent information is available.

Road network features (i.e., VMT density and access-road density) also exert positive ties with accessibility, despite their low statistical significance (t-stats=0.348 and 0.297, respectively). Consistent with previous findings (29), these positive ties suggest that greater street-network density may improve accessibility. More sophisticated measures of network design should be used that capture density, connectivity (e.g., link-to-node ratio [29]), and street patterns (e.g., gridded vs. cul-de-sacs). An attempt by the authors to identify intersections based on Tiger/Line® street maps was unsuccessful because some crossing road segments (from the Tiger/Line® shapefiles) were mis-represented as a whole segment, rather than split-into two segments. As a result, the ArcGIS operation yields sizable discrepancies. For example, a great number of three-link intersections was counted erroneously as a cul-de-sac (end point).

The SEM part of the model explains how neighborhood accessibility and SES influence diabetes risk. At the TAZ level, enhancing accessibility reduces diabetes risk in statistically significant ways (t stats = -7.19), so as increasing socio-economic status (t stats = -7.49). The SEM model is able to gauge self-selection bias via the coefficient \(\beta_s=0.642\) (t stats = 6.528), suggesting that people with a higher socio-economic status (SES) tend to live in more accessible neighborhoods.

As shown by the elasticities in Table 2, a one percent increase in the shares of multi-family parcels within a half mile of bike lanes is accompanied by a 0.049 percent reduction in diabetes risk, all else being equal. With a one percent rise in land-use balance, diabetes risk is estimated to decrease by 0.28 percent, holding everything else constant. A one percent increase in land value, as a proxy for the SES factor, correlates with a 0.37 percent drop in diabetes risk. A plausible explanation is that people with higher socio-economic status may have access to better quality of health care or perhaps enjoy more healthful lifestyles. For example, income was found to be a strong predictor of knowledge and acceptance of vaccinations (13). High- and low-income individuals differ more in their motivation to eat healthy food than in their ability to access and afford healthy food (12).

Figure 4 illustrates the spatial distribution of latent metrics for SES, accessibility, and diabetes risk. The central part of Travis County (where Austin is the county seat) enjoys greater

\(^2\)A Wald test is computed by

\[
\text{Wald statistic} = \sqrt{\frac{0.324}{0.050}} = 56.31 > \chi^2_{0.05}(1) = 3.841.
\]

The resulting test statistic is greater than the threshold value from the chi-square distribution table with degree of freedom 1. The test suggests that the coefficients of the two variables are significantly different.
accessibility, compared with the areas at the periphery. The central and western parts of the county exhibit higher SES that is conducive to health in contrast to the south-eastern side of the county, which shows lower SES. The predicted diabetes risk mirrors the observed diabetes prevalence, with concentration of lower risk in the central and western parts and clusters of higher risk in the eastern part.

**FIGURE 3** Latent Factors for SES, Accessibility, and Diabetes Risk (Clockwise)

![SES](image1)

![Access](image2)

![Diabetes](image3)
Results suggest that the latent-variable SEM model and map visualizations can be a powerful tool to decipher the relationship between transportation accessibility and health outcomes. Understanding such relationships is crucial to operationalizing health considerations in transportation decision making. For example, the coefficient estimates can help transportation and health practitioners to anticipate the changes in public health for a change in transportation systems (e.g., building a new bike route or adopting a mixed-use development plan). The graphical outputs (Figure 3) visualize the variation in socio-economic, accessibility, and diabetes risk over space, which may help identify under-served areas that are low in accessibility, areas that are socio-economically disadvantaged, and areas that bore a disproportionate burden of disease.

As stated in Methodology, the latent-variable SEM model is compared to alternative models to test its performance. Results suggest that the latent-variable SEM model have stronger statistical significance of individual coefficients and better goodness-of-fit for the overall prediction than the results from SEM and spatial models, respectively. The Deviance Information Criteria (DIC) of the latent-variable SEM model is also lower than those of the other two models, indicating a better overall goodness-of-fit. The spatial model (i.e., CAR) outputs suggest that a strong spatial pattern exists in diabetes risk; however, this argument should be interpreted with reservation since the TAZ-level diabetes counts are imputed from aggregates.

<table>
<thead>
<tr>
<th>TABLE 3 Model Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Latent-Variable SEM*</td>
</tr>
<tr>
<td>Land Entropy</td>
</tr>
<tr>
<td>Job pop ratio</td>
</tr>
<tr>
<td>VMTDen</td>
</tr>
<tr>
<td>AccsRdDen</td>
</tr>
<tr>
<td>BKSingFam%</td>
</tr>
<tr>
<td>BMulFam%</td>
</tr>
<tr>
<td>Land Value</td>
</tr>
<tr>
<td>medHHInc</td>
</tr>
<tr>
<td>PopDen</td>
</tr>
<tr>
<td>ρ</td>
</tr>
<tr>
<td>DIC</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
</tbody>
</table>

* Notes: For comparability among models, the coefficients in the latent-variable SEM model, which measures the direct effects of predictors on diabetic risk, is computed as the product of the coefficient in the measurement model (which establishes linkage between the observed covariates and the corresponding latent factors) and the coefficient in the SEM model (which estimates the effect of latent factors associated with SES and accessibility on diabetic risk). The variance, or the standard error squared, of the coefficient estimate that is just calculated is obtained by the property of variance, \((σ₁² + σ₂² + σ₁μ₂ + σ₂μ₁)^2\), where σ₁ and σ₂ are the corresponding standard errors from the measurement model and the SEM model, μ₁ and μ₂ the coefficient estimates (in Table 2).
CONCLUSIONS
This work quantified the statistical associations between diabetes counts and contributing factors related to neighborhood accessibility and socio-economic status (SES) after purging residential self-selection bias via a structural equation model (SEM) with latent variables. The method and results could help researchers and practitioners in the transportation and health arenas to anticipate the changes in public health for a change in transportation systems (e.g., building a new bike route or adopting a mixed-use development plan). Results can be displayed visually to show the variation in socio-economic, accessibility, and diabetes risk over space, which may help practitioners identify areas that are low in accessibility, areas that are socio-economically disadvantaged, and areas with high diabetes risk.

The latent-variable approach seems to be successful in describing unobserved factors underlying diabetes prevalence via observed (but limited or imperfect) covariates. Market land value explains the SES factor in a statistically and practically significant sense. Results also suggest that wealthier neighborhoods (characterized by higher market land value) are associated with lower diabetes risk. Based on estimates of elasticity, a one percent increase in average market land value is associated with a 0.366% reduction (elasticity = -0.366) in diabetes risk, holding everything else constant. A plausible explanation is that people with higher socio-economic status may have more knowledge about health issues, better health care, and healthier lifestyles (12, 13).

Proxies for neighborhood accessibility include land-use balance, job-population ratio, vehicle-miles traveled (VMT) density, access-road density, and shares of single- and multi-family parcels within a half mile of bike lanes. Land-use balance and proximity to bike lanes contribute to greater accessibility with practical and statistical significance. Providing biking facilities near multi-family developments (i.e., high density) would yield greater health benefits than providing these opportunities near single-family developments (i.e., low density). These differential effects are consistent with previous findings on the positive effect of residential density on the propensity to walk and bike (27) and on the reduction of vehicle ownership (17).

After teasing out self-selection bias, the SEM part of the model is able to explain the direction and magnitude of the effects of neighborhood accessibility on diabetes rates. On average, forty-eight percent of multi-family dwelling units are within a half mile distance from bike lanes in the county. A one percent increase at this average level could reduce diabetes rates by 0.7%. This rate would then translate into a reduction of 70 diabetic adults for a TAZ with 1,000 population. For a one percent rise in land-use balance, diabetes risk is estimated to decrease by 0.28 percent and diabetes rates drop by exp(-0.28) = 0.76 per person.

The paper also generates visuals of latent factors for SES, accessibility, and diabetes risk across TAZs (Figure 3). The spatial variation in accessibility (with the center enjoying greater accessibility, and lower scores elsewhere) is less than the spatial variation in the socio-economic factor over space in Travis County. In addition, the study region appears to divide into two parts: a lower-income and high diabetes risk (on the east side) vs. a higher-income and low diabetes risk (on the west side). Hence, to alleviate diabetes prevalence, emphasis should be placed not only on enhancing accessibility across demographics (e.g., expanding bike systems especially in
high-density neighborhoods), but also on elevating awareness of health issues and promoting healthful lifestyles for the less-privileged.

This work attests to the importance of accessibility and socio-economic status (SES) in understanding diabetes prevalence and demonstrates the potential of the latent-variable SEM model to quantify the effects of accessibility and SES on diabetes rates. This model could be applied to other public health settings thanks to its ability to describe unobserved factors underlying a disease and to purge the bias due to residential self-selection. The paper also reflects some aspects of the tiered approaches to data collection for different levels of health-related analysis for local transportation governments (30).

A limitation of this study is missing information about nutrition, attitude toward health, and physical activities; hence, model predictions should be interpreted with caution. Future enhancement could also calculate and use new variables that can better capture the connectivity of road networks, such as intersection density instead of lane-mile density, and street patterns (grid vs. cul-de-sacs). Nonetheless, the method is successful in characterizing the effects of the accessibility and SES factors on diabetes risk and can be used by transportation and health practitioners to anticipate the reduction in diabetes rates for a change in the transportation system and to visualize disparities in diabetes risk, accessibility, and wealth across neighborhoods.

ACKNOWLEDGEMENTS
The author is grateful to Rebecca Gleason and Catherine Castokis for their many useful comments and edits on an earlier draft. Ms. Jessie Patton-Levine, Epidemiology and Health Statistics Unit of the Health and Human Services Department of Austin, Texas, provides the raw diabetes estimates. The author bears sole responsibility for any errors.
REFERENCES
Outweigh the Risk? Environmental Health Perspectives 118 (8): 1109-1116.
benefits of cycling in urban environments compared with car use: health impact assessment
study. British Medical Journal 343: 4521-4528
Policy 29 (19): 121-131
Heart Disease and Diabetes. Geneva:WHO. URL:
Department of Health, by the University of Florida Department of Urban & Regional Planning
and Center for Health & the Built Environment. Accessed 10/9/2013 at URL:
http://www.doh.state.fl.us/Planning_eval/HIA/Training/FLORIDA_HIA_FINAL_REPORT.pdf
Report, Health Impact Assessment: Proposed Cleanup Plan for the Lower Duwamish Waterway
from Alaska. Northwest Public Health, Fall/Winter
Compact. Accessed 10/9/2013 at URL:
http://www.massdot.state.ma.us/GreenDOT/HealthyTransportationCompact.aspx
attitudes and behavior towards access, availability and motivation to eat fruit and vegetables.
Public Health Nutrition 6(2): 159-168
behaviors of low-income women considered high priority for receiving the novel influenza A
Transport Policy Institute. URL: http://www.vtbi.org/railben.pdf
environment and travel behavior? Evidence from northern California. Transportation Research
Part D,10: 427-444.
Logit analysis. University of California Transportation Center. URL:
http://www.uctc.net/papers/604.pdf


