Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 3245864, 11 pages
https://doi.org/10.1155/2018/3245864

Research Article

Generalized Estimating Equation Model Based Recursive Partitioning: Application to Distracted Driving

Hyoshin Park and Nigel Pugh

Department of Computational Science & Engineering, North Carolina A&T State University, 1601 E Market Street, Greensboro, NC 27411, USA

Correspondence should be addressed to Hyoshin Park; hpark1@ncat.edu

Received 31 December 2017; Revised 4 April 2018; Accepted 29 April 2018; Published 12 June 2018

Academic Editor: Emanuele Crisostomi

Traditional statistical methods have used a coarse aggregation of data across subjects that may not be representative of any single individual. Even though Generalized Estimating Equations procedure extends generalized linear model to allow for analysis of repeated measurements or other correlated observations, the nonlinear relationship between independent variables and dependent variable could significantly hinder the model's performance. In this study, we propose Generalized Estimating Equation based tree model that combines the advantages of both models by separating the data set recursively into subsets with significantly different parameter estimates. For the best application of the proposed model, distracted driving on intersection is analyzed in this study. Previous studies have focused on evaluating the singular effect of individual geometry and human characteristic variables on driving behaviors. Interactions between variables associated with red-light running (e.g., cell phone usage, cell phone interface, and driver age groups) present different levels of distraction on red-light running. As an indicator of the sensitivity to distractions (referring to the distance being an impairment due to a secondary task), the drivers' distance to the intersection onset of yellow interval is partitioned into two groups (i.e., close and far distance) that maximally differentiate the distraction behavior. Proposed cell phone impact zones produce more significant impacts of distraction on red-light running, compared against dilemma zone. We identify interactions that are sensitive to red-light running and different as a function of the level of the speed and yellow interval duration. Drivers are more vulnerable to cell phone distractions when their location is near the stop line of the intersection at onset of yellow interval.

1. Introduction

Generalized Estimating Equations (GEE) procedure extends the generalized linear model to allow for analysis of repeated measurements or other correlated observations. The major benefit of the repeated measure within subject and independence between subjects have attracted researchers to study crash estimation problems [1]. However, the nonlinear relationship between an independent variable and dependent variable could significantly hinder the model's performance. This is due to the nonlinear differences in structural parameters among the observed variables. There is a need for a new multivariate statistical method focusing on individuals or subgroups.

One of the major causes of nonlinearity is the coarse aggregation of data across subjects that may not be representative of any single individual. We propose GEE tree model that combines the advantages of GEE and the decision tree by separating the data set recursively into subsets with significantly different parameter estimates in a GEE. Decision tree [2, 3] is used to find the partitions of the covariate space associated with differences in the outcome variable. For the best application of GEE tree model, distracted driving on intersection is analyzed in this study.

1.1. Intersection Safety. Approximately 10% of intersections in USA were signalized, whereas fatalities at signalized intersections constituted 32% of the total number of fatalities in 2007 [4]. This disproportionately high percentage of signalized-intersection fatalities has been increasing from 32% in 2007 to 38% in 2014 [5, 6]. Drivers are often exposed to serious traffic crashes at signalized intersections where little space is available to avoid the consequences of unsafe stopping or illegally continuing into the intersection. This dangerous

maneuver is remarkable in areas where drivers likely have difficulties deciding whether to stop (a risk of rear-end crashes) or proceed (a risk of right-angle crashes) during the yellow phase. Studies have concentrated on developing prediction models utilizing observational and simulator data to recognize red-light running infringements before they happen, so the threatened driver and/or the infrastructure can be alerted [7]. An overwhelming 96% of drivers surveyed reported that they fear being hit by a red-light runner [8].

Studies have focused on developing prediction models using observational and simulator data to identify red-light running violations before they occur, so the endangered driver and/or the infrastructure can be notified. The type of dilemma zone has been distinguished, depending on when drivers enter the signalized intersection: a type I dilemma zone, a range where a vehicle approaching the intersection during the yellow interval can neither safely clear the intersection nor stop comfortably at the stop line [9, 10], and a type II dilemma zone, known as an indecision zone, where drivers have the choice either to decelerate and stop or to continue. Various drivers′ stopping probability and zone boundaries have been found [11, 12]. We refer to [13] for a comprehensive review of definitions and evaluation of dilemma zone models and algorithms in the literature. In this paper, the objective is to minimize sampling errors generated from installation. We propose new split indicators that show more superior clustering than the dilemma zone types in describing cell phone distraction impact on red-light running. Mobile phone use while driving involves many different tasks, for example, texting, browsing, taking pictures, and using GPS. This paper is just limited to mobile phone conversations, which are less demanding tasks [14, 15].

1.2. Driving Distraction. Previous studies have shown that drivers' responses to yellow light and distractions are different across gender and age. [16]. Distracted driving is any activity that could divert a driver's attention away from the primary task. For instance, anxious drivers experience performance decrements occurring in central and peripheral tasks [17]. In addition to this cognitive distraction (drivers' minds off the road), using a cell phone can distract drivers to take their eyes off the road (visual) as well as their hands off the steering wheel (physical) [18]. Most of all, the risk of a crash or near-crash was higher (17%) when the driver interacted with a cell phone [19]. The number of mobile-connected devices exceeded the world′s population in 2014 [20]. Findings from generalized linear mixed models for speed have revealed that age, gender, and distractions caused due to mobile phone conversations and texting are the critical factors that influence the mean speed of the drivers [21]. To address drivers from cell phone use, a number of states and jurisdictions in USA have distracted driving laws and designed national advertising campaigns (e.g., U Drive. U Text. U Pay). Nevertheless, about 77% of drivers still answer incoming calls while driving [22], as a consequence of a low risk awareness of phone use. Several studies have been conducted on the effect of age on dual-task performance. The subgroup of teenage drivers are often engaged in distracting activities, potentially putting themselves in danger [23].

The impact of different cell phone types on driving performance remains uncertain. The impact of distraction differs with the type of cell phone used: hand-held (HH), headset (HS), and hands-free (HF) [24]. Drivers were slower during periods without cell phone usage and had an increased time headway to the lead vehicle while talking on HF cell phones. On the contrary, HH and HF cell phone conversations have similar influence on response time and braking behavior [25]. It is important to note how the yellow laws operate in the state of Iowa. The yellow laws are generally restrictive. A yellow light means vehicle traffic that was related to the green light movement is now terminated. The driver should not enter the intersection and stop. If the driver cannot stop safely, then the driver is to cautiously proceed through the intersection [26].

1.3. Cell Phone Distraction at Intersection. Coupling signalized intersections with distraction have the potential to create a major safety concern that hampers critical decisions. For example, dialing outgoing calls or receiving incoming calls from cell phones may impair driving while drivers are at an indecision zone, deciding whether to proceed through the intersection or stop. Distracted young drivers may have gap acceptance behavior at roundabouts with elevated crash risk [27].

However, the effect of cell phone distraction on safety measures may not be the same across all levels of driver age or gender. This effort of identifying interactions is absent in previous studies. In this study, we aim to fill the gap by analyzing the relationship between cell phone distractions, age, gender, speed, yellow interval duration, and red-light running behaviors. To construct the process of drivers′ decision-making, GEE is used based on the data collected from National Advanced Driving Simulator (NADS).

The drivers′ distance to the intersection onset of yellow interval may present considerably different parameters of the estimated model. An average singular value for all scenarios may underestimate or overestimate the true parameters of the model. Important features may be lost in the interpolation process and lead to spurious patterns. The drivers′ distance is partitioned into two groups (i.e., close and far distance from an intersection stop line at the onset of yellow interval) that maximally differentiate the distraction behavior. The findings enhance the understanding of how distractions naturalistically affect driver behaviors in a particular zone (i.e., the two zones in this study). The result is compared with distraction impact of dilemma zone [28]. This knowledge is crucial for the development of engineering countermeasures such as new traffic control with signal timing, traffic sign, and safety notification to reduce crashes and training interventions targeted at improving safety-specific skill behaviors.

In summary, literature clearly agrees with the substantial influence of cell phone distraction on driving. However, recent studies have noted the variable impact of differing driver conditions on behavior. Even though geometric variables have an impact on stopping situations, deceleration rate may be based on selection of comfort of individuals. For example, drivers may brake harder if necessary regardless of speed at the onset of yellow light. We introduce a distance

indicator that is more sensitive to the distraction behavior of particular driver groups. Most of the studies to date use regression analysis that only evaluates the effect of individual independent variables on the dependent variable. To fill this gap, our contribution is demonstrated as follows: (1) identify interactions between six variables associated with red-light running; (2) consider heterogeneity by splitting data into groups based on different distances to stop line; (3) evaluate the moderation impact of cell phone distraction on different age and gender groups with the measurement error in the model; and (4) compare the performance results.

2. Methodology

We analyze driver behavior using definitions [29] by classifying position based on distance measures. A *critical crossing distance* (*CCD*) is associated with the duration of yellow phase (τ), approaching speed of the vehicles (V_0), maximum acceleration rate of vehicles (a_1), intersection width (ω), average vehicle length (l), and driver reaction time (δ_1). It is imperative to note that approaching speed of the vehicles is used to compute the CCD distance. Due to speed already being accounted for, we can work with distance measures. If stopping is chosen instead, a minimum *safe stopping distance* (*SSD*) is necessary to secure a safe stop. The *SSD* depends on the decision-making time of the driver (δ_2) and maximum deceleration rate of approaching vehicles (a_2). The formulation is as follows:

$$CCD = V_0 - (w + l) + \frac{1}{2}a_1 (\tau - \delta_1)^2$$
$$SSD = V_0\delta_2 + \frac{V_0^2}{2a_2} \tag{1}$$

We use the classified rules (Table 1) discussed in [30]. Drivers′ process is classified according to their position on the road and their driving behavior at the particular intersection when facing the yellow light.

2.1. GEE with Multiple Predictors. The main objective of GEE model is to predict a red-light running behavior, as a dependent variable, $\mathbf{Y}_i(Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$, and covariate $\mathbf{X}_i(X_{i1}, X_{i2}, \ldots, X_{in_i})$ for n_i observations in each ith subject and j^{th} response, using estimator $\boldsymbol{\beta}$. Let $R(\alpha)$ be a symmetric matrix that fulfills the requirement of being a correlation matrix and let μ_i be mean $\boldsymbol{\mu}_i(\mu_{i1}, \mu_{i2}, \ldots, \mu_{in_i})$. Let ϕ be a scale parameter to be estimated and let γ be a known variance function of μ_{ij} to calculate $\mathrm{Var}(X_{ij} \mid Y_{ij}) = \gamma(\mu_{ij})\phi$.

Considering independence working assumption [31], the variance-covariance matrix for $\mathbf{Y_i}$ is noted as

$$V_i = A_i^{1/2} R(\alpha) A_i^{1/2} \tag{2}$$

GEE yields asymptotically consistent $\widehat{\beta}$ with misspecified $R(\alpha)$, and the estimate of β is obtained by solving the following estimating equation:

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{K} D_i' V_i^{-1} (Y_i - \boldsymbol{\mu}_i) = 0, \tag{3}$$

Table 1: Drivers′ process classification rules in case of yellow signal.

Distance to STOP line	Reaction	Classification
1. Less than CCD	Crossed	Normal
2. Greater than CCD but less than SSD	Crossed	Aggressive
3. Greater than SSD	Crossed	Aggressive
4. Less than CCD	Stopped	Conservative
5. Greater than CCD but less than SSD	Stopped	Normal
6. Greater than SSD	Stopped	Normal
7. Less than SSD	Crossed	Normal
8. Greater than SSD but less than CCD	Crossed	Aggressive
9. Greater than CCD	Crossed	Aggressive
10. Less than SSD	Stopped	Conservative
11. Greater than SSD but less than CCD	Stopped	Normal
12. Greater than CCD	Stopped	Normal

where $D_i = \partial_{\boldsymbol{\mu}_i}/\partial_{\boldsymbol{\beta}'}$. Under mild regularity condition [31], $\boldsymbol{\beta}$ is asymptotically normally distributed with solution $\widehat{\boldsymbol{\beta}}_G$, and $K^{1/2}(\widehat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})$ is multivariate Gaussian with zero mean and covariance matrix V_G as follows:

$$V_G = \lim_{K\to\infty} K \left\{\sum_{i=1}^{K} D_i' V^{-1} D_i\right\}^{-1} \widehat{W} \left\{\sum_{i=1}^{K} D_i' V^{-1} D_i\right\}^{-1} \tag{4}$$

with

$$\widehat{W} = \sum_{i=1}^{K} D_i' V^{-1} \mathrm{Cov}(\mathbf{Y}_i) V^{-1} D_i \tag{5}$$

The variance estimate $\widehat{V}_G$ can be obtained by replacing α, β, and α with their consistent estimates.

Various correlation structures (independent, exchangeable, autoregressive AR, unstructured, and m-dependent) [31] are compared to enhance the efficiency of the parameter estimates. For the model selection, quasi-likelihood under the independence model criterion (QIC) is used in this study.

We analyzed distraction behavior in two separate models. In the first model for the '*Stop*' vehicle group, the dependent variable is the distance between the stop line (i.e., single white line painted across the road at an intersection) and actual stopped locations of vehicles on the red light. Note that stopping beyond a stop line could be contrary to state laws in USA. Vehicles that successfully stopped before the stop line are coded as '0', while unsuccessful stopping behavior is scaled to the distance they *illegally* stopped after the line. This magnitude is represented by braking distances calculated when δ is 0. Note that, at real intersections, some drivers may habitually stop after the stop line occupying crosswalk space. In the simulation environment, we do not observe this habitual behavior; therefore stopping after the stop line will be regarded as illegal.

The second model indicates the '*Go*' group vehicles, when drivers were trying to pass the intersection before the yellow light turns red (coded as '0'). An interval is calculated as the time difference between the yellow light turning red and aberrant vehicles illegally passing the intersection on the red light.

In this study, we find the effect of explanatory variables on safety behavior (i.e., red-light running) that has been moderated by cell phone distraction. In addition, we have two covariates (speed and yellow interval duration) that are used for controlling purposes. To decompose significant interactions from the full model, we utilize a backwards elimination procedure. We begin with the full factorial design in the first step. If the three-way interaction is significant, we test the simple interaction effects in the second step (no four-way interaction impact was observed in this study). Then, we check the main effects in the final step. On the contrary, if three-way interaction is not significant, it is dropped from the final model. If any simple interaction is significant, we include it; otherwise, we only have the main effect in the final model. This procedure of decomposing interaction effects is tested for 'Stop' and 'Go' vehicle groups separately.

2.2. Decomposition of Distraction Impact. Previous studies divided the data set into different dilemma zone types that can be used to explain drivers' aggressive or conservative behaviors. However, dilemma zone is not optimal criteria to represent the impact of cell phone distractions on red-light running behaviors. Drivers tend to adjust to their geometric circumstances, and this adjustment would let them successfully maneuver as they originally planned. For example, drivers brake hard when they are traveling at a high speed or they approach the intersection. Alerted drivers can decelerate or accelerate at a greater rate, indicating that releasing the drivers' attention could lead to distracted driving [32]. Therefore, the data partitioned by dilemma zone type may not produce statistically significant distraction results. This might be the main reason why [33] did not find any meaningful impact of driver characteristics by considering dilemma zone types.

In this paper, we assume that distractions are more sensitive to the level of drivers, gender, or age difference, when they are approaching close to the intersection onset of yellow interval. Considering the fact that it takes a few seconds to pass the intersection after they see a light turning to yellow, a glimpse of distraction impact from a cell phone might cause drivers to neither successfully stop nor go. We investigate the interaction of distraction impact on optimal distances from the stop line at intersections that would produce meaningful interpretation.

The decision tree [2, 3] is modified following procedure that identifies interactions for each cluster. The main purpose is to split the original GEE model into multigroup models according to covariates that present different cell phone distractions of drivers. Starting from the closet distance to the stop line, we optimally split the whole data set into two groups. Each group is assigned a priority defined to be the proportion of examples misclassified by the node. The tree maintains a queue of leaf nodes ordered by priority, and it successively expands the node at the head of the queue into a fork with two children. We identify nodes that offer the greatest chance of increasing the information gain $G(\sigma)$ based on the distances far from the stop line and close to the stop line.

$$G(\sigma) = R(\sigma)(1 - F(\sigma)) \tag{6}$$

where $R(\sigma)$ is the number of original samples reaching the node divided by total number of original training samples and $F(\sigma)$ is the number of correctly classified samples in the node divided by total number of all the samples in the node [2, 3].

3. Empirical Analysis

3.1. Data Description. The data set comes from a study previously conducted to examine the effects of mobile telephone use on driving performance of drivers in three age groups: young (ages 18-25 years), middle (ages 30-45 years), and older (ages 50-60 years). The study was conducted at the University of Iowa NADS, where participants were asked to drive through a signalized intersection. The traffic signal would transition from green to yellow to red to green again. Participants were engaged in one of three secondary task conditions: baseline, outgoing call, and incoming call. The data was collected at 240 frames per second. Note that we had evenly distributed samples. Identified correlations between two independent variables do not necessary imply that one will cause the other.

Among 49 participants, we obtained 1,158 observations. All participants had three main drives, after they familiarized themselves with the driving simulator. The yellow light event was designed to present the driver with a choice of whether to stop or go at a signalized intersection. As the driver approached the intersection, the traffic signal changed to yellow. The traffic signal was always green prior to the arrival of the driver. As the driver reached a specific point before the intersection, the traffic signal was triggered to change to yellow. The light remained yellow for 4.00 seconds and then transitioned to red for another 5.00 seconds before cycling back to green. Each participant had three drives consisting of three equivalent segments that exposed the participant to three cell phone interfaces. For each visit, the participant experienced a different order of treatment. The incoming and outgoing calls were started prior to the arrival at each segment.

3.2. Preliminary Analysis. The sensitivity of the combined factors against the classification of driver aggressiveness level is shown in Figure 1. We facilitate the understanding of driver aggressiveness to explain conservative driving behavior. When a younger female driver is exposed to a mobile phone with hands-free or headset, she is more likely to stop, even though she has the opportunity to safely cross the intersection. If she is within the *CCD* and decides to brake by applying a high deceleration rate, rear-end accidents may occur. Moreover, we explain aggressive driving behavior. For example, when an older male driver is exposed to an outgoing or incoming call, he chooses to speed up and cross the intersection farther than the *CCD* or the *SSD*, depending on whether there is an option zone or not.

In previous studies, the probability of a vehicle running a red light was analyzed. However, these red-light runners broke the rule, even though they had enough distance to the stop line. The important behavior is the estimation of the magnitude of red-light running from the data. In this study,

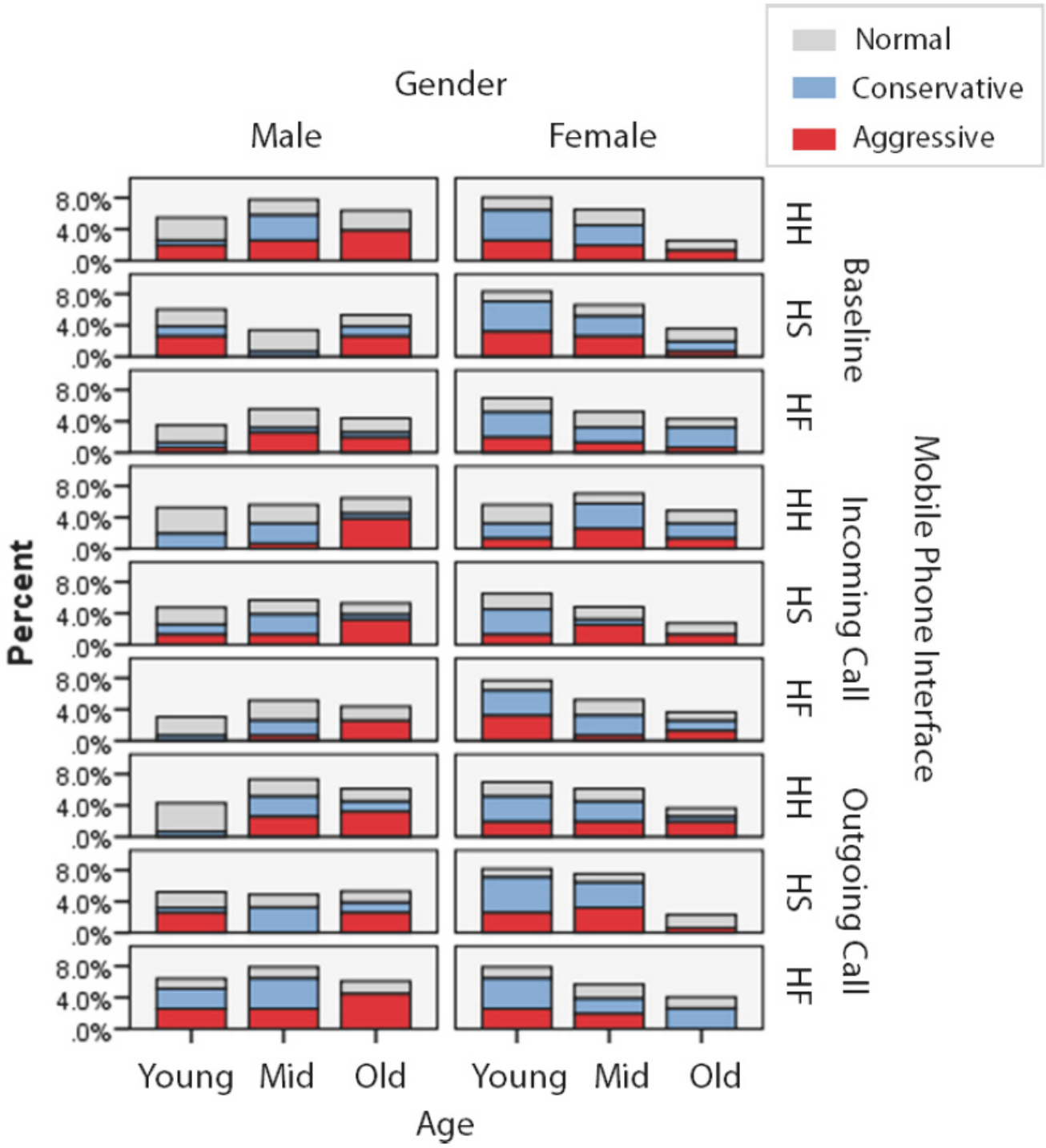


Figure 1: Sensitivity of driver classification to aggressiveness classes.

authors use (1) to calculate the timing difference between when a vehicle is at the stop line and when the signal turns from yellow to red, based on the velocity at the stop line. We assume that cell phone distraction would interact with other driver characteristics and lead to more/less impact on red-light running.

Successfully stopping at the intersection is an important behavior. Figure 1 shows the percentage of observed stop, fail to stop, and go, with different variables. More people successfully stop with longer yellow intervals: 4 sec (60%); lower speed at the yellow onset: 40 mph (60%); longer distance from stop line at the yellow onset: 250 ft (70%);

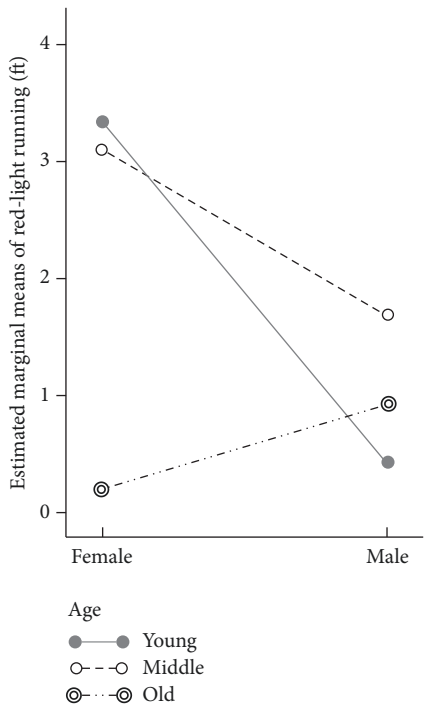


Figure 2: Two-way interaction between gender and age influencing red-light running.

female (57%); young and middle age (50%); and no phone call. However, the number of drivers who intend but fail to stop increases with shorter yellow intervals, higher speed, longer distance, female, old age, outgoing phone call, and hands-held interface.

4. Analysis of Distraction Interaction

4.1. 'Stop' and Red-Light Running. When there is a statistically significant interaction, the results of main effects are not really the true indicators of trends. Table 2 presents the test of the difference between the means and has significant prediction result: $p \leq 0.001$, when a vehicle stops at the intersection. In general, a difference in level between gender groups ($p = 0.002$), female (2.4 ft) and male (1.0 ft), and a difference in level between age groups ($p = 0.001$), young (1.8 ft), middle (2.2 ft), and old (0.5 ft), indicate the main effect on the distance from the stop line. This indicates impact of each predictor in estimating marginal means of the distance from the stop line. When we follow the procedure of previous studies, the main effects of phone use and phone interface are not significant (value in the right column). However, we do not conclude the result without considering the contribution of interaction on the model. We can have a more improved model when we consider the contribution of interaction impact on red-light running behaviors ($p \leq 0.001$) than we would if we considered the model with main effects only ($p = 0.031$). The considerable impact of interaction proves that reporting only main effects can be misleading.

The correlation structures are compared to find the best model in Table 2. The 1-dependent structure obtained a smaller QIC (15.4) than the unstructured (18.6), independent (15.8), exchangeable (15.8), and autoregressive (AR) (15.5). Pairs of measurements with separation greater than K are assumed to be uncorrelated. In this study, $k = 1$-dependent structures provide better (smaller) QICs for GEE models. The remainder of the paper only presents k-dependent structure.

We interpret two significant interactions between driver-character predictors in Figure 2. The first interaction (two-way) occurs between gender and age groups ($p < 0.001$). To be more specific, young and middle-age females stopped farther away from the stop line as compared to males in the same age groups, who stopped closer to the stop line. On the contrary, females in the old age group stopped closer to the stop line, while in the old age group males stopped farther away. The older age group presents a smaller discrepancy between females and males as compared to those in the young and middle-age groups. This finding of a significant difference in gender depending on different age groups indicates that there is indeed significant interaction. Different patterns of slope make each age group intersect with the others and present a crossover interaction.

A second interaction (three-way) occurs (Figure 3) between cell phone usage, interface, and age groups ($p = 0.038$). To portray this three-way interaction, we plot two-way interactions separately. When a driver has an HF phone (a), the shapes of lines for young, middle, and old age groups have similarities that do not indicate any significant difference between no, incoming, and outgoing cell phone use groups. The similarity between age groups does not present interaction with cell phone use groups. However, there is a three-way interaction, because two-way interaction plots (age∗cell phone use) corresponding to different levels of the cell phone interface present significant differences. Age group lines are not parallel; instead, they cross each other one time when a driver has an HS phone (b), and they cross two times when a driver has an HH phone (c). We find that this difference in two-way interactions indicates that there is a statistically significant three-way interaction. These values are based on covariates evaluated at value of speed (62.20 fts) and yellow interval (4.04 s).

4.2. 'Go' and Red-light Running. Table 3 presents the results of the analysis ($p = 0.001$) when a vehicle goes through the intersection without a stop. The interpretation of interaction effects in the analysis variance is based on the given time for a driver before red-light running. Following the procedure in Section 3.1, a two-way interaction occurs without a three-way interaction.

In Figure 4, the effect of phone interface is qualified depending on which age group is considered ($p = 0.029$). One simple effect of the old age group with outgoing call (leading to higher distance than no and incoming call) is different from the effect of the young age group with outgoing call (leading to higher distance than no call and lower distance than incoming call). For outgoing calls, the old age group leads to higher distance than the young age group (3 ft difference), but the change is greatly attenuated for no and incoming calls (almost the same). The effects of gender on distance of red-light running vehicles with use of HH, HF,

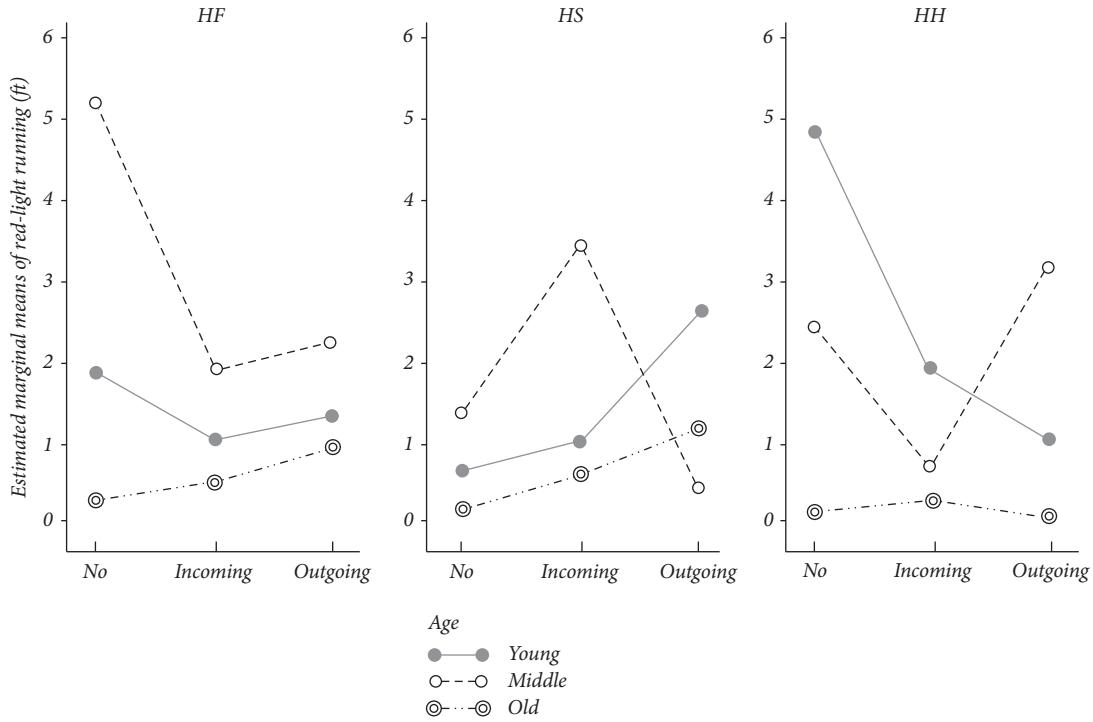


Figure 3: Three-way interaction between cell phone usage, interface, and age influencing red-light running.

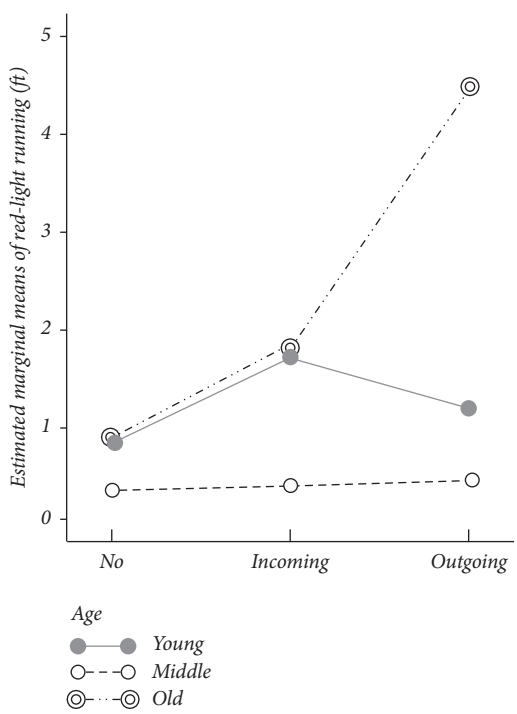


Figure 4: Two-way interaction between cell phone interface and age.

and HS cell phones are not the same. Regardless of an age difference, the phone call condition exhibits safer behavior when in close proximity. The middle-age group scores almost the same for no, incoming, and outgoing calls. Covariates appearing in the model are evaluated at value of speed (62.66 mph) and yellow interval (4.03 s).

5. Application of GEE Tree

5.1. Analysis of Different Distances to Stop Line. Each distance is a candidate separator to present distraction behaviors of different groups. We propose a set of two groups which has a powerful presentation of human interaction and red-light running behaviors compared to various zone types.

Figure 5 shows that data partitioning using dilemma zone type does not produce any significant main effect nor interaction impact of cell phone distraction to the model. The model for the group of the type I dilemma zone does not produce statistically significant results of red-light running. The model for the group of the type II dilemma zone has age, gender, and age∗gender as influential sources without any distraction indicator.

The proposed two groups (i.e., long and short distance) are separated by distances (between 199 ft and 202 ft, in green-shaded region) that share 11 samples. When the distance is closer (from 199-202 ft to 109 ft, in blue-shaded region), there are two-way interactions (phone interface ∗ gender and age ∗ gender) that differ as a function of the level of the speed (58.69 fts) and yellow interval (4.00 s).

However, when the distance is farther away (from 199–202 ft to 284 ft, in yellow-shaded region), no significant interaction effect occurs. Instead, each variable, phone interface, gender, and speed has significant main effects on the

Table 2: Result of model and parameter estimation ('Stop' sample), QIC = 15.4.

Source	Three, two-way interaction, main effects		Main effects
	df	*Sig.*	*Sig.*
Model	55	.000	.031
Intercept	1	.311	.901
Yellow	1	.273	.238
Speed	1	.298	.019
Phone use	2	.388	.250
Phone interface	2	.736	.614
Age*	2	.001	.016
Gender*	1	.002	.502
Phone use ∗ interface	4	.276	-
Phone use ∗ age	4	.653	-
Phone use ∗ gender	2	.394	-
Phone interface ∗ age	4	.257	-
Phone interface ∗ gender	2	.484	-
Age ∗ gender*	2	.001	-
Phone use ∗ interface ∗ age*	8	.038	-
Phone use ∗ interface ∗ gender	4	.390	-
Phone use ∗ age ∗ gender	4	.499	-
Phone interface ∗ age ∗ gender	4	.862	-

*Best results obtained.

Table 3: Result of model and parameter estimation ('Go' sample), QIC = 23.6.

Source	df	Sig.
Model	27	.000
Intercept	1	.000
Phone use	2	.730
Phone interface*	2	.044
Age*	2	.001
Gender*	1	.001
Yellow*	1	.000
Speed	1	.380
Phone use ∗ age	4	.664
Phone interface ∗ age*	4	.029
Age ∗ gender	2	.793
Phone use ∗ interface	4	.956
Phone use ∗ gender	2	.506
Phone interface ∗ gender	2	.247

*Best results obtained.

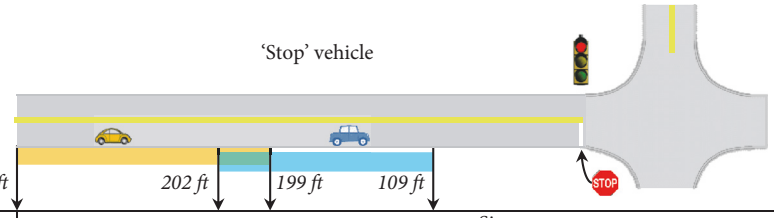
model. In both groups (long and short), the impact of phone interface on red-light running is universal, no matter how close the drivers are to the stop line.

Figure 6 shows that dilemma zone separator has no meaningful impact associated with cell phone distraction on red-light running. The type I dilemma zone model is not statistically significant. The type II dilemma zone model has only age and speed as main effects without any interaction.

Compared to the type II dilemma zone model ($p = 0.031$, QIC = 34.9), the proposed models present better performance (short: $p = 0.008$, QIC = 13.9; long: $p \leq 0.001$, QIC = 15.1). The separator for the 'Go' model has 15 samples of distances (i.e., between 214 ft and 217 ft) that can be used for both groups. For the short distance group, a significant interaction between phone interface and age was observed ($p = 0.014$). This group presents enhanced interaction effect compared to the model that considered the whole 'Go' samples in Table 3 ($p = 0.014$). By separating the long distance group (i.e., no cell phone interaction impact) from the whole 'Go' model, we can have better interpretation of red-light running behavior associated with distraction.

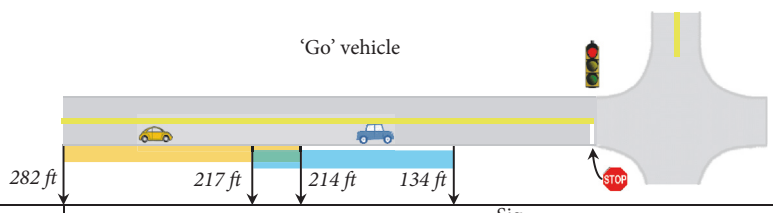
6. Discussion

Results of the proposed model show that there is a cell phone impact zone where drivers are more sensitive to secondary tasks, compared to the transnational definition of a dilemma zone [13]. Training interventions may be designed and evaluated, targeted at teaching drivers safe practices of navigating intersections and quicker decision making in dilemma zones. Note that the combinations in groups using 15 samples result in the same statistical results in the significance of the model: from the more conservative case (long distance: 217-282 ft and short distance: 134-217 ft) to the less conservative case (long distance: 214–282 ft and short distance: 134–214 ft). To represent a reasonable engineering practice, we always pick the best number (e.g., 215 ft) to report the best estimation of parameter, while also keeping the range that presents the statistically significant result. The safety implications found in this exploration are particularly imperative. One example is the detection of distracted driving [14]. In our paper, the findings of which interaction variables have a statistically



Source	Sig.			
	Long distance	Short distance	Type I dilemma zone	Type II dilemma zone
Model	.006	.000	Statistically not significant	.001
Intercept	.260	.019		.252
Phone use	.625	.365		.511
Phone interface	.008	.046		.766
Age	.094	.006		.002
Gender	.018	.004		.001
Yellow	.459	.031		.237
Speed	.046	.048		.297
Phone use*Age	No interactions	.769		.873
Phone interface*Age		.540		.685
Age*Gender		.005		.003
Phone use*Interface		.189		.289
Phone use*Gender		.248		.467
Phone interface *Gender		.023		.514

Figure 5: Different distances to stop line (Stop).



Source	Sig.			
	Long distance	Short distance	Type I dilemma zone	Type II dilemma zone
Model	.008	.000	Statistically not significant	.031
Intercept	.015	.000		.901
Phone use	.670	.083		.250
Phone interface	.368	.243		.614
Age	.085	.048		.016
Gender	.004	.027		.502
Yellow	.326	.000		.238
Speed	.009	.000		.019
Phone use*Age	No interactions	.216		No interactions
Phone interface*Age		.014		
Age*Gender		.252		
Phone use*Interface		.560		
Phone use*Gender		.607		
Phone interface *Gender		.149		

Figure 6: Different distances to stop line (Go).

significant effect on the driver in the respective dilemma zone type (type I or type II) show when drivers are most vulnerable to mobile phone conversations distractions. This information combined with distinguishing different distances from the intersection where driver distractions are sensitive will assist in avoiding or reducing the amount of collisions. In the future, appropriate engineering countermeasures can be developed and designed to incorporate these findings to further investigate avoiding collisions and educating drivers on improving driving safety behavior practices.

7. Conclusion and Future Research

We present analytical models to interpret red-light running associated with the cell phone distractions. Results of a model considering main effects without interaction effects can be misleading. In general, a GEE that considers interactions in the model provides superior performance of prediction. The first category, 'Stop' vehicles for both long and short distance to the intersection, is influenced by different levels of phone interface (HH, HS, and HF). In particular, vehicles having short distance are more sensitive to phone interface with different levels of genders. On the contrary, dividing the whole 'Stop' samples based on the dilemma zone types does not provide us any significant cell phone distraction impacts on red-light running. The second category, 'Go' vehicles, presents stronger impacts of interaction between phone interface and age for the shorter distance group.

The age group analyzed by this paper does not distinguish those over the age of 65. Future research includes using this older-65 population with cognitive declines and slow decision-making/response times. In this paper, one intersection was used to investigate the cell phone impact zone. However, different geometric conditions in different intersection areas would cause drivers to have different stopping behaviors. Furthermore, several intersections can be grouped with assumption of similar driving behaviors among different areas. The proposed method provides a manual to transportation authorities to locate the "signal ahead" sign at an appropriate distance from the intersection. This study is limited only to the cell phone effects, but future study can include the secondary task [34, 35].

In future work, data can also be cross-validated with real-world data. Further, the impact of cell phones on distraction behavior of drivers in automated and connected vehicles at signalized intersections can be investigated. Differential impact of driving experience can also be considered in subsequent modeling attempts. Advanced computing models also can be used to extract the important variables considering nonlinear relationship [36].

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the official views or policies of NADS or the US Department of Transportation-National Highway Traffic Safety Administration or any of the above organizations, nor do the contents constitute a standard, specification, or regulation of these organizations. An earlier version of this work was presented at the First Symposium of Center for Advanced Transportation Mobility (CATM 2017).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The data used for this study was from a study conducted at the University of Iowa NADS, as part of a wireless urban arterial project funded by the US Department of Transportation-National Highway Traffic Safety Administration (NHTSA). The authors would like to express their gratitude for providing descriptions and clarifications. This paper is partially supported by start-up fund provided by North Carolina A&T State University and USDOT University Transportation Centers (Contract 69A3551747125), Chancellor's Distinguished Fellows, and Department of Education Title III Program.

References

[1] X. Wang, M. Abdel-Aty, and P. Brady, "Crash Estimation at Signalized Intersections: Significant Factors and Temporal Effect," *Transportation Research Record*, vol. 1953, pp. 10–20, 2006.

[2] H. Park and A. Haghani, "Real-time prediction of secondary incident occurrences using vehicle probe data," *Transportation Research Part C: Emerging Technologies*, vol. 70, pp. 69–85, 2016.

[3] H. Park, A. Haghani, and X. Zhang, "Interpretation of Bayesian neural networks for predicting the duration of detected incidents," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 20, no. 4, pp. 385–400, 2016.

[4] Federal Highway Administration (FHWA), "Safety at Signalized Intersections," FHWA/TX-03/4022-2, FHWA, Washington, DC, USA, 2009.

[5] United States Department of Transportation (US DOT), *Traffic Safety Facts: A Compilation of Motor Vehicle Crash Data from The Fatality Analysis Reporting System And The General Estimates System*, National Highway Traffic Safety Administration, Washington, DC, USA, 2007.

[6] United States Department of Transportation (US DOT), *Traffic Safety Facts: A compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system*, National Highway Traffic Safety Administration, Washington, DC, USA, 2014.

[7] A. Jahangiri, H. Rakha, and T. A. Dingus, "Red-light running violation prediction using observational and simulator data," *Accident Analysis & Prevention*, vol. 96, pp. 316–328, 2016.

[8] R. A. Retting and A. F. Williams, "Characteristics of red light violators: Results of a field investigation," *Journal of Safety Research*, vol. 27, no. 1, pp. 9–15, 1996.

[9] D. Gazis, R. Herman, and A. Maradudin, "The problem of the amber signal light in traffic flow," *Operations Research*, vol. 8, pp. 112–132, 1960.

[10] FHWA, "Traffic signal timing manual," FHWA-HOP-08–024, Federal Highway Administration (FHWA), Washington, DC, USA, 2008.

[11] FHWA, "Intelligent detection-control system for rural signalized intersections," FHWA/TX-03/4022-2, Federal Highway Administration, Washington, DC, USA, 2002.

[12] FHWA, "Traffic control systems handbook," FHWA/TX-03/4022-2, Federal Highway Administration (FHWA), Washington, DC, USA, 2005.

[13] D. Hurwitz, "Application of driver behavior and comprehension to dilemma zone definition and evaluation," *Dissertations. Paper*, vol. 112, 2009.

[14] O. Oviedo-Trespalacios, M. M. Haque, M. King, and S. Washington, "Understanding the impacts of mobile phone distraction on driving performance: A systematic review," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 360–380, 2016.

[15] D. G. Kidd, J. Tison, N. K. Chaudhary, A. T. McCartt, and T. D. Casanova-Powell, "The influence of roadway situation, other contextual factors, and driver characteristics on the prevalence of driver secondary behaviors," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 41, pp. 1–9, 2016.

[16] M. M. Haque, A. D. Ohlhauser, S. Washington, and L. N. Boyle, "Decisions and actions of distracted drivers at the onset of yellow lights," *Accident Analysis & Prevention*, vol. 96, pp. 290–299, 2016.

[17] C. M. Janelle, R. N. Singer, and A. M. Williams, "External distraction and attentional narrowing: Visual search evidence," *Journal of Sport & Exercise Psychology*, vol. 21, no. 1, pp. 70–91, 1999.

[18] National Safety Council, *Understanding The Distracted Brain: Why Driving While Using Hands-Free Cell Phones Is Risky Behavior*, National Safety Council Press, Washington, DC, USA, 2012.

[19] Insurance Institute for Highway Safety, *Relationship of Near-Crash/Crash Risk to Time Spent on A Cell Phone While Driving*, Highway Loss Data Institute, Arlington, VA, USA, 2014.

[20] Cisco, "Visual networking index: global mobile data traffic forecast update in 2014-2019," 2015.

[21] N. R. V. Pushpa Choudhary, "Mobile phone use during driving: Effects on speed and effectiveness of driver compensatory behaviour," *Accident Analysis and Prevention*, vol. 106, Article ID S000145751730235, pp. 370–378, 2017, http://www.sciencedirect.com/science/article/pii/S000145751730235.

[22] J. Tison, N. Chaudhary, and L. Cosgrove, "National phone survey on distracted driving attitudes and behaviors," DOT HS 811 555, NHTSA, Washington, DC, USA, 2011.

[23] E. J. Westlake and L. N. Boyle, "Perceptions of driver distraction among teenage drivers," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 15, no. 6, pp. 644–653, 2012.

[24] National Highway Traffic Safety Administration (NHTSA), "Traffic Safety Facts Research Note," DOT HS 812 101, National Center for Statistics and Analysis, Washington, DC, USA, 2014.

[25] M. M. Haque and S. Washington, "The impact of mobile phone distraction on the braking behaviour of young drivers: A hazard-based duration model," *Transportation Research Part C: Emerging Technologies*, vol. 50, pp. 13–27, 2015.

[26] The Iowa Legislative Services Agency, *Section 321.257- Official traffic control signal*, The Iowa Legislature, Des Moines, IA, USA, 2017.

[27] M. M. Haque, O. Oviedo-Trespalacios, A. K. Debnath, and S. Washington, "Gap acceptance behavior of mobile phone-distracted drivers at roundabouts," *Transportation Research Record*, vol. 2602, pp. 43–51, 2016.

[28] D. Moore and D. Hurwitz, "Fuzzy logic for improved dilemma zone identification," *Transportation Research Record*, no. 2384, pp. 25–34, 2013.

[29] L. James, *Traffic engineering handbook*, Institute of Transportation Engineers, Washington, DC, USA, 5th edition, 1999.

[30] P. Papaioannou, "Driver behaviour, dilemma zone and safety effects at urban signalised intersections in Greece," *Accident Analysis & Prevention*, vol. 39, no. 1, pp. 147–158, 2007.

[31] K. Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.

[32] T. J. Gates, D. A. Noyce, L. Laracuente, and E. V. Nordheim, "Analysis of driver behavior in dilemma zones at signalized intersections," *Transportation Research Record*, vol. 2030, pp. 29–39, 2007.

[33] S. M. Lavrenz, V. Dimitra Pyrialakou, and K. Gkritza, "Modeling driver behavior in dilemma zones: A discrete/continuous formulation with selectivity bias corrections," *Analytic Methods in Accident Research*, vol. 3-4, pp. 44–55, 2014.

[34] H. Park, S. Gao, and S. Samuel, "Modeling Effects of Forward Glance Durations on Latent Hazard Detection," *Transportation Research Record*, vol. 2663, pp. 90–98, 2017.

[35] H. Park, "Explicit forward glance duration hidden Markov model for inference of hazard detection," in *Proceedings of 97th Annual Meeting of Human Factors and Ergonomics Society (HFES, 2017)*, Austin, TX, USA, October 2017.

[36] H. Park, A. Haghani, S. Samuel, and M. A. Knodler, "Real-time prediction and avoidance of secondary crashes under unexpected traffic congestion," *Accident Analysis & Prevention*, vol. 112, pp. 39–49, 2018, https://www.sciencedirect.com/science/article/pii/S000145751730421.