AN INTEGRATED LATENT CONSTRUCT MODELING FRAMEWORK FOR PREDICTING PHYSICAL ACTIVITY ENGAGEMENT AND HEALTH OUTCOMES

Megan M. Hoklas
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Phone: 512-471-4535, Fax: 512-475-8744
Email: mhoklas@utexas.edu

Subodh K. Dubey
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: subbits@gmail.com

Chandra R. Bhat (corresponding author)
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Phone: 512-471-4535, Fax: 512-475-8744
Email: bhat@mail.utexas.edu

Venu M. Garikapati
Arizona State University
School of Sustainable Engineering and the Built Environment
Tempe, AZ 85287-3005
Tel: (480) 965-3589; Fax: (480) 965-0557
Email: venu.garikapati@asu.edu

Ram M. Pendyala
Georgia Institute of Technology
School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: 404-894-2201; Fax: 404-894-2278; Email: rampendyala@gmail.com

Dae Hyun You
Arizona State University
School of Sustainable Engineering and the Built Environment
Tempe, AZ 85287-3005
Tel: (480) 965-3589; Fax: (480) 965-0557
Email: dhyou@asu.edu

Abstract
The health and well-being of individuals is related to their activity-travel patterns. Individuals who undertake physically active episodes such as walking and bicycling are likely to have improved health outcomes compared to individuals with sedentary auto-centric lifestyles. Activity-based travel demand models are able to predict activity-travel patterns of individuals at a high degree of fidelity, thus providing rich information for transportation and public health professionals to infer health outcomes that may be experienced by individuals in various geographic and demographic market segments. However, models of activity-travel demand do not account for the attitudinal factors and lifestyle preferences that affect activity-travel and mode use patterns. Such attitude and preference variables are virtually never collected explicitly in travel surveys, rendering it difficult to include them in model specifications. This paper applies Bhat's (2014) Generalized Heterogeneous Data Model (GHDM) approach, whereby latent constructs representing the degree to which individuals are health conscious and inclined to pursue physical activities may be modeled as a function of observed socio-economic and demographic variables and then included as explanatory factors in models of activity-travel outcomes and walk and bicycle use. The model system is estimated on the 2005-2006 National Health and Nutrition Examination Survey (NHANES) sample, demonstrating the efficacy of the approach and the importance of including such latent constructs in model specifications that purport to forecast activity and time use patterns.

Keywords: activity-travel and health, generalized heterogeneous data model, attitudes and lifestyle preferences, physical activity and health outcomes, activity-travel behavior models

1. INTRODUCTION

Recent advances in travel demand modeling have focused on the microsimulation of human activity-travel patterns with a view to better understand how, where, when, and why individuals pursue activities and allocate time to various travel and activity episodes (Arentze and Timmermans, 2008). Models of activity-travel behavior have traditionally used an array of observed explanatory variables to forecast activity-travel and mode usage patterns under a wide variety of scenarios. These include such variables as built environment attributes, network level of service variables, and household and person socio-economic and demographic variables. A missing ingredient in the modeling of human activity-travel patterns continues to be attitudinal constructs that capture the lifestyle preferences and proclivity of individuals (Mokhtarian and Salomon, 2001). In the absence of such constructs, these aspects of human lifestyle are not explicitly accounted for and are presumed to be absorbed in the random error terms of the econometric model chain.

This paper strives to overcome this limitation by presenting a latent variable structural equations modeling approach that includes latent constructs representative of the lifestyle choices or proclivity of individuals. The motivation for this study is derived from the increasing interest in drawing a connection between human activity-travel patterns and public health outcomes. Public health continues to be a top priority globally, with the rise of obesity in both children and adults contributing to studies of active living and how the built environment and the transportation system can be designed to foster active and healthy lifestyles (Frank, 2000). Research has shown that walking and biking are the most popular forms of physical activity and therefore planners are striving to create more walking and biking friendly-communities (Steinman et al, 2010). These strategies are receiving particular attention because individuals who are overweight or obese are at elevated risk of health problems such as coronary heart disease, high blood pressure, stroke, and cancer (National Heart, Lung, and Blood Institute, 2012). According to the U.S. Department of Health and Human Services (1996), the body reacts positively to physical activity leading to reduced risk of adverse medical conditions (Blair et al, 2001). Kimm et al (2005) found that individuals who are less physically active on a regular basis had an increase in BMI (body-mass index), a critical indicator of health. The common recommendation is that individuals should complete at minimum 30 minutes of physical activity at a moderate level five times a week to promote good health (U.S. Department of Health and Human Services 2008). The goal of these programs and recommendations is to promote healthier lifestyles, physical activity, and good nutrition habits.

Given this interest in the nexus between transportation and public health, it is encouraging to note that activity-based microsimulation models are able to offer rich information about people's activity and travel patterns in time and space at a high degree of fidelity. Specifically, activity-based microsimulation models are able to offer measures of walking, bicycling, and physical (recreational/sport) activity engagement that can be used to assess the potential health implications of an individual's activity-travel and mode usage pattern. However, there may also be unobserved lifestyle preferences and attitudes that impact the amount of walking, bicycling, and physical activity that a person undertakes. Those who are more fitness conscious, love the outdoors, and are active-living oriented are likely to undertake more physical activity than those who are less fitness conscious and more prone to sedentary lifestyles . Travel survey data sets do not include information about such lifestyle preferences and attitudes and hence these variables do not make their way into the activity-based model systems.

This study attempts to address this gap by formulating a latent variable structural equations model system where latent constructs (or variables) representative of an individual's lifestyle are modeled as a function of observed socio-economic and demographic variables routinely collected in travel surveys. These latent constructs, together with socio-economic and demographic variables, are then used to predict an array of activity outcomes that describe how people spend time and pursue activities of various types (including walking, bicycling, and physically active episodes). In this research, attitudinal constructs describing health consciousness and physical activity propensity are modeled as latent variables. The model system then utilizes the latent variables as explanatory factors (together with socio-economic and demographic attributes) to predict a series of activity and time use outcomes that are important to public health. The central idea is that activity-based models in practice could potentially be enhanced in the future to include such latent lifestyle constructs, with a view to better predicting how individuals pursue activities, use modes, and spend time for various types of activities.

The remainder of this paper is organized as follows. The next section describes the data source used, the data formation process, and the hypothesized impact of the latent variables. The third section details the methodology used in this effort. The fourth section presents estimation results for the latent constructs and the activity outcomes. The fifth and final section offers conclusions and directions for future research.

2. DATA DESCRIPTION

The data set for this paper is derived from the 2005-2006 National Health and Nutrition Examination Survey (NHANES), which incorporates information from both interviews and physical health examinations into the survey. The NHANES is administered by the National Center for Health Statistics (NCHS) as part of the Centers for Disease Control and Prevention's (CDC) initiative to produce national health statistics. The survey is administered to a nationally representative sample of approximately 5,000 people each year. The 2005-2006 data set, which covers a period of two years, solicited responses from a total of 10,348 individuals. The primary reason for choosing this specific version of the NHANES is that the 2005-2006 survey included a 'Physical Activity and Individual Activities Data' section, which is not included in newer versions of the survey. This section of the survey offered detailed information on specific leisure-time activities pursued by those 12 years of age and older, including data on the specific activity the respondent participated in, the intensity level of the activity, the number of episodes completed within the past month, and average duration of an episode for that activity. The initial dataset downloaded from the NHANES website consisted of responses from a total of 6,408 individuals. After extensive cleaning and filtering of the data, the estimation data set consisted of responses from 4,111 individuals.

Seventeen dependent variables, or outcomes (also sometimes referred to as indicators), related to health or physical activity participation were considered for this study. The first health outcome considered is the blood pressure of the individuals. Systolic and diastolic blood pressure readings were taken for each individual, with repetitive measurements taken to capture the best blood pressure readings. Body mass index is another health outcome used in this study. Trained health technicians measured the body mass index of each participant in kg/m^2 (defined as weight defined by square of the height). Using body mass index groupings provided by the CDC, individuals were placed into one of three categories: underweight and normal weight, overweight, or obese. The next health outcome, self-rated health, refers to how participants of the survey rated their general level of health on a five-point scale (poor, fair, good, very good, and

excellent). Another health measure used in this study is an indicator of whether individuals were covered by health insurance or any kind of health care plan. In the final data set used for analysis, 21.2 percent were not covered by any type of insurance or health care plan. Finally, there are three health indicators which specify whether a doctor has ever told the individuals that they have allergies, asthma, or diabetes.

Along with the seven health outcomes, there were ten physical activity participation outcomes included in this study. The physical activities from the survey were classified into four categories: bicycling, walking, moderate "other", and vigorous "other". The "other" category included an assortment of 46 leisure activities (other than walk and bike). Moderate intensity is defined by the survey as causing light sweating or a moderate increase in the heart rate, while vigorous intensity is associated with heavy sweating or a large increase in the heart rate. The first four outcome variables measure the average duration (in minutes per month) of engaging in the four physical activity categories. For modeling purposes, a logarithmic transformation is used for the four continuous duration variables to avoid prediction of negative values.

The next four outcome variables, specified as count variables, describe the number of times the individual participated in each of the physical activity groups over the period of a month. The final two indicators considered are the average number of hours that the participants watch television and the number of hours they use the computer in a day. Both indicators had six response categories, i.e., 0 hours, ≤1 hour, >1-2 hours, >2-3 hours, >3-4 hours, and >4 hours. The explanatory variables included in this study are age, ethnicity, education status, gender, and household annual income. The age and income variables were aggregated into four categories each. The categories for age are: 12-19 years, 20-44 years, 45-64 years, and 65 years or greater, while the categories for income are < \$25,000, \$25,000-\$44,999, \$45,000-\$64,999, and ≥ \$65,000. A complete profile of the survey sample used in this study is shown in Table 1. It is found that about one-half of the individuals in the sample have normal blood pressure. However, about two-thirds of the sample are either overweight or obese. More than one-in-five (22 percent) spend four or more hours per day watching television. The statistics on the monthly durations and counts of participation in the four physical activity categories are not provided in Table 1 to conserve on space. But, essentially, the average monthly duration in bicycling was about 75 minutes, in walking was about 230 minutes (just shy of four hours), in moderate "other" activities was 387 minutes (about 6.5 hours), and in moderate "vigorous" was 430 minutes (about 7 hours). For the corresponding counts, about 87% of respondents did not have any participation episodes in bicycling (average count of 1.66), 68% did not have any participation episodes in walking (average count of 5.68), 58% did not have any participation episodes in moderate "other" (average count of 7.84), and 61% did not have any participation episodes in vigorous "other" (average count of 8.85).

Two latent variable constructs were developed in this study: Health Consciousness and Physical Activity Propensity. The first latent variable "Health Consciousness" depicts an attitude or awareness towards one's health. Those who are more health-conscious presumably lead a lifestyle that promotes good health and focus on improving their health. The second latent variable "Physical Activity Propensity" reflects an individual's natural inclination to participate in physical activities and is representative of the lifestyle (active or not) that an individual adopts. Individuals who have a higher physical activity propensity will tend to participate in more physical activities, and choose to bicycle or walk.

Hoklas, Dubey, Bhat, Garikapati, Pendyala, You

3. MODEL STRUCTURE

This section presents the model structure and methodology adopted in this paper. First, the modeling framework is described to provide an understanding of how the various indicators and latent constructs are related in an integrated model system. Following the presentation of the framework, the paper offers a brief overview of the modeling methodology and formulation.

3.1 Model Framework

The model framework is presented in Figure 1. There are two sets of exogenous variables. The first set constitute socio-economic and demographic variables such as age, ethnicity, gender, education, and household income. The second set are exogenous health variables such as having asthma, allergies, or diabetes, and having health insurance or a health care plan. While it may be argued that the health conditions (asthma, allergies, and diabetes) are outcomes, it is often the case that these health conditions are not entirely under the control of the individual and may be significantly influenced by family history and heredity (University of Maryland Medical Center, 2011).

There are two health and physical activity related latent constructs that are considered in this paper. Both of these constructs, health consciousness and physical activity propensity, are modeled as a function of socio-economic and demographic variables as well as exogenous health variables. In this particular study, it was found that exogenous health variables were not significant in the physical activity propensity latent variable equation and hence an arrow from exogenous health variables to the physical activity propensity is suppressed in the figure.

The ultimate goal of the equations system is to offer a framework that allows the prediction of the physical activity outcomes, activity and time allocation patterns, and mode usage (walking and bicycling), while explicitly incorporating latent variables or constructs that reflect health consciousness and physical activity propensity (lifestyle). As shown in the figure, the health outcomes are modeled as a function of exogenous health variables, socio-economic and demographic variables, and the two latent constructs of health consciousness and physical activity propensity. Similarly, the ten physical activity outcomes are also modeled as a function of these four entities. A rather notable limitation of this study is the absence of built environment attributes and network level of service characteristics as explanatory factors. The households (respondents) are not geo-coded to any level of geography making it impossible to match such secondary attributes to the records in the data set. On the other hand, this is a unique data set that includes both a series of measures related to health, and a series of variables related to physical activity engagement and bicycling and walking. Reflecting the influence of latent health and lifestyle constructs on health outcomes and physical activity indicators requires a data set that includes these two types of information. For this reason, this data set has been utilized for this research, and future research efforts should aim to include contextual variables in the modeling effort.

The set of health and physical activity outcomes (three health outcomes and ten physical activity outcomes) constitute a mix of dependent variable types including nominal variables, ordered variables, and continuous variables. The modeling methodology needs to accommodate this mixture of variable types in a simultaneous equations modeling framework. In this paper, we use Bhat's (2014) Generalized Heterogeneous Data Model (GHDM) framework for the analysis. There are two components to the proposed model structure: (1) the latent variable structural equations model, and (2) the latent variable measurement equation model.

Only a brief account of the modeling methodology is provided here in the interest of brevity. The reader is referred to Bhat (2014) for details. For ease of notation, a cross-sectional model is presented. Where appropriate, index q for decision-makers is supressed ($q=1,2,\dots Q$), and it is assumed that all error terms are independent and identically distributed across decision-makers.

3.2 Latent Variable Structural Equations Model (SEM)

Let l be an index for latent variables ($l=1,2,\dots L$). Consider the latent variable z_l^* and write it as a linear function of covariates:

$$z_l^* = \boldsymbol{\alpha}_l' \boldsymbol{w} + \eta_l, \tag{1}$$

where $\boldsymbol{w}$ is a $(\tilde{D}\times1)$ vector of observed covariates, $\boldsymbol{\alpha}_l$ is a corresponding $(\tilde{D}\times1)$ vector of coefficients, and η_l is a random error term assumed to be standard normally distributed for identification purposes. Next, define the $(L\times\tilde{D})$ matrix $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2,\dots\boldsymbol{\alpha}_L)'$, and the $(L\times1)$ vectors $\boldsymbol{z}^* = (z_1^*, z_2^*,\dots, z_L^*)'$ and $\boldsymbol{\eta} = (\eta_1,\eta_2,\eta_3,\dots,\eta_L)'$. Unlike much of the earlier research in integrated choice latent variable (ICLV) modeling, this effort allows a multivariate normal (MVN) correlation structure for $\boldsymbol{\eta}$ to accommodate interactions among the unobserved latent variables: $\boldsymbol{\eta} \sim MVN_L[\boldsymbol{0}_L, \boldsymbol{\Gamma}]$, where $\boldsymbol{0}_L$ is an $(L\times1)$ column vector of zeros, and $\boldsymbol{\Gamma}$ is a $(L\times L)$ correlation matrix. In matrix form, equation (1) may be written as:

$$\boldsymbol{z}^* = \boldsymbol{\alpha}\boldsymbol{w} + \boldsymbol{\eta} \tag{2}$$

3.3 Latent Variable Measurement Equation Model Components

Consider a combination of continuous, ordinal, group, count, and nominal outcomes (indicators) influenced by the underlying latent variable vector $\boldsymbol{z}^*$. These outcomes may be a function of a set of exogenous variables as well. Let there be H continuous outcomes ($y_1, y_2,\dots, y_H$) with an associated index h ($h=1,2,\dots,H$). Let $y_h = \boldsymbol{\gamma}_h'\boldsymbol{x} + \boldsymbol{d}_h'\boldsymbol{z}^* + \varepsilon_h$ be the usual linear regression function, where $\boldsymbol{x}$ is an $(A\times1)$ vector of exogenous variables (including a constant) as well as possibly the observed values of other endogenous continuous variables, other endogenous ordinal variables, other endogenous count variables, and other endogenous nominal variables (introduced as dummy variables). $\boldsymbol{\gamma}_h$ is a corresponding compatible coefficient vector.[1] $\boldsymbol{d}_h$ is an $(L\times1)$ vector of latent variable loadings on the hth continuous outcome, and ε_h is a normally distributed measurement error term.

---

[1] In joint variable systems with nominal, ordinal, grouped, and count dependent variables, the structural effects of one limited-dependent variable on another can only be in a single direction. That is, it is not possible to have correlated unobserved effects underlying the propensities determining two limited-dependent variables, as well as have the observed limited-dependent variables themselves structurally affect each other in a bi-directional fashion. This creates a logical inconsistency problem (see Maddala, 1983, page 119 for a good discussion). In our empirical analysis, we considered alternative direction of endogenous effects, and the ones we will present provided the best data fit. It is critical to note that, regardless of which directionality of effects among the endogenous variables is specified (or even if no relationships are specified), the system is a joint bundled system because of the correlation in unobserved factors impacting the underlying propensities.

Hoklas, Dubey, Bhat, Garikapati, Pendyala, You

Next, consider N ordinal outcomes (indicator variables) for the individual, and let n be the index for the ordinal outcomes $(n = 1, 2, ..., N)$. Also, let J_n be the number of categories for the nth ordinal outcome $(J_n \geq 2)$ and let the corresponding index be $j_n (j_n = 1, 2, ..., J_n)$. Let $\tilde{y}_n^*$ be the latent underlying variable whose horizontal partitioning leads to the observed outcome for the nth ordinal variable. Assume that the individual under consideration chooses the a_n^{th} ordinal category. Then, in the usual ordered response formulation for the individual:

$$\tilde{y}_n^* = \tilde{\gamma}_n' x + \tilde{d}_n' z^* + \tilde{\varepsilon}_n, \text{ and } \tilde{\psi}_{n,a_n-1} < \tilde{y}_n^* < \tilde{\psi}_{n,a_n}, \tag{3}$$

In equation (3) x is a fixed vector of variables as defined earlier, $\tilde{\gamma}_n$ is a corresponding vector of coefficients to be estimated, $\tilde{d}_n$ is an $(L \times 1)$ vector of latent variable loadings on the nth continuous outcome, the $\tilde{\psi}$ terms represent thresholds, and $\tilde{\varepsilon}_n$ is the standard normal random error for the nth ordinal outcome. For each ordinal outcome, $\tilde{\psi}_{n,0} < \tilde{\psi}_{n,1} < \tilde{\psi}_{n,2} ... < \tilde{\psi}_{n,J_n-1} < \tilde{\psi}_{n,J_n}$; $\tilde{\psi}_{n,0} = -\infty, \tilde{\psi}_{n,1} = 0$, and $\tilde{\psi}_{n,J_n} = +\infty$.

Consider V grouped outcomes (indicator variables) for the individual, and let v be the index for the grouped outcomes $(v = 1, 2, ..., V)$. Also, let J_v be the number of categories for the vth grouped outcome $(J_v \geq 2)$ and let the corresponding index be $j_v (j_v = 1, 2, ..., J_v)$. Let $\vec{y}_v^*$ be the latent underlying variable whose horizontal partitioning leads to the observed outcome for the vth grouped variable. Assume that the individual under consideration chooses the a_v^{th} grouped category. Then, in the grouped response formulation for the indivdual:

$$\vec{y}_v^* = \vec{\gamma}_v' x + \vec{d}_v' z^* + \vec{\varepsilon}_v, \text{ and } \vec{\psi}_{v,a_v-1} < \vec{y}_v^* < \vec{\psi}_{v,a_v}, \tag{4}$$

where x is a fixed vector of variables as defined earlier, $\vec{\gamma}_v$ is a corresponding vector of coefficients to be estimated, $\vec{d}_v$ is an $(L \times 1)$ vector of latent variable loadings on the vth grouped outcome, the $\vec{\psi}$ terms represent thresholds, and $\vec{\varepsilon}_b$ is the standard normal random error for the vth grouped outcome. For each grouped outcome, $\vec{\psi}_{v,0} < \vec{\psi}_{v,1} < \vec{\psi}_{v,2} ... < \vec{\psi}_{v,J_v-1} < \vec{\psi}_{v,J_v}$; $\vec{\psi}_{v,0} = -\infty$, and $\vec{\psi}_{v,J_v} = +\infty$, which are observed thresholds that do not need to be estimated.

Let there be C count variables for a household, and let c be the index for the count variables $(c = 1, 2, ..., C)$. Let the count index be $k_c (k_c = 0, 1, 2, ..., \infty)$ and let r_c be the actual observed count value for the household. Then, following the recasting of a count model as a generalized ordered-response probit formulation (Castro et al, 2011; Bhat et al, 2013), a generalized version of the negative binomial count model may be written as:

$$\breve{y}_c^* = \breve{d}_c' z^* + \breve{\varepsilon}_c, \breve{\psi}_{c,r_c-1} < \breve{y}_c^* < \breve{\psi}_{c,r_c}, \tag{5}$$

$$\breve{\psi}_{c,r_c} = \Phi^{-1} \left[\frac{(1-\upsilon_c)^{\theta_l}}{\Gamma(\theta_c)} \sum_{t=0}^{r_c} \left(\frac{\Gamma(\theta_c+t)}{t!} (\upsilon_c)^t \right) \right] + \varphi_{c,r_c}, \upsilon_c = \frac{\lambda_c}{\lambda_c + \theta_c}, \text{ and } \lambda_c = e^{\breve{\gamma}_c' x} \tag{6}$$

where $\breve{y}_c^*$ is a latent continuous stochastic propensity variable associated with the count variable c that maps into the observed count r_c through the $\breve{\psi}_c$ vector (which is a vertically stacked column vector of thresholds $(\breve{\psi}_{c,-1}, \breve{\psi}_{c,0}, \breve{\psi}_{c,1}, \breve{\psi}_{c,2}, ...)'$. $\breve{d}_c$ is an $(L \times 1)$ vector of latent variable

Hoklas, Dubey, Bhat, Garikapati, Pendyala, You

loadings on the cth count outcome, and $\breve{\varepsilon}_c$ is a standard normal random error term. $\breve{\gamma}_c$ is a column vector corresponding to the vector $\boldsymbol{x}$ (including a constant) of exogenous observable covariates. Φ^{-1} in the threshold function of Equation (6) is the inverse function of the univariate cumulative standard normal. θ_c is a parameter that provides flexibility to the count formulation, and is related to the dispersion parameter in a traditional negative binomial model $(\theta_c > 0 \,\forall\, c)$.

$\Gamma(\theta_c)$ is the traditional gamma function; $\Gamma(\theta_c) = \int_{\tilde{t}=0}^{\infty} \tilde{t}^{\,\theta_c-1} e^{-\tilde{t}} \, d\tilde{t}$.

Finally, let there be G nominal (unordered-response) variables for an individual, and let g be the index for the nominal variables ($g = 1, 2, 3, \ldots, G$). Also, let I_g be the number of alternatives corresponding to the gth nominal variable ($I_g \geq 3$) and let i_g be the corresponding index ($i_g = 1, 2, 3, \ldots, I_g$). Consider the gth nominal variable and assume that the individual under consideration chooses the alternative m_g. Also, assume the usual random utility structure for each alternative i_g. Then,

$$U_{gi_g} = \boldsymbol{b}'_{gi_g} \boldsymbol{x} + \vartheta'_{gi_g}(\boldsymbol{\beta}_{gi_g} \boldsymbol{z}^*) + \varsigma_{gi_g}, \tag{7}$$

where $\boldsymbol{x}$ is the same fixed vector of exogenous variables as earlier, $\boldsymbol{b}_{gi_g}$ is an $(A \times 1)$ column vector of corresponding coefficients, and ς_{gi_g} is a normal error term. $\boldsymbol{\beta}_{gi_g}$ is an $(N_{gi_g} \times L)$-matrix of variables interacting with latent variables to influence the utility of alternative i_g, and ϑ_{gi_g} is an $(N_{gi_g} \times 1)$-column vector of coefficients capturing the effects of latent variables and their interaction effects with other exogenous variables. If each of the latent variables impacts the utility of the alternatives for each nominal variable purely through a constant shift in the utility function, $\boldsymbol{\beta}_{gi_g}$ will be an identity matrix of size L, and each element of ϑ_{gi_g} will capture the effect of a latent variable on the constant specific to alternative i_g of nominal variable g. Let $\boldsymbol{\varsigma}_g = (\varsigma_{g1}, \varsigma_{g2}, \ldots \varsigma_{gI_g})'$ ($I_g \times 1$ vector), and $\boldsymbol{\varsigma}_g \sim MVN_{I_g}(0, \boldsymbol{\Lambda_g})$. Taking the difference with respect to the first alternative, the only estimable elements are found in the covariance matrix $\breve{\boldsymbol{\Lambda}}_g$ of the covariance matrix of the error differences, $\breve{\boldsymbol{\varsigma}}_g = (\breve{\varsigma}_{g2}, \breve{\varsigma}_{g3}, \ldots, \breve{\varsigma}_{gI_g})$ (where $\breve{\varsigma}_{gi} = \breve{\varsigma}_{gi} - \breve{\varsigma}_{g1}$, $i \neq 1$). Further, the variance term at the top left diagonal of $\breve{\boldsymbol{\Lambda}}_g$ ($g=1,2,\ldots,G$) is set to unity to account for scale invariance.

A comprehensive account of the model structure, sufficient identification conditions, and estimation procedure is beyond the scope of this paper, but is available in Bhat (2014).

4. MODEL ESTIMATION RESULTS

This section presents model estimation results. Results are presented first for the latent variable structural equation estimation, followed by the results for latent variable measurement equation components (non-nominal and nominal outcomes).

Hoklas, Dubey, Bhat, Garikapati, Pendyala, You

4.1 Latent Variable Structural Equation Model Results
Table 2 provides the results for the effects of individual-specific variables on the two latent constructs in the structural equation model. In the case of the 'Physical Activity Propensity' latent variable, only age is found to be significant in representing or capturing this trait of an individual. It was found that individuals who had the highest propensity to be physically active are in the 12-19 years age group (the youngest age group), which then decreased in magnitude as the individual's age increases. Telama and Yang (2000) found that after age 12, the frequency of participating in at least 30 minutes of physical activity declined continually up to age 27. This trend may be consistent with the notion that individuals may find it difficult to allocate time for leisure activities due to work commitments, responsibilities, and/or maintaining and caring for their family. In addition, those in the 12-19 year age group may participate in organized sports activities inside or outside of their educational setting. Such activities are likely to diminish with advancing age.

With respect to the "Health Conscious Attitude" latent variable, the results indicate that the higher the education status of an individual, the higher their level of health consciousness. This may be reflective of educated individuals having better knowledge and greater awareness of the ingredients necessary to lead a healthy life, and having the resources necessary to implement these ingredients (Kenkel, 1991). The gender variable also contributes to "health consciousness", with men being less health-conscious than women. Wardle et al (2004) report similar findings stating that women attach a greater importance to healthy eating and are more likely to be conscious of their nutritional intake than their male counterparts. The model also indicates that individuals who have health insurance coverage have a higher level of health consciousness than those who are uninsured.

The correlation coefficient between the "Health Conscious Attitude" and "Physical Activity Propensity" latent constructs is statistically significant and positive at a value 0.122, indicating that individuals who are more health-conscious are also more inclined to participate in physical activity than their less health-conscious peers.

4.2 Latent Variable Measurement Equation Results (Non-Nominal Outcomes)
Table 3 provides results for the latent variable measurement equation components, which consists of twelve indicator variables (four continuous, three group, four count, and one ordinal). The set of twelve indicator variables includes two health outcome variables and ten physical activity and activity/time-use variables. The results of each component are discussed in this section.

The first component corresponds to the natural logarithm of walk duration, one of the continuous variables. It is found that those who have a normal weight body mass index (BMI) tend to walk for less duration compared to those with higher (overweight and obese) BMI values. Evidence in the US supports this as it has been reported that the highest percentage of people walking are from the overweight and obese BMI categories (Simpson et al, 2003). Those with allergies and asthma are likely to devote less time to walking, presumably because walking outdoors aggravates these conditions creating significant discomfort to the individual (Foty et al, 2013). Walk duration is also seen to vary by race and gender. The "Health Conscious" latent variable is found to have a positive impact on walk duration. This latent construct also has a positive influence on time spent bicycling (Bike Duration) and in fact, is the only explanatory factor that appears significant in the model specification. Clearly, health conscious individuals allocate more time to walking and bicycling. The error variances of these two continuous

Hoklas, Dubey, Bhat, Garikapati, Pendyala, You

components is fixed to 1.0, because they were very close to 1.0 in most specifications, and we noticed that fixing these led to accelerated convergence.

The time allocated to moderate physical activities is significantly lower for those who fall within the normal or overweight BMI categories. It appears that obese individuals allocate more time to moderate physical activities, presumably in an attempt to improve their health. It is found that Hispanic males allocate less time to moderate physical activity, while Caucasian males and females have a positive impact on activity duration. Those with a normal or overweight BMI spend less time on vigorous activities compared to their obese BMI counterparts, who may be attempting to rectify their health through physical activity engagement. Those with allergies spend less time pursuing vigorous physical activities. Caucasian, African-American, and Hispanic females, and Caucasian males, were all found devote less time to vigorous activities. Only African-American males were found to pursue longer periods of vigorous physical activities.

The self-reported health rating results suggest that those who are obese, diabetic, or have asthma are likely to report a lower state of health, while those who participate in at least 30 minutes of physical activity on a regular basis are likely to report a higher level of health well-being. These findings are consistent with those reported by Okosun et al (2001) who found an inverse association between the proportion of individuals reporting excellent health status and the proportion of those who were obese. As expected, health conscious individuals are more likely to report a higher health rating.

The first grouped variable, blood pressure, is influenced by a number of indicators and socio-demographic variables. Blood pressure is higher for individuals who are overweight, obese, or diabetic. The number of hours watching TV (sedentary activity) is associated with higher levels of blood pressure, a finding consistent with that reported by Sugiyama et al (2007). Along similar lines, the results indicate that participating in any type of physical activity for 30 minutes or more per day will lower blood pressure. Finally, the results indicate that females in general have lower blood pressure than their male counterparts (see Reckelhoff, 2001, for similar findings). It is interesting to note that individuals with asthma report lower levels of blood pressure; it appears that those with asthmatic conditions may be adopting healthy lifestyles to minimize the adverse impacts of their asthmatic symptoms and this in turn leads to lower blood pressure readings. Another interesting finding is that neither of the latent constructs significantly affects blood pressure; it is likely that blood pressure is more of a medical outcome as opposed to an activity/time-use outcome. While the latter is clearly impacted by latent constructs (representing lifestyle), the former is affected by the activity/time-use patterns rather than the latent constructs per se.

The number of hours spent watching television is affected by the "Physical Activity Propensity" latent variable. Those who have a higher propensity to engage in physical activities devote less time to watching television, a finding consistent with expectations. On the other hand, the time allocated to activities on the computer is affected by the education status of the individual rather than any health or physical activity related variables. Consistent with findings in the American Time Use Survey (Bureau of Labor Statistics 2013), individuals with a higher level of education spend more time on the computer.

The number of walking episodes (a count variable) is affected by several variables. As expected, those who are health conscious (latent construct) are inclined to undertake a greater number of walking episodes. Those who are normal weight report fewer walking episodes while those who are overweight report a higher number of walking episodes. It appears that the

overweight individuals are attempting to shed some of the weight through walking activities (Simpson et al, 2003). Hispanic males and African-American males are likely to pursue a greater number of walking episodes; this may, at least in part, be due to the lower levels of auto ownership in Hispanic and African-American households resulting in males in these households walking more than other groups. Caucasians (whether male or female) show lower levels of walking, as do African-American females.

For the number of biking episodes (a count variable), it is found that gender plays an important part as does age. Females pursue fewer biking episodes than their male counterparts, regardless of race, a finding consistent with that reported in past studies (Garrard et al, 2008). The 12-19 years age group reports a higher number of biking episodes, consistent with the notion that teenagers (with lower levels of auto ownership) are likely to use the bicycle to meet their mobility needs. Winters et al (2007) found that individuals in the 12 to 19 year age group were three times more likely to bicycle than their older counterparts and that bicycling rates decreased steadily with age. Health conscious individuals are likely to pursue a greater number of bicycling episodes.

Those in the normal weight and overweight categories are likely to pursue fewer moderate physical activity episodes in comparison to their obese counterparts. Individuals in the obese category may be attempting to pursue a higher number of such episodes in an attempt to improve their health. Individuals with allergies and asthma report undertaking fewer moderate activity episodes, consistent with the notion that such conditions limit the ability of individuals to undertake physical activities particularly outdoors. African-American and Hispanic individuals (both male and female) were found to report a larger number of vigorous physical activity episodes, relative to their Caucasian counterparts. The relationship between race and moderate physical activity episode engagement is less clear. For both moderate and vigorous physical activity engagement, the physical activity propensity (latent construct) is found to be an important and significant predictor. The significant dispersion parameters on the count dependent variables show that there is considerable heterogeneity in the population with respect to the frequencies of these variables (variance is greater than mean).

4.3 Latent Variable Measurement Equation Results (Nominal Outcome)

Table 4 displays estimation results for the nominal outcome variable, body mass index. The base category is that corresponding to normal weight. The constants suggest that there is a negative baseline propensity associated with being overweight and a positive baseline propensity associated with being obese. In other words, within this sample, individuals are more likely to fall into the obese category relative to other categories all else being equal. The results also indicate that watching more than three hours of television per day contributes to obesity, a finding consistent with extant literature (Hu et al, 2003). Individuals with asthma or allergies are likely to be obese. Being diabetic is also associated with being overweight and obese, a finding consistent with that reported in previous studies (e.g., Resnick et al, 2000). It is interesting to note that higher income individuals are more likely to be overweight. Higher income individuals may be more auto-centric in their mode choice, spend more time at work (to earn income) reducing time available for healthy physical activities, and have larger caloric intakes as they dine out more often than their lower income counterparts.

The table also presents the results of the loading of the latent variables on the nominal body mass index outcome, with the base category being obese BMI. The two latent variables, Health Conscious Attitude and Physical Activity Propensity, were statistically significant for the

overweight BMI group, albeit with opposite signs. Health conscious individuals show a reduced propensity to be overweight, as expected; however, it appears that they exhibit an equal proclivity towards being normal weight or obese. It is possible that health consciousness is associated with these two categories for different reasons. Health conscious individuals maintain good health and are of normal weight, or at the other extreme, obese individuals are health conscious as they attempt to improve their health condition. Physical activity propensity is positively associated with being normal weight or overweight, as opposed to obese, a finding that is intuitive and reasonable. Physical activity increases daily energy expenditures which can assist in weight loss and maintaining a healthy weight (Hills et al, 2011).

4.4 Measures of Fit
The performance of the GHDM model structure used here may be compared to that of model which does not consider latent constructs (NL-GHDM). The model which does not consider latent construct does not account for dependencies across different modeling dimensions (non-nominal and nominal outcome variables). The composite log-likelihood value for the GHDM model (with 132 parameters) is -490386.23, while the corresponding value for the NL-GHDM model (with 110 parameters) is -534146.07. The two models (GHDM and NL-GHDM) may be compared using the adjusted composite likelihood ratio test (ADCLRT) statistics that is approximately chi-squared distributed (see Bhat, 2011 for a detailed discussion). The ADCLRT statistic value is 203.88, which is larger than the critical chi-square value with 22 degree of freedom at any level of significance. This clearly illustrates the need to consider the dependencies across various modeling dimensions, which can be done efficiently using Bhat's (2014) GHDM model.

5. CONCLUSIONS
There is increasing interest in drawing connections between activity-travel indicators and public health outcomes. Activity-based models of travel are increasingly providing richer disaggregate information about the types of activities Many of the indicators related to physical activity participation, sedentary activity participation (such as watching television or sitting at the computer for extended periods), and extent of bicycling and walking are measures that public health professionals would be interested in connecting to health outcomes such as body mass index (BMI), blood pressure, and overall state of health.

However, despite the widespread recognition of the importance of attitudes and lifestyle preferences on activity engagement patterns and mode use, activity-based models fail to include such variables in the model specification. Engagement in physical activities, and the use of bicycle and walk modes, are likely to be influenced by the lifestyle preferences and attitudes of individuals. However, such lifestyle preferences and attitudes are rarely, if ever, measured in surveys rendering it difficult to explicitly include such measures in activity model specifications.

This study constitutes an initial attempt to fill this gap by adopting a GHDM model system in which latent constructs that describe an individual's health consciousness and physical activity propensity are modeled as a function of observed socio-economic and demographic characteristics. The resulting latent constructs, together with socio-economic and demographic variables, are then used to predict a number of activity engagement outcomes (describing frequency and duration of participation in various types of activities – both physically active and sedentary) and health outcomes (such as body mass index, self-reported health well-being, and blood pressure). The entire system of equations is estimated simultaneously through the use of

Hoklas, Dubey, Bhat, Garikapati, Pendyala, You

the maximum approximate composite marginal likelihood (MACML) estimation approach that greatly simplifies the evaluation of the likelihood function and brings about computational efficiency in the estimation of simultaneous equations model systems that involve a mixture of dependent variable types.

The findings of the paper show that latent constructs, health consciousness and physical activity propensity, are related to socio-economic and demographic variables. These latent constructs play a significant role in shaping activity-travel and mode use patterns, with those who are more health conscious or inclined towards physically active lifestyles reporting higher levels of physical activity engagement and better health outcomes. Given the significance of the latent variables in explaining activity engagement and mode use, activity-based microsimulation models may be enriched in terms of the model specification through the inclusion of such latent variables that are themselves functions of observed socio-economic and demographic variables collected in travel surveys. There has been a reluctance historically to include attitudinal and lifestyle preference variables in model specifications because such variables are not typically measured in travel surveys, and more importantly, they are difficult to forecast into the future. However, the approach proposed in this paper, where latent variables are functions of observed variables and can be included in models of activity-travel behavior, offers a mechanism by which such latent attitudinal and lifestyle constructs can be included in models of activity-travel demand.

The study is not without its limitations. Due to the nature of the study, the survey data set used for this modeling effort had to include both activity-travel indicators as well as health indicators. The 2005-2006 National Health and Nutrition Examination Survey (NHANES) offered such a data set, but this data set suffered from the drawback that it did not include any built environment, contextual, or network level of service variables – all of which invariably affect activity-travel indicators and possibly health outcomes as well. Future research and data collection efforts should attempt to include all of the variables of interest so that contextual variables may be accounted for in the model specification.

REFERENCES
Arentze, T., and H. Timmermans. Social Networks, Social Interactions, and Activity-Travel Behavior: A Framework for Microsimulation. *Environment and Planning B: Planning and Design*, Vol. 35, 2008, pp. 1012-1027.
Bhat, C.R., S. Astroza, R. Sidharthan, M. Jobair Bin Alam, WH Khushefati. A Joint Count-Continuous Model of Travel Behavior with Selection Based on a Multinomial Probit Residential Density Choice Model. Technical Paper: University of Texas at Austin, 2013.
Bhat, C.R. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B*, Vol. 45, No. 7, 2011, pp. 923-939.
Bhat, C.R. (2014) A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables, Technical Paper, Department of Civil, Architectural, and Environmental Engineering, The University of Texas at Austin, available at http://www.caee.utexas.edu/prof/bhat/abstracts/GHMD.pdf.
Blair, S., Y. Cheng, and S. Holder. Is physical activity or physical fitness more important in defining health benefits? *Medicine & Science in Sports and Exercise*, Vol. 33, No. 6 (supplementary), 2001, S379-S399.

Bureau of Labor Statistics. American Time Use Survey, 2013. http://www.bls.gov/news.release/atus.t11.htm. Accessed May 23, 2014.

Frank, L. D. Land Use and Transportation Interaction Implications on Public Health and Quality of Life. *Journal of Planning Education and Research*, Vol. 20, No. 1, 2000, pp. 6-22.

Foty, R., K. Nelligan, T. To, D. Stieb and S. Dell. Children with Asthma are Less Likely to Walk to School. *Pediatric Allergy, Immunology, and Pulmonology*, Vol. 26, No. 2, 2013, pp. 81-87.

Garrard, J., G. Rose and S. Lo. Promoting Transportation Cycling For Women: The Role of Bicycle Infrastructure. *Preventive Medicine*, Vol. 46, No. 1, 2008, pp. 55-59.

Hills, A., L. Andersen, and N. Byrne. Physical Activity and Obesity in Children. *British Journal of Sports Medicine*, Vol. 45, No. 11, 2011, pp. 866-870.

Hu, F., T. Li, G. Colditz, W. Willett and J. Manson. Television Watching and Other Sedentary Behaviors in Relation to Risk of Obesity and Type 2 Diabetes Mellitus in Women. *The Journal of the American Medical Association*, Vol. 289, No. 14, 2003, pp. 1785-1791.

Kenkel, D. S. Health Behavior, Health Knowledge, and Schooling. *Journal of Political Economy*, 1991, pp. 287-305.

Kimm, S., N. Glynn, E. Obarzanek, A. Kriska, S. Daniels, B. Barton, and K. Liu. Relation between the Changes in Physical Activity and Body-Mass Index during Adolescence: A Multicenter Longitudinal Study. *The Lancet*, Vol. 366, no. 9482, 2005, pp. 301-307.

Mokhtarian, P. L., and I. Salomon. How Derived is the Demand for Travel? Some Conceptual and Measurement Considerations. *Transportation Research Part A*, Vol. 35, No. 8, 2001, pp. 695-719.

National Heart, Lung, and Blood Institute. What are the health risks of overweight and obesity?, July 13, 2012. http://www.nhlbi.nih.gov/health/health-topics/topics/obe/risks.html, Accessed April 15, 2014.

Okosun, I., S. Choi, T. Matamoros, and A. Dever. Obesity Is Associated with Reduced Self-Rated General Health Status: Evidence from a Representative Sample of White, Black, and Hispanic Americans. *Preventive Medicine*, Vol. 32, No. 5, 2001, pp. 429-436.

Reckelhoff, J. Gender Differences in the Regulation of Blood Pressure. *Hypertension*, Vol. 37, No. 5, 2001, pp. 1199-1208.

Resnick, H., P. Valsania, J. Halter, and X. Lin. Relation of Weight Gain and Weight Loss on Subsequent Diabetes Risk in Overweight Adults. *Journal of Epidemiology and Community Health*, Vol. 54, No. 8, 2000, pp. 596-602.

Simpson, M., M. Serdula, D. Galuska, C. Gillespie, R. Donehoo, C. Macera, and K. Mack. Walking Trends among U.S. adults: The Behavioral Risk Factor Surveillance System, 1987-2000. *American Journal of Preventive Medicine*, Vol. 25, No. 2, 2003, pp. 95-100.

Steinman, L., M. Doescher, D. Levinger, C. Perry, L. Carter, A. Eyler, and C. Voorhees. Master Plans for Pedestrian and Bicycle Transportation: Community Characteristics. *Journal of Physical Activity and Health*, Vol. 7, No. 1, 2010, pp. S60-S66.

Sugiyama, T., G. Healy, D. Dunstan, J. Salmon, and N. Owen. Is Television Viewing Time a Marker of a Broader Pattern of Sedentary Behavior? *Annals of Behavioral Medicine*, Vol. 35, No. 2, 2008, pp. 245-250.

Telama, R. and X. Yang. Decline of physical activity from youth to young adulthood in Finland. *Medicine and Science in Sports and Exercise*, Vol. 32, No. 9, 2000, pp. 1617-1622.

University of Maryland Medical Center. *Asthma*. June 2011, http://umm.edu/health/medical/altmed/condition/asthma. Accessed August 1, 2014.

U.S. Department of Health and Human Services. Physical Activity and Health: A Report of the Surgeon General. DIANE Publishing, 1996.

U.S. Department of Health and Human Services. Physical Activity Guidelines for Americans, 2008.

Wardle, J., A. Haase, A. Steptoe, M. Nillapun, K. Jonwutiwes, and F. Bellisie. Gender Differences in Food Choice: The Contribution of Health Beliefs and Dieting. *Annals of Behavioral Medicine*, Vol. 27, No. 2, 2004, pp. 107-116.

Winters, M., M. Friesen, M. Koehoorn, and K. Teschke. Utilitarian Bicycling: A Multilevel Analysis of Climate and Personal Influences. *American Journal of Preventive Medicine*, Vol. 32, No. 1, 2007, pp. 52-58.
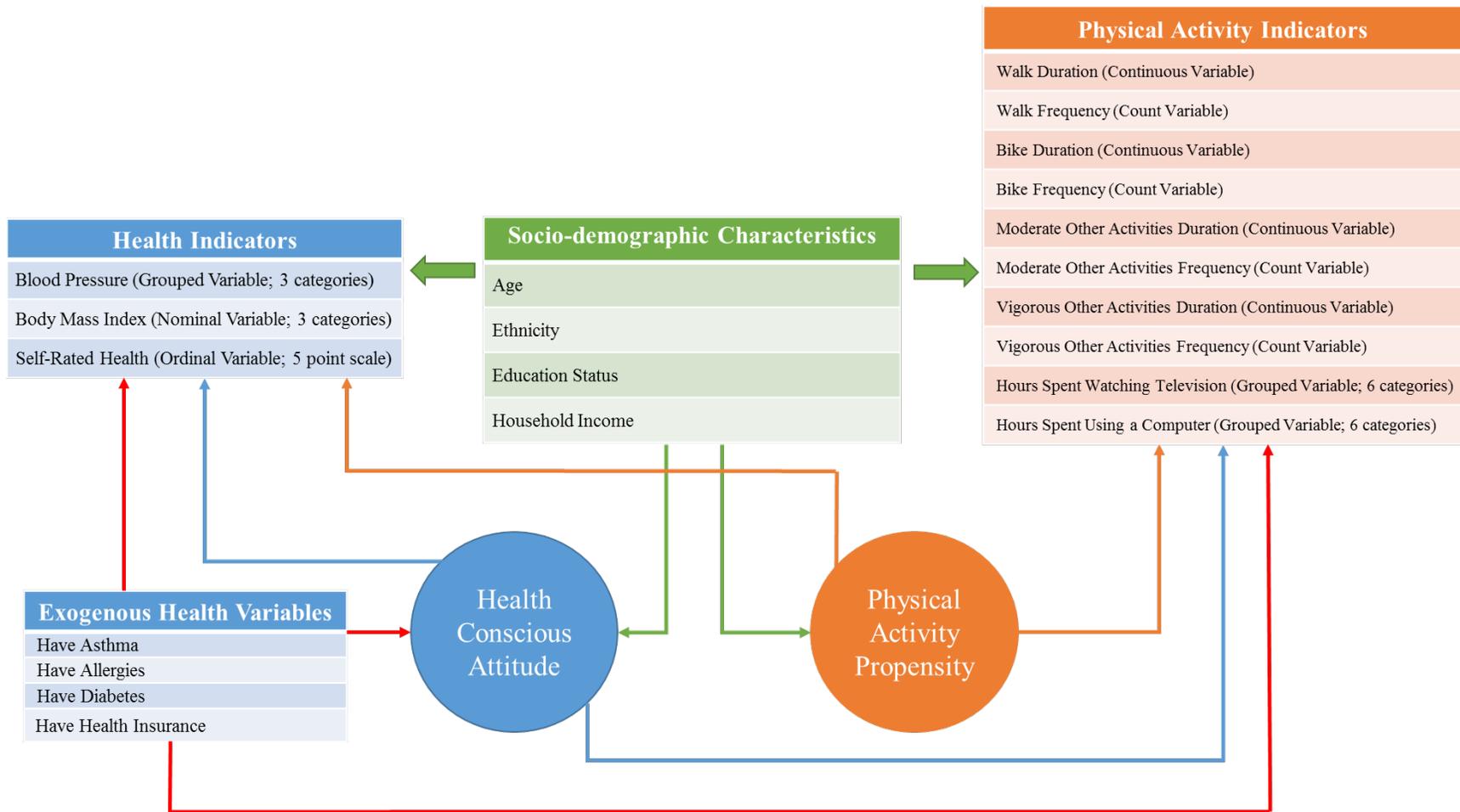
FIGURE 1. Model Framework

TABLE 1. Descriptive Characteristics of the Sample

Demographic Variables	Sample
Gender	
Female	50.64%
Male	49.36%
Age (years)	
12-19	23.45%
20-44	34.61%
45-64	26.76%
65 or older	15.18%
Ethnicity	
Caucasian	46.07%
African-American	25.15%
Mexican American or Other Hispanic	24.40%
Other Race (Included Multi-Racial)	4.38%
Household Size	
One Person	11.21%
Two People	27.85%
Three People	18.92%
Four People	17.15%
Five People	12.89%
Six People	6.59%
Seven People or More	5.38%
Education Status	
Less than 9th Grade	17.78%
9th through 11th Grade	20.19%
High School Graduate or GED	20.80%
Some College or Associates Degree	24.47%
College Graduate or Above	16.74%
Income Level	
Less than $25,000	28.07%
$25,000 - $44,999	23.40%
$45,000 - $64,999	17.25%
$65,000 or More	31.28%
Health Variables	**Sample**
Blood Pressure	
Normal	49.70%
Prehypertension	34.42%
High Blood Pressure Stage 1	11.43%
High Blood Pressure Stage 2 or Crisis	4.45%
Body Mass Index (BMI)	
Underweight or Normal weight	36.29%
Overweight	30.24%
Obese	33.47%
Covered by Health Insurance	
Yes	78.81%
No	21.19%
Diabetic	
Yes	8.78%
No	91.22%
Have Asthma	
Yes	14.13%
No	85.87%

TABLE 1. Descriptive Characteristics of the Sample (continued)

Have Allergies	
Yes	31.16%
No	68.84%
Physical Activity Variables	**Sample**
Hours watching TV in a day	
None	1.73%
≤ 1 Hour	30.94%
2 Hours	26.49%
3 Hours	17.39%
4 Hours	10.61%
≥5 Hours	12.84%
Hours using Computer in a day	
None	38.90%
≤1 Hour	42.74%
2 Hours	9.24%
3 Hours	3.53%
4 Hours	2.19%
≥5 Hours	3.41%

TABLE 2. Estimation Results of the Latent Variable Structural Equation Model

Variable	Coefficient	t-stat
Physical Activity Propensity (PAP)		
Age (base is 65 and above)		
12 – 19 years (Yes=1, No=0)	1.450	43.126
20 – 44 years (Yes=1, No=0)	0.587	20.192
45 – 64 years (Yes=1, No=0)	0.283	9.314
Health-conscious (HC)		
Education Status (base is 11th grade or less)		
High school (Yes=1, No=0)	0.035	2.607
Some college degree (Yes=1, No=0)	0.093	4.279
Bachelor or post-graduate degree (Yes=1, No=0)	0.327	13.402
Gender (base is male)		
Female (Yes=1, No=0)	0.177	5.847
Insurance Status (base is not-insured)		
Insured (Yes=1, No=0)	0.068	3.620
Correlation Between PAP & HC		
Correlation coefficient between PAP & HC latent constructs	0.122	11.247

TABLE 3. Estimation Results for Latent Variable Measurement Equation - Non-nominal Outcomes

Variable	Coefficient (t-stat)	Variable	Coefficient (t-stat)
Natural logarithmic of Walk Duration (Continuous variable)		**Natural logarithmic of Vigorous Activity Duration (Continuous variable)**	
Constant	-2.853(-54.024)	Constant	-0.901(-27.905)
Body mass index (base is obese)		*Body mass index (base is obese)*	
Normal-weight (Yes=1, No=0)	-0.277(-14.381)	Normal-weight or Overweight (Yes=1, No=0)	-1.819(-87.901)
Overweight (Yes=1, No=0)	3.090(28.877)	Allergy (Yes=1, No=0)	-0.089(-6.316)
Allergy (Yes=1, No=0)	-0.060(-5.843)	*Race and Gender combination*	
Asthma (Yes=1, No=0)	-0.078(-5.789)	Caucasian male (Yes=1, No=0)	-0.177(-9.684)
Race and Gender combination		African-American male (Yes=1, No=0)	0.056(3.307)
Hispanic male (Yes=1, No=0)	0.182(6.758)	Hispanic female (Yes=1, No=0)	-0.223(-11.078)
Caucasian male (Yes=1, No=0)	-0.189(-8.669)	Caucasian female (Yes=1, No=0)	-0.200(-9.978)
Hispanic female (Yes=1, No=0)	0.256(10.538)	African-American female (Yes=1, No=0)	-0.26(-10.715)
Caucasian female (Yes=1, No=0)	-0.293(-17.01)	*Latent Variable*	
Latent Variable		Physical activity propensity	1.322(62.375)
Health-conscious	1.938(44.752)	Variance	0.433(62.625)
Variance	1.000(fixed)	**Health Rating (Ordinal Variable)**	
Natural logarithmic of Bike Duration (Continuous variable)		Constant	2.122(68.763)
Constant	-2.151(-38.051)	*Body mass index (base is normal and overweight)*	
Latent Variable		Obese (Yes=1, No=0)	-0.429(-81.244)
Health-conscious	0.095(6.536)	Diabetic (Yes=1, No=0)	-0.671(-93.08)
Variance	1.000(fixed)	Asthma (Yes=1,No=0)	-0.207(-34.659)
Natural logarithmic of Moderate Activity Duration (Continuous variable)		Participates in any type of physical activity for 30	0.331(66.818)
Constant	-1.115(-26.803)	*Latent Variable*	
Body mass index (base is obese)		Health-conscious	0.183(33.731)
Normal-weight (Yes=1, No=0)	-2.072(-46.152)	*Threshold*	
Overweight (Yes=1, No=0)	-1.252(-29.555)	Threshold 1 (fair and good)	1.171(25.927)
Allergy (Yes=1, No=0)	-0.124(-4.948)	Threshold 2 (good and very good)	2.235(41.621)
Race and Gender combination		Threshold 3 (very good and excellent)	3.369(54.109)
Hispanic male (Yes=1, No=0)	-0.037(-1.032)	**Blood Pressure (Group Variable)**	
Caucasian male (Yes=1, No=0)	0.175(5.888)	Constant	1.091(73.438)
Caucasian female (Yes=1, No=0)	0.097(2.747)	*Body mass index (base is normal weight)*	
Latent Variable		Overweight (Yes=1, No=0)	0.069(42.906)
Physical activity propensity	1.160(44.919)	Obese (Yes=1, No=0)	0.101(68.630)
Variance	0.794(57.839)	Diabetic (Yes=1, No=0)	0.113(75.575)
		TV-hours (base: do not watches television)	
		Watches television for 0 – 3 hours per day	0.055(14.338)
		Watches television for 4 and more hours per day	0.091(23.229)
		Female (Yes=1, No=0)	-0.061(-61.594)
		Participates in any type of physical activity for 30 min	-0.047(-37.789)
		Asthma (Yes=1,No=0)	-0.010(-6.978)
		Variance	0.045(24.747)

TABLE 3. Estimation Results for Latent Variable Measurement Equation - Non-nominal Outcomes (Continued)

Variable	Coefficient (t-stat)	Variable	Coefficient (t-stat)
TV Hours (Group Variable)		**Number of Moderate Activity Episodes (Count variable)**	
Constant	2.133(52.042)	Constant	1.226(31.359)
Latent Variable		*Body mass index (base is obese)*	
Physical activity propensity	-0.198(-49.674)	Normal-weight (Yes=1, No=0)	-1.408(-52.948)
Variance	2.085(25.171)	Overweight (Yes=1, No=0)	-1.121(-42.073)
Computer Hours (Group Variable)		Allergy (Yes=1, No=0)	-0.079(-5.938)
Constant	-0.122(-24.277)	Asthma (Yes=1,No=0)	-0.084(-6.153)
Education status (base is high school or less)		*Race and Gender combination*	
Some college degree (Yes=1, No=0)	0.465(63.140)	Hispanic male (Yes=1, No=0)	-0.201(-11.196)
Bachelor or post-graduate degree (Yes=1, No=0)	0.586(56.336)	Caucasian male (Yes=1, No=0)	0.025(1.578)
Variance	2.225(32.597)	Hispanic female (Yes=1, No=0)	0.173(9.885)
Number of Walking Episodes (Count variable)		Caucasian female (Yes=1, No=0)	0.173(9.885)
Constant	-1.703(-18.035)	African-American male (Yes=1, No=0)	0.244(11.201)
Body mass index (base is obese)		*Latent Variable*	
Normal-weight (Yes=1, No=0)	-0.242(-14.977)	Physical activity propensity	0.820(81.009)
Overweight (Yes=1, No=0)	2.532(36.808)	Dispersion parameter	0.082(49.097)
Race and Gender combination		**Number of Vigorous Activity Episodes (Count variable)**	
Hispanic male (Yes=1, No=0)	0.090(2.176)	Constant	-0.426(-6.523)
Caucasian male (Yes=1, No=0)	-0.158(-4.368)	*Body mass index (base is obese)*	
African-American male (Yes=1, No=0)	0.109(2.808)	Normal-weight (Yes=1, No=0)	-1.539(-60.241)
Caucasian female (Yes=1, No=0)	-0.137(-3.798)	Overweight (Yes=1, No=0)	-1.495(-58.35)
African-American female (Yes=1, No=0)	-0.167(-4.688)	Allergy (Yes=1, No=0)	-0.023(-3.537)
Latent Variable		*Race and Gender combination*	
Health-conscious	2.337(68.932)	Hispanic male (Yes=1, No=0)	0.025(1.665)
Dispersion parameter	0.059(12.155)	Caucasian male (Yes=1, No=0)	-0.066(-5.205)
Number of Biking Episodes (Count variable)		African-American male (Yes=1, No=0)	0.225(17.851)
Constant	-0.144(-1.964)	Hispanic female (Yes=1, No=0)	0.085(5.831)
Race and Gender combination		Caucasian female (Yes=1, No=0)	-0.081(-6.387)
Hispanic male (Yes=1, No=0)	0.595(7.057)	African-American male (Yes=1, No=0)	0.032(1.999)
Caucasian male (Yes=1, No=0)	0.396(5.122)	*Latent Variable*	
African-American male (Yes=1, No=0)	0.500(5.908)	Physical activity propensity	1.688(82.873)
Hispanic female (Yes=1, No=0)	-0.828(-10.899)	Dispersion parameter	0.039(19.125)
Caucasian female (Yes=1, No=0)	-0.208(-2.799)		
African-American male (Yes=1, No=0)	-0.275(-3.612)		
Age (base is greater than 19 years old)			
12 – 19 years (Yes=1, No=0)	0.385(6.060)		
Latent Variable			
Health-conscious	0.150(29.227)		
Dispersion parameter	0.033(96.724)		

**TABLE 4. Estimation Results for Latent Variable Measurement Equation –
Nominal Outcome**

Body Mass Index (Base: Normal weight)	Overweight		Obese	
Variable	**Coefficient**	**t-stat**	**Coefficient**	**t-stat**
Constant	-1.135	-6.961	0.204	3.852
TV-hours (base is < 3 hours per day)				
More than 3 hours per day			0.239	7.712
Computer hours (base is < 3 hours per day)				
More than 3 hours per day			-0.036	-2.904
Allergy (Yes=0, No=0)			0.306	9.053
Asthma (Yes=0, No=0)			0.267	8.808
Income (base is less than 25,000)				
25,000 – 44,999	0.160	3.857		
45,000 and more	0.452	9.037		
Diabetic (Yes=1, No=0)	0.558	8.29	0.314	8.692
Race and Gender combination				
Hispanic female (Yes=1, No=0)	-1.124	-10.834	0.288	8.241
African-American male (Yes=1, No=0)	-0.250	-5.275	0.380	8.835
Hispanic male	-0.771	-9.701		
Effect of Latent Constructs on Nominal Outcome				
Variable	**Normal weight**		**Overweight**	
Physical activity propensity	2.630	10.239	2.781	10.740
Health-conscious			-5.488	-13.162
Error Difference Matrix				
	Overweight		**Obese**	
Overweight	1.00	(fixed)	0.17	1.70
Obese			0.03	1.72