MODELING THE INFLUENCE OF FAMILY, SOCIAL CONTEXT, AND SPATIAL PROXIMITY ON NON-MOTORIZED TRANSPORT MODE USE

Nazneen Ferdous
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Tel: 512-471-4535; Fax: 512-475-8744
Email: nazneen.ferdous@mail.utexas.edu

Ram M. Pendyala
Arizona State University
Department of Civil and Environmental Engineering
Room ECG252, Tempe, AZ 85287-5306
Tel: (480) 727-9164; Fax: (480) 965-0557
Email: ram.pendyala@asu.edu

Chandra R. Bhat*
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Tel: 512-471-4535; Fax: 512-475-8744
Email: bhat@mail.utexas.edu

Karthik C. Konduri
Arizona State University
Department of Civil and Environmental Engineering
Room ECG252, Tempe, AZ 85287-5306
Tel: (480) 727-9164; Fax: (480) 965-0557
Email: karthik.konduri@asu.edu

*corresponding author

ABSTRACT

This paper presents a joint model of walking and bicycling activity duration using a hazard based specification that recognizes the interval nature of time reported in activity-travel surveys. The model structure takes the form of a multilevel hazard-based model system that accounts for the range of interactions and spatial effects that might affect walking and bicycling mode use. In addition to the individual-specific factors, family (household-specific) interactions, social group (peer) influences, and spatial clustering effects are also considered as potential factors that contribute to heterogeneity in non-motorized transport mode use behavior. The model system presented is capable of accommodating grouped duration responses often encountered in activity-travel surveys. A composite marginal likelihood estimation approach is adopted to estimate parameters in a computationally tractable manner. The model system is applied to a survey sample drawn from the recent 2009 National Household Travel Survey in the United States. Model results show that there are significant unobserved family-level, social group, and spatial proximity effects that contribute to heterogeneity in walking and bicycling activity duration. The unobserved effects were also found to have a differential impact on bicycling activity duration, thus suggesting the need to treat and model walking and bicycling separately in transportation modeling systems.

1. INTRODUCTION

Transportation professionals are interested in the analysis of walk and bicycle mode use from a variety of perspectives, including the reduction of fuel consumption and greenhouse gas (GHG) emissions (*1*), enhancement of public health and well-being (*2*), and the design of livable urban spaces. Although there is considerable literature on non-motorized travel behavior, previous research in this domain is limited in its treatment of walk and bicycle mode choice in several ways. For example, several studies have examined overall physical activity participation among adults and children in the context of the built environment, but these studies do not explicitly separate walking and bicycling from other physically active episodes of participation (*e.g.*, *3*, *4*). Other studies have lumped walking and bicycling together into a single category of non-motorized mode use, without sufficiently recognizing that there may be trade-offs across the use of these two modes of transport and important differences in the factors that influence their use (*e.g.*, *5*). Yet other studies have examined walking or bicycling in isolation of the other, thus preventing the ability to model or understand the use of these non-motorized physically active modes in a comprehensive way. There are numerous studies exclusively dedicated to the study of the choice of walking (*e.g.*, *6-8*), and others that exclusively focus on bicycling (*e.g.*, *9-11*).

The importance of considering walking and bicycling mode use in a unified framework has not gone unrecognized in the literature. However, many of these studies have restricted their focus to examining walking and bicycling habits of either children/adolescents, particularly in the context of their travel to and from school (*e.g.*, *12*), or adults in the context of their commute or short-distance trip making (*e.g.*, *13, 14*). Ogilvie *et al.* (*15*), Pikora *et al.* (*16*), and Saelens *et al.* (*17*) provide more extensive reviews of studies in this topic area.

In general, past research considers specific demographic segments, and describes or models non-motorized mode use of individuals in isolation of their social, familial, and spatial context. Sener *et al.* (*18*) jointly considered physical activity participation of all members in a family, but their analysis was limited by the consideration of all physical activities together as a single choice. The current paper uses a hazard-based duration model structure consistent with the use of time allocation as a measure of non-motorized mode use. Specifically, a proportional hazard specification is employed to capture activity participation behavior of individuals (*19*, *20*). The model system recognizes the presence of individual-specific unobserved factors that can affect the amount of non-motorized mode use as a whole, as well as the amount of time specifically allocated to bicycling vis-à-vis walking. The model also incorporates the effects of unobserved common household-specific attributes that can influence walking and cycling activity durations of all individuals in a household. Similarly, social group-specific and spatial-cluster specific unobserved factors that impact walking and cycling activity durations are also included in the model system. The data used in this paper is drawn from the San Francisco Bay Area subsample of the 2009 National Household Travel Survey (NHTS) (*21*). The data set includes detailed attitudinal information on walking and bicycling making it particularly appealing for this study. The sample data set is further enhanced by merging variables describing built environment attributes from a variety of secondary data sources.

The design of a behavioral model system that accounts for these myriad effects is motivated by work in social ecology theory that considers the critical role played by social networks in shaping individual choices in a variety of domains. The spatial distribution of social relationships has been found to be an important aspect of the vitality and fabric of urban neighborhoods (*22*, *23*). A specification that captures these multiple effects and interactions leads to a multi-level cross-cluster structure that recognizes and preserves between-cluster heterogeneity. Further, the relations between walking and cycling activity durations within and across individuals are correlated through these various unobserved factors. For example, consider individuals from a "health conscious" household. Individuals from this household are likely to have a higher propensity to engage in walking and bicycling activities for longer time periods. Also, between these two activities, if the household has an intrinsic preference for walking over bicycling, then the participation durations of walking activity of all individuals in the household are likely to be affected. Thus, ignoring unobserved common household-specific factors and considering only observed exogenous variables may lead to inconsistent parameter estimates (*24*). This, in turn, can result in less accurate assessment of the impact of policy measures designed to promote walking and/or bicycling at individual as well as household levels.

The multivariate cross-cluster model system proposed in the current paper requires the evaluation of a more than thousand-dimensional integral (number of individuals in the data set multiplied by number of activity types). As this is computationally prohibitive, a composite marginal likelihood (CML) approach that requires no use of simulation techniques is employed for parameter estimation (*25*, *26*). The CML approach entails the development of a surrogate likelihood function that involves easy-to-compute, low-dimensional, marginal likelihoods. The CML estimates are consistent and asymptotically normally distributed, and the approach provides accurate inferential conclusions.

The rest of the paper is organized as follows. The detailed modeling methodology is presented in the next section. The third section provides a brief overview of the data used in the study, while the fourth section presents model estimation results. Concluding thoughts and directions for further research are offered in the fifth and final section.

2. THE MODEL STRUCTURE

This section presents the modeling methodology. In the interest of brevity, a few details of the CML approach are omitted in this paper. More complete details of the approach may be found in Bhat *et al.* (*26*).

2.1 Mathematical Formulation

Let $\lambda_{qijlm}(\tau)$ represent the hazard at continuous time τ of ending time investment in activity type m (m = 1, …, M) for the q^{th} (q = 1, 2, …, Q) individual belonging to household i (i = 1, 2, …, I), social cluster j (j = 1, 2, …, J), and spatial cluster l (l = 1, 2, …, L). That is, $\lambda_{qijlm}(\tau)$ represents the conditional probability that individual q will stop investing additional time in activity type m during an infinitesimally small time period after time τ, given that the individual has not yet stopped investing time in activity type m until time τ:

$$\lambda_{qijlm}(\tau) = \lim_{\Delta \to 0^+} \frac{P(\tau < T_{qijlm} < \tau + \Delta \mid T_{qijlm} > \tau)}{\Delta}, \tag{1}$$

where T_{qijlm} is the index representing the continuous time of participation in activity m for individual q belonging to household i, social cluster j, and spatial cluster l. Next, the hazard rate $\lambda_{qijlm}(\tau)$ may be written using a proportional hazard formulation as a function of a vector of covariates x_{qm} specific to individual q and activity type m:

$$\lambda_{qijlm}(\tau) = \lambda_{m0}(\tau)\exp(\beta_m' x_{qm} + \alpha_{qijlm} + \omega_{qm}), \tag{2}$$

where β_m is a vector of coefficients specific to activity m, α_{qijlm} is a scalar term associated with individual q, household i, social cluster j, spatial cluster l, and activity type m, and ω_{qm} is an unobserved idiosyncratic factor affecting the hazard for individual q and activity m (ω_{qm} may represent unobserved factors such as the q^{th} individual's intrinsic liking or aversion for activity type m). ω_{qm} is assumed to be independent of x_{qm} and α_{qijlm}, and normally distributed with a mean of zero (an innocuous normalization for identification purposes) and variance σ_m^2. Equation (2) represents the micro-level model for individual q in household i, belonging to social cluster j and spatial cluster l, participating in activity m. Next, the scalar term α_{qijlm} is allowed to vary across individuals, households, social groups, and spatial clusters in a higher-level macro-model:

$$\alpha_{qijlm} = \varsigma' h_{qijl} + v_q + u_i + u_{im} + w_j + w_{jm} + z_l + z_{lm}, \tag{3}$$

where h_{qijl} is a vector of observed variables specific to individual q or household i or social cluster j or spatial cluster l or to the combination of these higher level macro-units, ς is a corresponding parameter vector to be estimated, v_q is an individual-specific random term that captures unobserved variation across individuals in the hazard function for all activity types (v_q may include intrinsic individual-specific factors such as motivation for physical activity that affects the duration of participation of the individual in all types of walking and bicycling activities), u_i is a household-specific random term that captures unobserved variation across households in the hazard function for all activity types (u_i may include intrinsic household-specific lifestyle factors impacting all individuals in the household in their attitudes and perspectives toward all types of walking and bicycling activities), u_{im} is another household-specific random term that captures unobserved variation across households in the hazard function specific to activity type m (u_{im} includes intrinsic household-specific factors that makes individuals in a household more inclined to participate in specific types of physical activity such as bicycling), w_j and w_{jm} are similar social-cluster specific random terms, and z_l and z_{lm} are similar spatial-cluster specific random terms. Consider that the above random terms are realizations from independent and identically normally distributed terms across individuals (for v_q), across households (for u_i and u_{im}), across social clusters (for w_j and w_{jm}), and across spatial clusters (for z_l and z_{lm}). Thus, the distributions of the error terms are:

$v_q \sim N[0,\theta^2]$, $u_i \sim N[0,\mu^2]$, $u_{im} \sim N[0,\mu_m{}^2]$, $w_j \sim N[0,\eta^2]$, $w_{jm} \sim N[0,\eta_m{}^2]$, $z_l \sim N[0,\delta^2]$, and $z_{lm} \sim N[0,\delta_m{}^2]$

If one were to define $\gamma_m = (\beta_m', \varsigma')'$ and $d_{qijlm} = (x_{qm}', h_{qijl}')'$, then the micro- and macro-models of Equations (2) and (3) can be combined into a single equation as follows:

$$\lambda_{qijlm}(\tau) = \lambda_{m0}(\tau)\exp(\gamma_m' d_{qijlm} + v_q + u_i + u_{im} + w_j + w_{jm} + z_l + z_{lm} + \omega_{qm}), \tag{4}$$

The proportional hazard formulation of Equation (4) can be written equivalently in terms of the logarithm of the integrated hazard at continuous time T_{qijlm} as follows (*21*):

$$s^*_{qijlm} = -\ln \Lambda_0(T_{qijlm}) = -\ln \int_{\tau=0}^{T_{qijlm}} \lambda_{m0}(\tau) d\tau = \gamma'_m d_{qijlm} + v_q + u_i + u_{im} + w_j + w_{jm} + z_l + z_{lm} + \omega_{qm} + \varepsilon_{qm} \ , \quad (5)$$

ε_{qm} in the above equation occurs because of the intrinsic probabilistic nature of the hazard function. Further, when the relationship between the hazard function and covariates takes the proportional hazard form of Equation (4), it can be shown that ε_{qm} is standard extreme value distributed: $\text{Prob}(\varepsilon_{qm} < a) = G(a) = \exp[-\exp(-a)]$. In Equation (5), since each individual q is uniquely identified with a particular household i, social group j, and spatial cluster l, it is convenient from a presentation standpoint to suppress the indices i, j, and l in T_{qijlm} and d_{qijlm}. Thus, T_{qm} and d_{qm} will be used to represent T_{qijlm} and d_{qijlm} respectively.

Now, consider the case where time T_{qm} is unobservable on the continuous scale, but is observed in grouped (or discrete) intervals t_{qm}. In the empirical context of the current paper, this grouping is a result of individuals rounding off activity durations (often to the nearest fifth minute) when reporting time-use patterns in activity-travel surveys (*27*). The net result of such rounding is that there is clumping or "ties" in the data at durations of time that are integer multiples of five minutes. The presence of such ties renders usual parametric continuous baseline hazard models inappropriate, since these models use density function terms in the likelihood function that are appropriate only for continuous duration data. It is important to explicitly recognize the interval-level data arising from the grouping of underlying continuous times during the estimation process. To do so, consider k as an index for grouped time intervals (*i.e.*, t_{qm} = 0, 1, 2,…, k,…, K_m). Let $b_{m,k+1}$ be the upper bound on the continuous time scale corresponding to the grouped time interval k. Then, Equation (5) may be written in an equivalent grouped response form as follows:

$$s^*_{qijlm} = -\ln \Lambda_0(T_{qm}) = \gamma'_m d_{qm} + v_q + u_i + u_{im} + w_j + w_{jm} + z_l + z_{lm} + \omega_{qm} + \varepsilon_{qm}, \ \ t_{qm} = k \text{ if } \psi_{m,K_m-k} < s^*_{qm} < \psi_{m,K_m+1-k}, \quad (6)$$

where $\psi_{m,K_m+1-k} = -\ln \Lambda_0(b_{m,k})$ is the upper bound for interval k for activity type m $(\psi_{m,0} < \psi_{m,1} < \psi_{m,2} \cdots < \psi_{m,K_m+1};\ \psi_{m,0} = -\infty, \psi_{m,K_m+1} = +\infty)$.

In the above specification, if θ^2 (variance of v_q), μ^2 (variance of u_i), μ_m^2 (variance of u_{im}; $m = 1,...,M$), η^2 (variance of w_j), η_m^2 (variance of $w_{jm}; m = 1,...,M$), δ^2 (variance of z_l), and δ_m^2 (variance of $z_{lm}; m = 1,...,M$) are all simultaneously equal to zero, then it implies that there is no variation in the activity durations for different activity types based on unobserved factors that are specific to the individual, the household, the social cluster to which the individual belongs, and the spatial cluster to which the individual belongs. In this case, the cross-random grouped response (CRGR) model of Equation (6) collapses to the standard grouped response (SGR) model. The implication is that all unobserved heterogeneity is due to overall idiosyncratic factors associated with the propensity to participate in each activity type, and there are no common unobserved individual, household, social group, and spatial cluster factors impacting durations of participation in the activity types. Note also that the specification of Equation (6) generates a rich covariance pattern structure among the hazard functions for participation in different activity types. The (log) integrated hazards (LIHs) for any pair of activity types m and

m' ($m \neq m'$) for the same individual have a covariance of $\theta^2 + U_q\mu^2 + \eta^2 + \delta^2$, where $U_q = 1$ if the individual is in a household with more than one individual and zero otherwise. For two different individuals q and q', the covariance in the LIHs between any pairing of activity types m and m' across the two individuals is equal to $H_{qq'}\mu^2 + H_{qq'mm'}\mu_m^2 + R_{qq'}\eta^2 + R_{qq'mm'}\eta_m^2 + G_{qq'}\delta^2 + G_{qq'mm'}\delta_m^2$, where $H_{qq'} = 1$ if individuals q and q' are in the same household, $H_{qq'mm'} = 1$ if individuals q and q' are in the same household and m and m' are the same activity type, $R_{qq'} = 1$ if individuals q and q' are in the same social group, $R_{qq'mm'} = 1$ if individuals q and q' are in the same social group and m and m' are the same activity type, $G_{qq'} = 1$ if individuals q and q' are in the same spatial cluster, and $G_{qq'mm'} = 1$ if individuals q and q' are in the same spatial cluster and m and m' are the same activity type. The indicator variables above take the value of zero otherwise.

2.2 Estimation Approach

Let y_{qm} be the q^{th} individual's observed activity participation time (obtained in the grouped intervals) in activity type m. The conditional likelihood function for individual q's participation duration in activity type m (conditional on $v_q, u_i, u_{im}, w_j, w_{jm}, z_l, z_{lm}$ and ω_{qm}) can be written as:

$$L_{qm}\left| v_q, u_i, u_{im}, w_j, w_{jm}, z_l, z_{lm}, \omega_{qm} = G[\psi_{K_m+1-y_{qm}} - B_{qm}] - G[\psi_{K_m-y_{qm}} - B_{qm}] \right.$$

where $B_{qm} = \gamma_m' d_{qm} + v_q + u_i + u_{im} + w_j + w_{jm} + z_l + z_{lm} + \omega_{qm}$.

The likelihood function unconditional on ω_{qm} is:

$$L_{qm}\left| v_q, u_i, u_{im}, w_j, w_{jm}, z_l, z_{lm} = \int_{\omega_{qm}} (G[\psi_{K_m+1-y_{qm}} - B_{qm}] - G[\psi_{K_m-y_{qm}} - B_{qm}])dF(\omega_{qm}) \right.,$$

where $F(\omega_{qm})$ is the univariate cumulative normal distribution function corresponding to ω_{qm}. The likelihood function of the entire sample cannot be broken down as the product of the likelihood functions for each individual's choices of grouped time interval for each activity m, because the underlying latent values s^*_{qijlm} are correlated across individuals q and activities m (due to the presence of the $v_q, u_i, u_{im}, w_j, w_{jm}, z_l,$ and z_{lm} error terms). Further, since the various clusters are not hierarchical (*i.e.*, one cluster is not nested within the other), the analyst needs to consider the entire set of $Q \times M$ observations (q = 1, 2, …, Q; m = 1, …, M) as a single cluster in developing the likelihood function. To accomplish this, stack the s^*_{qijlm} values together vertically in the vector s^*, and let the implied variance-covariance of s^* due to the $v_q, u_i, u_{im}, w_j, w_{jm}, z_l,$ and z_{lm} (but not considering ω_{qm} and ε_{qm}) error terms be Ω. Thus Ω is a $[(M \times Q) \times (M \times Q)]$ variance-covariance matrix whose elements are parameterized based on $\theta^2, \mu^2, \mu_m^2, \eta^2, \eta_m^2, \delta^2,$ and δ_m^2. Define a multivariate normally distributed variable vector $g \sim MVN(0, \Omega)$. Then the likelihood function may be written as:

$$L = \int_g \prod_{q=1}^{q} \prod_{m=1}^{M} (L_{qm} | v_q, u_i, u_{im}, w_j, w_{jm}, z_l, z_{lm}) dF_{M \times Q}(g|\Omega)$$

The likelihood function above entails the evaluation of an integral of the order of $(Q \times M)$. The usual simulation techniques become impractical, if not infeasible, to evaluate such a multidimensional integral for even small to moderate Q. For this reason, the composite marginal likelihood (CML) technique is adopted in the current paper (see *26*, *28*). Specifically, in the current paper, a pairwise marginal likelihood estimation approach which is based on forming a surrogate likelihood function that compounds pairs of individual-activity type combinations is used. Then, by maximizing this surrogate (log) likelihood function, a consistent estimator of all relevant parameters characterizing the original high dimensional distribution is obtained. Let the parameter vector to be estimated be represented as:

$$\kappa = (\gamma_1', ..., \gamma_M';\ \psi_1', ..., \psi_M'; \sigma_1, ..., \sigma_M; \mu_1, ..., \mu_M; \eta_1, ..., \eta_M; \delta_1, ..., \delta_M, \theta, \mu, \eta, \delta)',$$

where $\psi_m = (\psi_{m,1}, \psi_{m,2}, ..., \psi_{m,K_m})'$. The pairwise marginal likelihood function includes two main components – one component that represents the likelihood of all pairs of activity type combinations within individuals, and the second component that represents the likelihood of pairs of individual-activity type combinations across individuals:

$$L_{CML}(\kappa) = \left[\prod_{q=1}^{Q} \prod_{m=1}^{M-1} \prod_{m'=m+1}^{M} \int_{\varepsilon_{qm'}=-\infty}^{+\infty} \int_{\varepsilon_{qm}=-\infty}^{+\infty} \left[\Phi_2(m_{y_{qm}+1}, m_{y_{qm'}+1}, \theta_{qmm'}) - \Phi_2(m_{y_{qm}}, m_{y_{qm'}+1}, \theta_{qmm'}) - \Phi_2(m_{y_{qm}+1}, m_{y_{qm'}}, \theta_{qmm'}) + \Phi_2(m_{y_{qm}}, m_{y_{qm'}}, \theta_{qmm'}) \right] dF(\varepsilon_{qm}) dF(\varepsilon_{qm'}) \right] \times$$
$$\left[\prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} \prod_{m=1}^{M} \prod_{m'=1}^{M} \int_{\varepsilon_{q'm'}=-\infty}^{+\infty} \int_{\varepsilon_{qm}=-\infty}^{+\infty} \left[\Phi_2(m_{y_{qm}+1}, m_{y_{q'm'}+1}, \theta_{qq'mm'}) - \Phi_2(m_{y_{qm}}, m_{y_{q'm'}+1}, \theta_{qq'mm'}) - \Phi_2(m_{y_{qm}+1}, m_{y_{q'm'}}, \theta_{qq'mm'}) + \Phi_2(m_{y_{qm}}, m_{y_{q'm'}}, \theta_{qq'mm'}) \right] dF(\varepsilon_{qm}) dF(\varepsilon_{q'm'}) \right] \quad (7)$$

where

$$m_{y_{qm}} = \left[\frac{\psi_{K_m - y_{qm}} - \gamma_m' d_{qm} - \varepsilon_{qm}}{\sqrt{\theta^2 + U_q \mu^2 + \eta^2 + \delta^2 + U_q \mu_m^2 + \eta_m^2 + \delta_m^2 + \sigma_m^2}} \right],$$

$$\theta_{qmm'} = \left[\frac{\theta^2 + U_q \mu^2 + \eta^2 + \delta^2}{\left(\sqrt{\theta^2 + U_q \mu^2 + \eta^2 + \delta^2 + U_q \mu_m^2 + \eta_m^2 + \delta_m^2 + \sigma_m^2} \right) \left(\sqrt{\theta^2 + U_q \mu^2 + \eta^2 + \delta^2 + U_q \mu_{m'}^2 + \eta_{m'}^2 + \delta_{m'}^2 + \sigma_{m'}^2} \right)} \right], m \neq m'$$

and

$$\theta_{qq'mm'} = \left[\frac{H_{qq'} \mu^2 + H_{qq'mm'} \mu_m^2 + R_{qq'} \eta^2 + R_{qq'mm'} \eta_m^2 + G_{qq'} \delta^2 + G_{qq'mm'} \delta_m^2}{\left(\sqrt{\theta^2 + U_q \mu^2 + \eta^2 + \delta^2 + U_q \mu_m^2 + \eta_m^2 + \delta_m^2 + \sigma_m^2} \right) \left(\sqrt{\theta^2 + U_{q'} \mu^2 + \eta^2 + \delta^2 + U_{q'} \mu_{m'}^2 + \eta_{m'}^2 + \delta_{m'}^2 + \sigma_{m'}^2} \right)} \right], q \neq q'.$$

In the CML function above, $F(.)$ is the univariate cumulative standard type I extreme value distribution. In Equation (7), the bi-dimensional integrations can be carried out using quadrature techniques or simulation techniques. However, an alternative is to use the normal scale mixture (NSM) representation of the extreme value distribution. In this approach, the non-normality of the error term ε_{qm} is removed by replacing it with a weighted mixture of normally distributed variables (*29*, *30*). In the interest of brevity, the details of this mixing approach are suppressed in this paper, but are available in Bhat (*30*).

The pairwise estimator $\hat{\kappa}_{CML}$ offers parameters that are consistent and asymptotically normal distributed with asymptotic mean κ and covariance matrix given by the inverse of Godambe's (*31*) sandwich information matrix $G(\kappa)$ (*32*):

$$F(\kappa) = [G(\kappa)]^{-1} = [H(\kappa)]^{-1} J(\kappa) [H(\kappa)]^{-1}, \text{ where}$$

$$H(\kappa) = E\left[-\frac{\partial^2 \log L_{CML}(\kappa)}{\partial\kappa\partial\kappa'} \right] \text{ and}$$

$$J(\kappa) = E\left[\left(\frac{\partial \log L_{CML}(\kappa)}{\partial\kappa} \right) \left(\frac{\partial \log L_{CML}(\kappa)}{\partial\kappa'} \right) \right]$$

The $H(\kappa)$ matrix can be estimated in a straightforward manner using the Hessian of the negative of $\log L_{CML}(\kappa)$, evaluated at the CML estimate $(\hat{\kappa}_{CML})$. However, the estimation of the $J(\kappa)$ matrix is more difficult. In the current paper, pure Monte Carlo computation is used to estimate the $J(\kappa)$ matrix. In this approach, B data sets (T^1, T^2,..., T^B) are generated where each dataset T^b (b = 1, 2, …, B) is a $Q \times M$ matrix of the dependent variables generated using the exogenous variables and the CML estimates $(\hat{\kappa}_{CML})$. Once these datasets are generated, the estimate of $J(\kappa)$ is given by:

$$\hat{J}(\hat{\kappa}) = \frac{1}{B} \sum_{b=1}^{B} \left[\left(\frac{\partial \log L_{CML}(\hat{\kappa})}{\partial\kappa} \right)_{T^b} \left(\frac{\partial \log L_{CML}(\hat{\kappa})}{\partial\kappa'} \right)_{T^b} \right]$$

3. DATA

In the current study, data collected for households drawn from Marin, Solano, and Sonoma counties in the San Francisco Bay Area under the National Household Travel Survey (NHTS) add-on program is used. Data from these three counties was used because the California NHTS add-on survey included detailed information on weekly walking and bicycling activity duration for all individuals 5 years of age or above and because the authors have access to extensive secondary data on built environment attributes for these locations. These built environment attributes and spatial context variables were appended to the survey sample to form a rich data set suitable to the analysis of walking and bicycling choice behavior.

The sample selected for analysis includes only individuals at least 5 years old who reported participating in at least one activity of interest (*i.e.*, either walking or bicycling) over a period of one week. The continuous walking and bicycling activity durations were divided into grouped intervals and indexed appropriately (as per the description in the previous section). Following this initial sample selection process, a series of steps were undertaken to determine associations across individuals and households. First, an indicator variable was generated to denote individuals of the same family. Then, a number of demographic factors such as age, household structure, and household income were used to define social grouping. A preliminary exploratory analysis suggested that using age to define social groups would yield the best model specification. It was found that age is the most significant variable in identifying homogeneous clusters of individuals who have very similar levels of walking and bicycling duration. All individuals were grouped into one of nine social groups based on age. Following this, the residential location of each household was geo-located to a TAZ (traffic analysis zone); thus, all households that reside in a TAZ belong to a spatial cluster.

The survey collected information about respondents' attitudes towards bicycling and walking and the factors that impact the amount of bicycling and walking they undertake (*19*). A factor analysis, using principal components estimation and varimax rotation, was performed to reduce the data and define a more compact set of influential factors. These factors generally captured lifestyle characteristics, neighborhood (built environment) characteristics, motorized traffic characteristics, and perceptions of personal safety (details of the factor analysis are available from the authors).

The final sample includes 882 individuals from 561 households. Of these individuals, 96.1% participate in some walking activity and 18.9% participate in some bicycling activity over a period of one week. Individuals who participate in these activities spend, on average, 204 minutes and 130 minutes per week in walking and bicycling activity respectively. Table 1 provides information on walking and bicycling activity durations for individuals who participate in these activities. The lengths of the discrete periods used in estimation (presented in the third column) increase for larger activity durations until termination for all individuals (except for the first period for walking activity which is 10 minutes long). The number of discrete periods used for walking is higher than for bicycling because of the more extensive number of individuals walking in the sample, thus providing adequate number of individuals in finer time periods. For the final discrete period, all spells longer than 840 minutes for walking and 240 minutes for bicycling are collapsed to a single period.

The discrete-period sample hazards (the sixth column) are estimated using the Kaplan-Meier non-parametric estimator (*33*). These discrete-period sample hazards cannot be compared directly across periods due to variation in the length of time periods. The discrete-period sample hazards are converted to continuous-time sample hazards under the assumption that the hazard is constant within each period *k*. Thus, the continuous-time sample hazard $\hat{\lambda}_{m0}(k)$ can be estimated as follows:

$$\hat{\lambda}_{m0}(k) = -\frac{\ln(1-\hat{\lambda}^{*}_{m0}(k))}{\Delta t(k)},$$

where $\hat{\lambda}^{*}_{m0}(k)$ is the discrete-period sample hazard in period *k* and $\Delta t(k)$ is the length of the period *k*. The continuous-time sample hazards are plotted in Figure 1, and show that the sample hazards are higher for bicycling activity duration compared to walking activity duration in the first 45 minutes. This implies that individuals who participate in walking activity tend to commit a certain minimum amount of time to pursue this activity. Also, walking activity duration hazards exhibit more widespread "peaks" than bicycling duration hazards. This indicates a more even distribution of walking activity durations across participating individuals in comparison to bicycling activity durations. The hazard function for walking duration exhibits three highest "peaks" at time periods containing 1 hour, 2-hour, and 3-hour walking activity durations per week. Other "peaks" in the plot of hazard function for walking can be observed at multiples of 30 minutes intervals. A similar trend, but to a lesser degree can be observed in the plot of the hazard function for bicycling duration. This pattern of hazard functions highlights the discrete interval nature of reporting of the underlying continuous time variable and the need to adopt an appropriate framework that can explicitly recognize this feature. The model system proposed in the current paper incorporates this ability.

4. EMPIRICAL ANALYSIS

Estimation results for the final model specification are presented in Table 2. The final specification includes some variables that are not statistically significant at the usual 5% level of significance because the effects of these variables are intuitive, and have the potential to guide future research that may have the benefit of larger sample sizes. The coefficients in the table indicate the effects of variables on the duration hazard for walking and cycling activity. A positive (negative) coefficient implies that the corresponding explanatory variable increases (decreases) the hazard rate and decreases (increases) the activity duration.

In general, the variables offer plausible behavioral interpretations. Males tend to allocate more time to bicycling, as do those who are employed. Those who are part-time employed allocate more time to walking than do other groups. It is possible that employed individuals walk for utilitarian purposes during the workday (say, for lunch or to accomplish quick errands) or before or after work to decompress. Individuals in "couple" households spend more time walking and bicycling; these individuals probably do not have the constraints associated with child care in the home, and find walking and bicycling a relaxing joint activity. However, it is found that individuals in households with children that are older (11-15 years) allocate more time to bicycling, presumably because these children are able to pursue outdoor activities and ride bicycles together with parents.

The perceived absence of walkable attractions and having busy lifestyles deter walking as evidenced by the positive coefficient associated with this attitudinal variable. Similarly, the absence of walk-friendly environment and facilities deters people from spending time on walking activities. Likewise, several bicycling related factors deter time allocation to bicycling. Busy lifestyles and the unavailability of bicycle paths/trails, inconvenience in terms of carrying things and lack of paved bicycle facilities, and perceived (lack of) safety are all associated with positive coefficients. These results suggest that there are myriad factors that affect the time allocation to walking and bicycling activities. On the one hand, busy lifestyles deter individuals from allocating time to walking and bicycling. This may not be a factor that can be easily manipulated by policy-makers, but it may be possible to ease lifestyle constraints by providing flexible work schedules and telecommuting options. However, more directly related to transportation planning and design is the finding that unavailability and poor perceived quality of walking and bicycling infrastructure, and lack of walking "attractions" (*i.e.*, appealing destinations), are clearly having an adverse impact on the amount of time spent walking and bicycling. Planners, designers, and policy makers may be able to enhance walking and bicycling use by addressing these issues.[1]

Of special relevance in the context of this paper is the bottom half of Table 2 which presents estimates of the heterogeneity parameters. These parameters capture heterogeneity effects due to unobserved individual, household, social group, or spatial neighborhood factors. Heterogeneity, *i.e.*, differences in behavior across individuals, may arise due to individual-specific effects, interaction effects, or agglomeration effects. Individual-specific effects constitute intrinsic individual attitudes, perceptions, beliefs, and values that affect walking and bicycling duration. Household-specific effects constitute family influences that are not typically measured or observed in surveys. Lifestyle choices at the household level, family-member interactions, and parental influence may impact the time spent by individuals for walking and bicycling. The argument can be extended to the context of social group and spatial clusters, except that these effects may be a combination of interaction effects and agglomeration effects. Individuals of different age groups may interact with their peers (outside the household), thus bringing about peer (social network) influence that impacts walking and bicycling activity. Finally, households (individuals) may choose to reside in neighborhoods that support their

[1] None of the many built environment variables considered entered into the final model specification. This is because the attitudinal variables potentially capture the effects of the built environment.

lifestyle and travel preferences, thus producing agglomeration effects where spatial clustering of households with similar walking and cycling propensity may occur.

An examination of these parameter estimates shows that heterogeneity effects are generally statistically significant at all levels of clustering. The overall heterogeneity parameter in each level is representative of heterogeneity in the use of non-motorized modes of travel as a whole, while the activity-specific heterogeneity parameter denotes whether there are unobserved factors that differentially impact walking and bicycling duration across individuals. At the individual-specific level, the overall heterogeneity parameter is significant indicating that there are significant unobserved individual-specific factors that affect non-motorized mode use as a whole. At the activity-specific level, it is found that such unobserved factors significantly impact bicycling activity duration. All other levels exhibit similar findings. Overall heterogeneity effects are statistically significant at all levels, although the significance is weaker in the context of household-specific heterogeneity. There are clear indications from the model results that significant social influence effects and neighborhood clustering effects exist in the amount of non-motorized mode use by individuals. However, when one examines the activity-specific heterogeneity parameters, it appears that the heterogeneity effects are more pronounced (significant) in bicycle activity duration as opposed to walking duration. At all levels, the activity-specific heterogeneity parameter is statistically insignificant for walking activity duration. In summary, the results show that heterogeneity exists, that heterogeneity exists due to interaction and agglomeration effects at multiple levels, and that the heterogeneity specific to bicycling activity duration is statistically significant. When model estimation results reported in this paper were compared against those obtained from a model specification in which heterogeneity due to a variety of effects was ignored, it was found that coefficient estimates on socio-economic and attitudinal variables noticeably differed in magnitude (although the signs on the coefficients were identical).

Baseline hazard plots were generated (not shown in the interest of brevity) and compared against the sample hazard plots shown in Figure 1. As in the case of the sample hazard rates, baseline hazards were calculated under the assumption that the hazard remains constant within each discrete time interval. The baseline hazard functions are found to be non-monotonic and characterized by multiple peaks, similar to the sample hazard functions. This finding clearly indicates that non-parametric hazard functions are preferred over parametric specifications for analyzing walking and bicycling activity durations. Another interesting finding is that there are clear differences between the baseline hazards and the sample hazards. For walking activity duration, the baseline hazard increases with increasing activity duration, while the sample hazard decreases with increase in activity duration (except for the first 45 minutes). For bicycling activity duration, the baseline hazard and sample hazard were found to be more similar in profile; however, the baseline hazard shows more distinct peaks than the sample hazard. These differences between the baseline and sample hazards suggest that it is important to recognize variations in activity durations due to both observed and unobserved factors using approaches such as that adopted in this paper.

The log-composite likelihood value for the fully specified independent grouped response probit model (IGRP) (that is, independent grouped response probit models for each activity type) at convergence is –6,647,007.8 and that for the fully specified multi-level cross-random grouped response probit model (MCGRP) presented in the table is –5,524,133.6. The composite likelihood ratio test (*CLRT*) statistic for comparing the MCGRP model with the IGRP model is 2,245,748.4. However, the *CLRT* statistic does not have the standard χ^2 asymptotic distribution under the null hypothesis, as in the case of the regular maximum likelihood inference procedure. Other measures can be used to determine whether the MCGRP model form is statistically superior to the IGRP model form. The t-statistics on θ, σ_m, μ, μ_m, η, η_m, δ, and δ_m parameter estimates are statistically significant, indicating that the MCGRP model is likely to be superior to the IGRP model which omits these statistically significant parameters. Further, one may compute $\bar{\rho}_c^2$ in the composite marginal likelihood approach for the MCGRP model and the IGRP models as $\bar{\rho}_c^2 = 1 - [(\log L_{CML}(\hat{\kappa}) - N) / \log L_{CML}(T)]$, where $\log L_{CML}(\hat{\kappa})$ is the composite marginal log-likelihood at convergence, N is the number of model parameters excluding the thresholds, and $\log L_{CML}(T)$ is the log-likelihood with only thresholds in the model. The value of $\bar{\rho}_c^2$ for the IGRP model and the MCGRP model are 0.17 and 0.31 respectively once again indicating that the IGRP model may be rejected in favor of the MCGRP model.

5. SUMMARY AND CONCLUSIONS

This paper offers a framework and methodology for modeling the time spent walking and bicycling by individuals, while explicitly recognizing heterogeneity arising from individual-specific factors, family or intra-household interactions, social group or peer influences, and spatial clustering effects. In the United States, walking and bicycling activity is often a lifestyle preference that is linked closely to personal and household attitudes, beliefs, values, and perceptions. These attitudes and preferences are likely to be shaped by not only one's own individual-specific beliefs, but also influences of other household members, social peers, and neighborhood elements.

In this study, the time allocated to walking and bicycling activity over a period of one week is modeled jointly using a hazard model specification, thus providing the ability to examine how effects of various factors differentially impact walking vis-à-vis bicycling. The methodology adopted in this paper is capable of accommodating grouped responses that typically occur in activity-travel survey data sets wherein durations (start and end times) are rounded to the nearest fifth minute. The multilevel cross-random model structure is presented in detail in the paper together with a model estimation approach that overcomes the challenge associated with evaluating a thousand-dimension integral of a multivariate density function. The composite marginal likelihood (CML) approach provides a tractable, easy to implement way to estimate parameters by transforming the large multidimensional integral to a low-dimensional integral.

The model is estimated on a survey sample data set derived from the California add-on of the United States National Household Travel Survey (NHTS) conducted in 2009. The subsample specific to three counties in the San Francisco Bay Area is extracted and analyzed in this paper. The continuous time hazard functions suggest that individuals tend to be more uniform in the allocation of time to walking than to bicycling. Higher hazards for bicycling at small duration (up to 45 minutes) suggest that individuals tend to commit a certain minimum amount of time to walking, thus reducing the hazard in those initial periods. The model estimation results show standard individual and household demographic and socio-economic variables impact walking and bicycling activity duration. More importantly, however, there are numerous attitudinal factors and perceptions that affect walking and bicycling activity duration. In addition to busy lifestyles and such constraints, it is found that perceptions of poor walking and bicycling infrastructure and concerns about safety adversely impact the amount of walking and bicycling undertaken by individuals. These findings are all consistent with expectations and point to the need for professionals and policy makers to consider neighborhood designs, land use configurations, and infrastructure investments that alleviate the concerns and enhance perceptions of bicycling and walking convenience and safety.

Another important finding in this study is the significance of heterogeneity effects at multiple levels in the determination of non-motorized mode use, largely attributable to heterogeneity in bicycle activity duration. Travel demand model systems, with virtually no exception, ignore many of the (unobserved) interaction effects, social context, and spatial clustering effects that bring about heterogeneity in behavior. In this paper, it is found that unobserved individual specific factors, intra-household interaction effects, social peer group influence, and spatial clustering effects are all significant determinants walking and bicycling activity duration. The finding that family effects are important suggests that public education campaigns targeted at parents may bring about changes in the non-motorized mode use of children due to "family" effects. Similarly, social peer group influences and spatial clustering effects should not be ignored in modeling non-motorized mode use. People tend to be influenced by the behavior of individuals they associate with (in this paper, association based on age group was found to offer a plausible specification), and households tend to locate in spatial clusters (zones or neighborhoods) consistent with their lifestyle and travel preferences. Integrated land use – transport model systems able to capture such effects through enhanced model specifications are likely to offer more accurate policy predictions that better inform decision makers.

ACKNOWLEDGEMENTS
The authors would like to acknowledge support from the Southern California Association of Governments for facilitating this research. The authors thank the California Department of Transportation for providing the add-on survey sample data set used in this study, and the Metropolitan Transportation Commission for providing secondary network and land use data sets. Finally, the authors acknowledge support from the Sustainable Cities Doctoral Research Initiative at the Center for Sustainable Development at The University of Texas at Austin. The authors are solely responsible for all errors and omissions.

REFERENCES

1. Gotschi, T. and K. Mills. Active Transportation for America, 2008. Available at: http://www.railstotrails.org/resources/documents/whatwedo/atfa/ATFA_20081020.pdf
2. Timperio, A., D. Crawford, A. Telford, and J. Salmon. Perceptions about the Local Neighborhood and Walking and Cycling among Children. *Preventive Medicine*, Vol. 38, No. 1, 2004, pp. 39-47.
3. Badland, H. and G. Schofield. Transport, Urban Design, and Physical Activity: An Evidence-based Update. *Transportation Research Part D*, Vol. 10, No. 3, 2005, pp. 177-196.
4. Frank, L.D., T.L. Schmid, J.F. Sallis, J. Chapman, and B.E. Saelens. Linking Objectively Measured Physical Activity with Objectively Measured Urban Form: Findings from SMARTRAQ. *American Journal of Preventive Medicine*, Vol. 28, No. 2, 2005, pp. 117-125.
5. Cao, X., P.L. Mokhtarian, and S.L. Handy. The Relationship between the Built Environment and Nonwork Travel: A Case Study of Northern California. *Transportation Research Part A*, Vol. 43, No. 5, 2009, pp. 548-559.
6. McGinnn, A.P., K.P. Evenson, A.H. Herring, S.L. Huston, and D.A. Rodriguez. Exploring Associations between Physical Activity and Perceived and Objective Measures of the Built Environment. *Journal of Urban Health*, Vol. 84, No. 2, 2007, pp. 162-184.
7. Forsyth, A., J.M. Oakes, B. Lee, and K.H. Schmitz. The Built Environment, Walking, and Physical Activity: Is the Environment more Important to Some People than Others? *Transportation Research Part D*, Vol. 14, No. 1, 2009, pp. 42-49.
8. Agrawal, A.W. and P. Schimek. Extent and Correlates of Walking in the USA. *Transportation Research Part D*, Vol. 12, No. 8, 2007, pp. 548-563.
9. Rietveld, P. and V. Daniel. Determinants of Bicycle Use: Do Municipal Policies Matter? *Transportation Research Part A*, Vol. 38, No. 7, 2004, pp. 531-550.
10. Hunt, J.D. and J. E. Abraham. Influences on Bicycle Use. *Transportation*, Vol. 34, No. 4, 2007, pp. 453-470.
11. Xing, Y., S.L. Handy, and P.L. Mokhtarian. Factors Associated with Proportions and Miles of Bicycling for Transportation and Recreation in Six Small US Cities. *Transportation Research Part D*, Vol. 15, No. 2, 2010, pp. 73-81.
12. Cooper, A.R., N. Wedderkopp, H. Wang, L.B. Andersen, K. Froberg, and A.S. Page. Active Travel to School and Cardiovascular Fitness in Danish Children and Adolescents. *Medicine & Science in Sports & Exercise*, Vol. 38, No. 10, 2006, pp. 1724-1731.
13. Plaut, P.O. Non-motorized Commuting in the US. *Transportation Research Part D*, Vol. 10, No. 5, 2005, pp. 347-356.
14. Kim, S. and G.F. Ulfarsson. Curbing Automobile use for Sustainable Transportation: Analysis of Mode Choice on Short Home-based Trips. *Transportation*, Vol. 35, No. 6, 2008, pp. 723-737.
15. Ogilvie, D., M. Egan, V. Hamilton, and M. Petticrew. Promoting Walking and Cycling as an Alternative to Using Cars: Systematic Review. *British Medical Journal*, Vol. 329, No. 7469, 2004, pp. 763-766.
16. Pikora, T., B. Giles-Corti, F. Bull, K. Jamrozik, and R. Donovan. Developing a Framework for Assessment of the Environmental Determinants of Walking and Cycling. *Social Science & Medicine*, Vol. 56, No. 8, 2003, pp. 1693-1703.

17. Saelens, B.E., J.F. Sallis, and L.D. Frank. Environmental Correlates of Walking and Cycling: Findings from the Ttransportation, Urban Design, and Planning Literatures. *Annals of Behavioral Medicine*, Vol. 25, No. 2, 2003, pp. 80-91.
18. Sener, I.N., N. Eluru, and C.R. Bhat. On Jointly Analyzing the Physical Activity Participation Levels of Individuals in a Family Unit Using a Multivariate Copula Framework. *Journal of Choice Modelling*, Vol. 3, No. 3, 2010, pp. 1-38.
19. Hensher, D.A. and F.L. Mannering. Hazard-based Duration Models and Their Application to Transport Analysis. *Transport Reviews*, Vol. 14, No. 1, 1994, pp. 63-82.
20. Bhat, C.R. and A.R. Pinjari. Duration Modeling. In D.A. Hensher and K.J. Button (eds), *Handbook of Transport Modelling*, 2nd edition, Elsevier Science, 2008, pp. 105-132.
21. NHTS Telephone (CATI) Questionnaire Extended Interview, 2009. Available at: http://nhts.ornl.gov/2009/pub/ExtendedInterview.pdf.
22. Greenbaum, S.D. and P.E. Greenbaum, The Ecology of Social Networks in Four Urban Neighborhoods. *Social Networks*, Vol. 7, No. 1, 1985, pp. 47-76.
23. Röper, A., B. Völker, and H. Flap. Social Networks and Getting a Home: Do Contacts Matter? *Social Networks*, Vol. 31, No. 1, 2009, pp. 40-51.
24. Bhat, C.R. A Multi-level Cross-classified Model for Discrete Response Variables. *Transportation Research Part B*, Vol. 34, No. 7, 2000, pp. 567-582.
25. Varin, C. On Composite Marginal Likelihoods. *Advances in Statistical Analysis*, Vol. 92, No. 1, 2008, pp. 1-28.
26. Bhat, C.R., I.N. Sener, and N. Eluru. A Flexible Spatially Dependent Discrete Choice Model: Formulation and Application to Teenagers' Weekday Recreational Activity Participation. *Transportation Research Part B*, Vol. 44, No. 8-9, 2010, pp. 903-921.
27. Bhat, C.R. A Generalized Multiple Durations Proportional Hazard Model with an Application to Activity Behavior During the Work-to-home Commute. *Transportation Research Part B*, Vol. 30, No. 6, 1996, pp. 465-480.
28. Varin, C., N. Reid, and D. Firth. An Overview of Composite Marginal Likelihoods. Working Paper 09-45, Centre for Research in Statistical Methodology (CRiSM), University of Warwick, U.K, 2010.
29. Choy, S.T.B. and J.S.K. Chan. Scale Mixtures Distributions in Statistical Modeling. *Australian & New Zealand Journal of Statistics*, Vol. 50, No. 2, 2008, pp. 135-146.
30. Bhat, C.R. The Maximum Approximated Composite Marginal Likelihood (MACML) Estimation of Multinomial Probit-Based Unordered Response Choice Models. Technical paper, Department of Civil, Architectural & Environmental Engineering, The University of Texas at Austin, 2010.
31. Godambe, V. P. An Optimum Property of Regular Maximum Likelihood Estimation. *The Annals of Mathematical Statistics*, Vol. 31, No. 4, 1960, pp. 1208-1211.
32. Zhao, Y. and H. Joe. Composite Likelihood Estimation in Multivariate Data Analysis. *The Canadian Journal of Statistics*, Vol. 33, No. 3, 2005, pp.335-356.
33. Kiefer, N.M. Economic Duration Data and Hazard Functions. *Journal of Economic Literature*, Vol. 26, No. 2, 1988, pp. 646-679.

LIST OF TABLES

TABLE 1 Walking and Bicycling Activity Durations and the Discrete Period Sample Hazards

TABLE 2 Model Estimation Results

LIST OF FIGURES

FIGURE 1 Continuous time sample hazard functions for walking and bicycling durations.

TABLE 1 Walking and Bicycling Activity Durations and the Discrete Period Sample Hazards

Discrete time period k[a]	Time interval t (mins)	Interval length (mins)	No. of individuals terminating activity participation in this time period (F_k)	No. of individuals "at risk" of terminating activity participation in this time period (R_k)	Discrete-period hazard $\left(H_k = \frac{F_k}{R_k}\right)$	Standard error of H_k[b]
Walking activity duration						
1	$0 < t \leq 10$	10	7	848	0.008	0.003
2	$10 < t \leq 15$	5	8	841	0.010	0.003
3	$15 < t \leq 20$	5	22	833	0.026	0.006
4	$20 < t \leq 30$	10	37	811	0.046	0.007
5	$30 < t \leq 40$	10	16	774	0.021	0.005
6	$40 < t \leq 50$	10	25	758	0.033	0.006
7	$50 < t \leq 60$	10	106	733	0.145	0.013
8	$60 < t \leq 80$	20	11	627	0.018	0.005
9	$80 < t \leq 100$	20	85	616	0.138	0.014
10	$100 < t \leq 120$	20	116	531	0.218	0.018
11	$120 < t \leq 150$	30	36	415	0.087	0.014
12	$150 < t \leq 180$	30	90	379	0.237	0.022
13	$180 < t \leq 210$	30	27	289	0.093	0.017
14	$210 < t \leq 240$	30	47	262	0.179	0.024
15	$240 < t \leq 300$	60	55	215	0.256	0.030
16	$300 < t \leq 360$	60	47	160	0.294	0.036
17	$360 < t \leq 420$	60	31	113	0.274	0.042
18	$420 < t \leq 480$	60	14	82	0.171	0.042
19	$480 < t \leq 600$	120	34	68	0.500	0.061
20	$600 < t \leq 720$	120	9	34	0.265	0.076
21	$720 < t \leq 840$	120	10	25	0.400	0.098
22	$840 < t$	> 120	15	15	1.000	-
Bicycling activity duration						
1	$0 < t \leq 15$	15	8	167	0.048	0.017
2	$15 < t \leq 30$	15	22	159	0.138	0.027
3	$30 < t \leq 45$	15	17	137	0.124	0.028
4	$45 < t \leq 60$	15	25	120	0.208	0.037
5	$60 < t \leq 90$	30	20	95	0.211	0.042
6	$90 < t \leq 120$	30	20	75	0.267	0.051
7	$120 < t \leq 180$	60	26	55	0.473	0.067
8	$180 < t \leq 240$	60	10	29	0.345	0.088
9	$240 < t$	> 60	19	19	1.000	-

[a] Note, in the estimated model k starts from 0 which represents non-participation in the activity.
[b] Standard error of H_k is estimated using Greenwood's formula.

TABLE 2 Model Estimation Results

	Walking activity		Bicycling activity	
	Estimates[a]	t-stat	Estimates[a]	t-stat
Individual socio-demographic variables				
Male (base: female)			-2.262	-3.39
Employment status (base: not employed)				
Full-time employed			-1.540	-1.79
Part-time employed	-0.588	-1.36	-1.587	-2.43
Household (HH) socio-demographic variables				
Couple (base: "other" households)	-0.702	-1.74	-0.670	-1.84
Presence of children aged between 11 to 15 years in the HH (base: no children in the HH)			-0.724	-1.46
Effects of individual attitudinal variables				
Walking				
Absence of "attractions" and busy life style related factors	1.372	1.99		
Unavailability of walk-friendly environment/facilities	3.033	3.35		
Bicycling				
Busy life style and absence of bicycle paths/trails			2.660	2.52
Inconvenience and lack of paved bicycle facilities			2.179	2.72
(Lack of) Safety			2.059	1.57
Heterogeneity parameters (standard deviation)				
Individual-specific heterogeneity				
Overall (θ)	0.873	3.80	0.873	3.80
Activity-specific (σ_m)	1.007	1.35	0.182	5.15
Household-specific heterogeneity				
Overall (μ)	0.143	1.64	0.143	1.64
Activity-specific (μ_m)			0.342	3.58
Social group-specific heterogeneity				
Overall (η)	0.442	3.80	0.442	3.80
Activity-specific (η_m)	0.257	1.31	0.610	5.01
Spatial cluster-specific heterogeneity				
Overall (δ)	0.690	3.79	0.690	3.79
Activity-specific (δ_m)	0.791	1.32	0.151	5.14
Mean log-likelihood	-5524133.6			

[a] A positive (negative) coefficient implies that the corresponding explanatory variable increases (decreases) the hazard rate and decreases (increases) the activity duration.

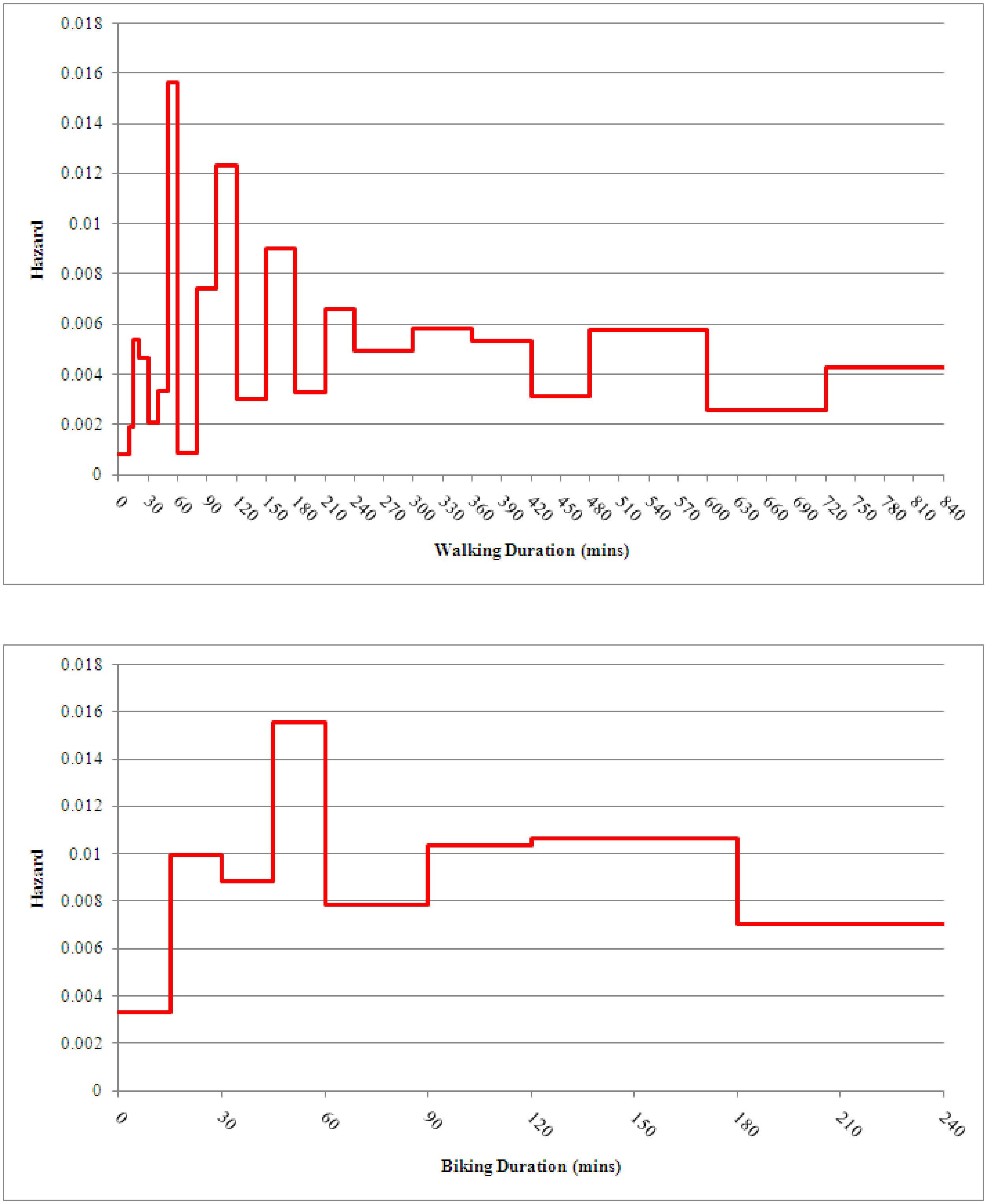

FIGURE 1 Continuous time sample hazard functions for walking and bicycling durations.