2014

STRIDE | Southeastern Transportation Research, Innovation, Development and Education Center

Final Report

Empirically-Based Performance Assessment & Simulation of Pedestrian Behavior at Unsignalized Crossings

(2012-016S)

Bastian Schroeder, Ph.D., North Carolina State University
Nagui Rouphail, Ph.D., North Carolina State University
Lily Elefteriadou, Ph.D., University of Florida
Virginia Sisiopiku, Ph.D., University of Alabama at Birmingham
Kati Salamati, Ph.D., North Carolina State University
Graduate Students: Elizabeth Hunter, Briana Phillips, Thomas Chase, Yinan Zheng, Shrikanth Mamidipalli

September 2014

This page intentionally left blank

STRIDE Project 2012-016S
-
Empirically-Based Performance Assessment and Simulation of Pedestrian Behavior at Unsignalized Crossings

Bastian Schroeder
Nagui Rouphail
Katy Salamati
Elizabeth Hunter
Briana Phillips
INSTITUTE FOR TRANSPORTATION RESEARCH AND EDUCATION
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NC

AND

Lily Elefteriadou
Thomas Chase
Yinan Zheng
UNIVERSITY OF FLORIDA TRANSPORTATION INSTITUTE
UNIVERSITY OF FLORIDA
GAINESVILLE, FL

AND

Virginia P. Sisiopiku
Shrikanth Mamidipalli
UNIVERSITY OF ALABAMA AT BIRMINGHAM
BIRMINGHAM, AL

Southeastern Transportation Research, Innovation, Development, and Education Center
Gainesville, FL
2014
http://www.stride.ce.ufl.edu/

U.S. DOT DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGMENTS

This work was sponsored by a grant from the Southeastern Transportation Research, Innovation, Development, and Education Center (STRIDE) at the University of Florida. The STRIDE center is funded through the U.S. Department of Transportation's University Transportation Centers Program. Additional financial support was provided by the North Carolina, Florida, and Alabama Departments of Transportation. The authors would like to thank STRIDE and NCDOT, FDOT, and ALDOT for their support of university-based research in transportation, and especially for the funding provided in support of this project.

The authors would also like to thank the reviewers of the STRIDE proposals and draft deliverables for their review time and input in this project. The authors would also like to express special gratitude to Dr. Scott Washburn at the University of Florida, who graciously assisted and advised the team in implementing the resulting algorithms in a new microsimulator he is developing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

This page intentionally left blank

ABSTRACT

The objective of this research was to provide an improved understanding of pedestrian-vehicle interaction at mid-block pedestrian crossings and develop methods that can be used in traffic operational analysis and microsimulation packages. Models describing driver yielding and pedestrian gap acceptance behavior were developed from field data collected at 27 mid-block pedestrian crossings in three states (Alabama, Florida, and North Carolina), encompassing two different types of land use: university campuses and downtown areas. The project included an in-vehicle driver behavior study with 15 drivers. This part of the data collection was performed in Florida and the results were used to develop nine simulation components describing various aspects of pedestrian-vehicle interaction. Specific outcomes for this research include: (a) a standalone model of pedestrian gap acceptance behavior at unsignalized crossings, (b) a driver yielding behavioral model, (c) models describing vehicle dynamics and driver behavior in advance of the crosswalk, (d) prototype algorithms incorporated and tested in a micro simulator, and (e) educational modules for dissemination of the research results. Key deliverables include the prototype algorithms implemented in simulation, a final report summarizing the research and findings, and educational modules on the research results that can be incorporated into university curricula, or serve as material for standalone professional development courses.

This page intentionally left blank

EXECUTIVE SUMMARY

The objective of this research was to develop new and improved algorithms for describing pedestrian and vehicle interaction at unsignalized midblock pedestrian crossings and to implement them in a traffic simulation environment. The algorithms developed address pedestrian and driver behavior at mid-block crosswalks, based on targeted empirical observations of naturally occurring and staged crossings. The models describe pedestrian gap selection, driver yielding behavior, and other behavioral processes. All models were developed to be compatible in form with algorithms used in microsimulation tools, and have been implemented in a micro simulator for illustrative purposes and testing.

The primary data collection included an observational study of pedestrian-vehicle interaction performed at 27 unsignalized mid-block crosswalks in Florida, Alabama, and North Carolina. The data collection protocol included video recording of pedestrian-vehicle interaction events, as well as measurements of vehicle speed and gap times in advance of the crosswalk. The response variables in the evaluation were the decision of a driver to yield to a pedestrian, and the decision of a pedestrian to accept a gap in conflicting traffic. Independent variables included the vehicle speed, necessary deceleration rate, gap length, pedestrian behavioral attributes, driver and vehicle characteristics, and site attributes.

The final recommended driver yielding model uses the form of a binary logit model. In the recommended model, increased vehicle speed (SPD) was seen to reduce the likelihood of yielding, as did an increased required deceleration rate (DECEL). Presence of adjacent yields (ADJ), low speed platoons (LSPLT), presence of multiple pedestrians (MUP), and female pedestrians (FEMALE) were seen to increase the likelihood of yielding. Drivers were also more likely to yield to pedestrians on-campus (CAMPUS) than off-campus. For the universal model, all variables except for DECEL were significant at the $p<0.05$ level.

The recommended gap acceptance model uses only two parameters, the size of the gap length in seconds, and a binary variable distinguishing between gaps and lag events (first arriving vehicle, without a prior lead vehicle to "open" the gap). An increase in gap length is associated with an increased probability of pedestrians crossing. A lag event has a negative coefficient, meaning that a pedestrian is less likely to accept a lag than a gap given the same length in seconds. This effect may be explained because the pedestrian may require some of the lag time to evaluate the available time to cross, after first arriving at the crosswalk. For a gap event, this "screening time" and decision making likely takes place before the gap "opens".

In addition, an instrumented vehicle experiment was conducted to provide insight on other simulation components and variables that define the interaction between vehicle(s) and pedestrian(s) at the crosswalk. The models are implemented in the simulation to supplement the yield and gap acceptance models described above. Overall, nine supplemental components were developed, for a total of eleven predictive models:

- Driver Yield Decision (from observational study),
- Pedestrian Gap Acceptance (from observational study),
- Pedestrian-Vehicle Conflict Identification,
- Driver Decision Distance,
- Driver Yield Type Check,
- Soft Yield Dynamics,
- Hard Yield Dynamics,
- Driver Yield Rejection Response,
- Driver Wait Time,
- Pedestrian Yield Recognition, and
- Pedestrian Yield Rejection.

The eleven resulting algorithms were implemented in a microsimulation model to evaluate their effectiveness and accuracy. The documentation of the behavioral models will allow for implementation of the algorithms in other simulation software packages as well. The behavioral models are based on empirical observations, and were derived from field observations of naturally occurring and staged pedestrian crossings.

The project resulted in the following products:

1. A driver yielding behavioral model developed from 27 unsignalized crosswalks in three states, sensitive to vehicle dynamics, pedestrian attributes, traffic condition, and geographical area (North Carolina, Florida, and Alabama);
2. A standalone model of pedestrian gap acceptance behavior at unsignalized crossings, sensitive to the available gap length and distinguishing gap and lag events, and developed from data collected at 24 unsignalized crosswalks in three states;
3. Nine supplemental model components describing various attributes of pedestrian-vehicle interaction, derived from an in-vehicle driver study with 15 participants conducted in Florida;
4. Implementation of all eleven algorithms in a simulation environment, which incorporates pedestrian-vehicle interactions at unsignalized crossings in a micro simulator; and
5. Educational modules for dissemination of the research results to researchers and students in the southeast and nationally, supported by seamless technology transfer through the available simulation modules.

This research has broad impact on the state of the practice of pedestrian analysis in the Southeast region, and likely beyond. The field of pedestrian analysis and modeling has documented gaps and limitations, and this research aimed to make significant improvements to the ability to model pedestrian traffic. In an age of increasing focus on accommodation of non-motorized road users in our transportation systems, engineers need tools to evaluate the impacts of different intersection treatments on both pedestrians and the conflicting vehicle stream. Oftentimes, engineering analyses include the use of microsimulation tools, which to this point had not been specifically calibrated for pedestrian-vehicle interaction behavior. The behavioral models resulting from this research will assist in evolving these microsimulation tools to the point where analysts can predict the operational characteristics of unsignalized pedestrian crossings.

This research delivered an improved understanding of pedestrian and driver behavior at unsignalized midblock crossing points and provides practitioners with enhanced tools for considering pedestrian presence. This goal is being achieved by developing algorithms for microsimulation tools to model the interaction between pedestrians and drivers. The pedestrian-vehicle interaction simulation was implemented in a new microsimulator developed by Dr. Scott Washburn at the University of Florida. This project adds pedestrian movement and interaction with vehicles to the existing model that is programmed in the C# programming language to demonstrate an application of the statistical interaction models in microsimulation. The simulation developed for this project is a proof of concept, as the models of interaction can be applied in any microsimulation that includes vehicles and pedestrians. The simulation is time step based and once the simulation is initialized, all vehicle movements and decisions are made prior to pedestrian movements and decisions, at each time step. Major sub models developed for the simulation are the results of various data collection and modeling efforts from this research and are presented in detail in the previous chapters.

Limitations of this research include those related to the scope and type of collected data, the modeling assumptions, and the implementation in simulation itself. All limitations are described in detail in Chapter 7: Conclusions and Recommendations.

CHAPTER 1: INTRODUCTION

Background

Pedestrian safety and access are key contributors to livability in modern urban and suburban infrastructure systems, which focus increasingly on walkability and multi-modal transportation. An increasing number of local and state agencies are placing high priority on providing adequate pedestrian facilities. Pedestrian facility improvement projects aim to create recreational pedestrian paths, to revitalize downtown areas, or to create safe walking routes to destinations. With increasing focus on these types of pedestrian facilities, questions about pedestrian safety, the interaction of pedestrians and motorized traffic, as well as their operational effects, need to be explored. Especially when these pedestrian facilities intersect with streets, engineers have to decide how to control the interaction of the pedestrian and vehicle modes at the crossing. Oftentimes, traffic signals are not warranted (FHWA, 2009) at crossings without the adequate volumes of pedestrian and vehicle traffic, and even if warranted, other considerations may conflict with the installation of a traffic signal. As a result, many pedestrian crossings in the United States and in other countries are unsignalized. However, a thorough understanding of pedestrian (and driver) behaviors in these environments, and especially at unsignalized street crossings is lacking.

The National Highway Traffic Safety Administration, NHTSA, lists a total of 4,743 pedestrian fatalities in 2012, and 76,000 pedestrian injuries in traffic collisions nationwide. The report further cites that the highest rates of pedestrian fatalities occurred during the hours of 4pm to 8pm (almost 26%) and 8pm to 12am (almost 28%) on weekdays, suggesting a relationship between pedestrian safety and heavy PM peak hour traffic (NHTSA, 2014). Pedestrian fatalities were highest in California (612), Texas (478), and Florida (476). When the states are sorted by "Pedestrian Fatalities per 100,000 Population," four of the top ten states were in the southeast (South Carolina, Louisiana, Florida, and North Carolina). These statistics underline the critical importance of improving pedestrian safety in the southeast region and the state included in the STRIDE consortium.

An earlier study by NHTSA reports that 45% of pedestrian fatalities in 2001 are caused by improper crossing of a roadway or intersection, or failure to yield right-of-way (NHTSA, 2003), which underlines the importance of proper pedestrian facilities for potential pedestrian crossing locations. The same document confirms that 70% of the pedestrian fatalities happen on roadway crossings and intersections with or without crosswalks, and found that 41% of fatalities happened on locations where a crosswalk was not available. No statistical data was found on the portion of pedestrian collisions at signalized and unsignalized intersections.

While recent national-level research has explored impacts of a range of pedestrian crossing treatments, and developed new pedestrian signal warrants (Fitzpatrick, et. al. 2006), the macroscopic approach used in that research is inadequate for simulation-based evaluation of driver and pedestrian interactions. In the current state of engineering practice, microscopic traffic simulation tools are routinely used to explore and quantify the performance of various intersection treatments. In one recent example, simulation was used to quantify the operational impacts of pedestrian signalization at modern roundabouts to vehicular traffic (Schroeder, et al. 2008). Through simulation, the authors were able to draw conclusions about innovative signal strategies and crosswalk configurations that may help alleviate pedestrian-induced vehicular delay, while providing safe crossing opportunities for pedestrians. The recent success of simulation is attributable to its ability to replicate variability (for example, in driver behavior), to consider system effects, and to produce sophisticated 3D and 4D visualization output to use in stakeholder outreach. However, consideration of pedestrian behavior in simulation is currently limited and is recognized as one of top ten key research needs in traffic simulation (FHWA, 2004). Of particular research need is the interaction between pedestrians and vehicles at unsignalized pedestrian crossings, and in particular, simulation algorithms to describe pedestrian gap acceptance and driver yielding behavior.

Previous work by members of the research team (Schroeder, et al. 2011) has defined a framework for describing the accessibility of an unsignalized crossing to pedestrians as a function of four factors: 1. The availability of crossing opportunities in the form of yields and crossable gaps, 2. the rate of utilization of these opportunities by pedestrians, 3. the delay incurred until an opportunity is utilized, and 4. the level of risk experienced when attempting a crossing. It also builds on prior work on driver yielding behavior (Schroeder and Rouphail, 2011a) and pedestrian gap acceptance behavior (Schroeder and Rouphail, 2011b); both focused on pedestrian behavior at midblock crossings. In the context of the aforementioned framework, this project developed behavioral models for the first two factors (i.e., availability and utilization of gaps and yields), incorporated them in a simulation environment, and estimated factor three (delay). While some research has explored the ability of simulation to also predict the fourth factor in the level of risk (e.g. FHWA Surrogate Safety Assessment Methodology, SSAM), the modeling of pedestrian risk is beyond the scope of this effort.

Project Objectives

The objective of this research was to develop new and improved algorithms for describing pedestrian and vehicle interaction at unsignalized crossings and to implement them in a traffic simulation environment. The algorithms developed address pedestrian and driver behavior at mid-block crosswalks, based on targeted empirical observations of naturally occurring and staged crossings (i.e., those where the pedestrian in the interaction is one of the researchers). The models describe pedestrian gap selection and driver yielding behavior, and are compatible in form with algorithms used in microsimulation tools.

The behavioral models were developed based on empirical observations, including both natural and staged pedestrian crossings. Staged crossings were used at sites where there were few naturally occurring pedestrians observed. Study sites were located on university campuses as well as at downtown areas.

This research also developed and implemented algorithms in a traffic micro simulator. Detailed documentation of these models is provided to allow for their implementation in other simulation software packages.

Project Impact and Products

This project addresses two main themes cited in the STRIDE mission statement, namely *safety and livability*. STRIDE's focus on *safety* considerations is addressed in relation to pedestrian crossings at unsignalized intersections, where shared right of way rules are often unclear and sometimes violated. The research also addresses human factors issues in formulating behavioral algorithms that go beyond traditional yield or gap acceptance models. The models are based on such explanatory variables as pedestrian level of assertiveness, presence of platoons behind the subject vehicle, crossing treatments, and driver required deceleration rates among other factors.

On the *livability* side, this research tackles the issue of pedestrian facility evaluation and modeling, including the effect of infrastructure and the built environment on both pedestrian and driver behavior. Pedestrian behavior is modeled as a function of the crossing width, prevailing speed limit, as well as land use factors surrounding the crossing. Driver behavior models are based on naturalistic driving observations from 15 drivers. The vehicle and pedestrian interaction system was modeled in microsimulation and is well documented so that developers of other micro simulators may also incorporate these models in other tools.

This research delivers an improved understanding of pedestrian and driver behavior at unsignalized crossing points and will provide practitioners with enhanced tools for considering pedestrian presence. This research enhances livability by more accurately considering access (and impediments to it) in urban street analyses. Results from this research may be used to support policy-level recommendations on multimodal transportation infrastructure design that considers pedestrian access. In summary, the research produced:

1. A series of pedestrian behavioral models regarding yield and gap acceptance by pedestrian user class, traffic condition, and geographical area (North Carolina, Florida, and Alabama) at midblock crosswalks;
2. A similar set of driver yielding behavioral models across the same factors as in (1);
3. Prototype algorithms which for the first time will incorporate pedestrian-vehicle interactions at unsignalized crossings in a micro simulator;
4. An ability to study the impact of pedestrian and vehicle interactions on system performance in a simulation environment, and obtain delays to vehicles and pedestrians, as well as the capacity of unsignalized crossings; and
5. Presentations, publications, and educational modules for dissemination of the research results to researchers and students in the southeast and nationally, supported by seamless technology transfer through the available simulation modules.

This research has a broad impact on the state of the practice of pedestrian analysis in the Southeast region, and beyond. The field of pedestrian analysis and modeling has documented gaps and limitations, and this research made significant contributions in our ability to model pedestrian-vehicle interactions. In an age of increasing focus on accommodation of non-motorized road users in our transportation systems, engineers need the tools to evaluate the impacts of different intersection treatments on both pedestrians and the conflicting vehicle stream. The behavioral models resulting from this research are expected to evolve microsimulation tools to the point where analysts can predict the operational characteristics of unsignalized pedestrian crossings, which is an area that is not well documented at the present time.

Report Organization

This report is organized in seven chapters and several appendices which provide more detailed documentation. Chapter 1 provided background on the project and a summary of the study objectives. Chapter 2 provides a literature review of prior work related to the subject matter. Chapter 3 presents the analysis framework for this project, followed by a discussion of the data collection methodology in Chapter 4. Chapter 5 presents the modeling results in the form of driver yielding models, pedestrian gap acceptance model, and other driver behavioral models. Chapter 6 presents the implementation of these models into a micro simulator followed by a conclusions and recommendation discussion in Chapter 7.

This page intentionally left blank

CHAPTER 2: LITERATURE REVIEW

Introduction

A thorough literature review was performed in this study to gain an in depth understanding of current practices and existing knowledge gaps with respect to modeling pedestrian and vehicle interaction. The review of literature concentrated on pedestrian stream models, current analysis methods for various facility types, pedestrian behavioral attributes, and approaches for data collection and performance estimation. This chapter provides a brief summary of recent published literature pertaining to pedestrian- vehicle interactions at unsignalized crossings with emphasis on the pedestrian gap acceptance and driver yielding behavior studies. The detailed literature review is available in Appendix A.

Background

Current traffic engineering analysis tools and capacity models are of limited use for evaluating the interaction of pedestrians and vehicles at unsignalized crossing facilities. The analysis methodologies for unsignalized intersections in the 2010 Highway Capacity Manual (HCM) are traditionally limited to boundary cases, which assume strictly enforced right-of-way rules (TRB, 2000). These assumptions mean that pedestrian operations are analyzed by either assuming pedestrian priority (100% driver yielding) or vehicle priority without yielding right-of-way to pedestrians (Schroeder and Rouphail, 2011). More complex interaction of the two modes in which some drivers yield to pedestrians and some pedestrians accept gaps in traffic is typically ignored in traditional HCM methods. This type of interaction was previously referred to as a *mixed-priority* crossing (Schroeder and Rouphail, 2010) and is very common in the field.

Changes in the 2010 HCM (TRB, 2010) have made an attempt at combining pedestrian gap acceptance and driver yielding behavior for pedestrian delay analysis, but the revised methodology is not based on empirical observations and has not been calibrated by field observations. In practice, alternative analysis tools in the form of microscopic simulation applications are frequently used to help overcome some of the limitations of the HCM procedures.

Gap Acceptance

The review of the literature confirmed that pedestrian crossing behavior has not been explored to the same degree that vehicle gap acceptance has been investigated. While similar in concept, there are a variety of pedestrian characteristics and caveats in the interaction between the pedestrian and vehicle modes that require separate pedestrian gap acceptance models.

Traditionally, literature on vehicle gap acceptance has used a constant value of *critical gap* (CG) that is calibrated for local conditions (Troutbeck and Brilon, 2002). By definition, the critical gap is the time between consecutive vehicles on the major road at which a vehicle waiting at the minor approach is equally likely to accept the gap or reject it. It can differ depending on the type of movement and the type of vehicle.

There are several ways for estimating CG from field data, including a graphical method (Troutbeck and Brilon, 2002), a regression method (Troutbeck and Brilon, 2002), a statistical method based on maximum likelihood estimation (Troutbeck, 2001), and the Ramsey-Routledge method (ITE, 2010). In application of these methods, the capacity of the minor street flow (in veh/hr) becomes a function of the CG on the minor approach t_c, the follow-up time on the minor approach t_f and the conflicting major street flow q_p as shown in Equation 1, adopted from the HCM2010 for Two-Way Stop-Controlled intersections (TRB, 2010).

$$Capacity = \frac{3{,}600 \times q_p \times e^{-q_p t_c}}{1 - e^{-q_p t_f}}$$

Equation 1

The *follow-up time* is defined as the time needed for additional vehicles in a stored queue to accept the same gap. The size of t_f is typically less than t_c, because some of the decision and acceleration times for subsequent vehicles occur during the initial gap.

In addition to deterministic gap acceptance, a report compiled for the Federal Highway Administration (FHWA) Next Generation Microsimulation (NGSIM) research effort (Cambridge Systematics, 2004) discusses *probabilistic* gap acceptance models, for which the driver response for an identical event (same speed, same gap in conflicting traffic) can be drawn from a probabilistic distribution of possible responses. Such *probit* models assume a mean CG with a random variance term depending on the specific coefficients defined for a driver and/or situation. Conceptually, probit models could represent inconsistent driver behavior and a heterogeneous population by drawing gap acceptance decisions from random distributions.

Alternatively, probabilistic behavior can be modeled in the form of a *binary or multinomial logit* model. A logit model could describe the likelihood of gap acceptance as a function of a number of different parameters (for example assertive vs. non-assertive pedestrians, gap time, and type of the arriving vehicle). It thus introduces greater complexity in the gap acceptance model, but in turn requires a lot of data for calibration. Logit Gap Acceptance Models have been proposed by Ben-Akiva and Lerman (1985) and Cassidy (1995) and Probit Models were suggested by Mahmassani and Sheffi (1981) and Madanat (1994).

Some researchers have proposed even more complex algorithms for modeling gap acceptance. Kita (1993) used neural networks to describe the process, under the assumption that gap acceptance is not a linear sequence of events, but that multiple factors affect the decision making process. This modeling approach is capable of removing consistency assumptions, but the authors upheld the assumption of homogeneity.

Models for Pedestrian Gap Acceptance

The deterministic gap acceptance model in the HCM2010 (TRB, 2010) offers a method for estimating critical gap t_c as a function of crosswalk length L, Pedestrian Walking Speed S_p and pedestrian start-up time t_s (Equation 2).

$$\boldsymbol{t_c = \frac{L}{S_p} + t_s}$$

Equation 2

Rouphail et al. (2005) described pedestrian gap acceptance as the sum of latency (response delay) and actual crossing times, an approach similar to the HCM method discussed above. The authors used field estimates of the median latency time in place of the HCM start-up time. The authors' research compared latency times of blind and sighted pedestrians and found that blind pedestrians exhibited significantly larger latency times, resulting in longer critical gap values and presumably more delay. The increased delay to blind pedestrians is consistent with research findings presented above.

Researchers have also attempted to use advanced gap acceptance models to describe pedestrian crossings. Sun et al. (2002) calibrated probit and binary logit models to describe both pedestrian gap acceptance and driver yielding from actual field data. The authors excluded about 25% of observations for later model validation and found that binary logit models performed best in both cases, correctly predicting 85.6% of gap acceptance and 87.1% of yielding decisions. For comparison, a probit model only resulted in 68.5% correctly predicted gap acceptance decisions, and a deterministic critical gap model actually achieved a surprising 81.5% correct predictions. Regression analysis found the important factors for pedestrian gap acceptance to be gap size, number of pedestrians waiting, and age of pedestrians. The authors recommended the binary- logit model for estimation, stating that the good performance of the deterministic model was likely due to an extraordinarily homogeneous pedestrian population.

From observations in China, Yang et al. (2006) derived a pedestrian gap acceptance formulation for the critical gap (CG) of pedestrians. This Equation (3) is shown below, where L is the length of the crossing, S is the walking speed and F is a factor of safety based on the pedestrian's confidence.

$$CG = L/S + F \quad \textbf{Equation 3}$$

Similar assumptions for pedestrian gap acceptance were used in the analysis of unsignalized pedestrian crossings at roundabouts and channelized right turn lanes by Rouphail et al. (2005) and Schroeder et al. (2006), respectively. Schroeder (2008) developed logistic regression-based gap acceptance models for unsignalized crossings to better describe the process of pedestrian gap acceptance by incorporating vehicle dynamics, pedestrian characteristics and concurrent events at the crosswalk.

Driver Yielding Behavior

Driver yielding behavior has been linked in research to operational characteristics such as vehicle speeds (Geruschat and Hassan 2005), as well as geometric characteristics of the crosswalk location, for example; entry versus exit leg at a roundabout (Rodegerdts, 2007, Ashmead et al., 2005). But to date, these isolated studies of driver yielding behavior at unsignalized crosswalks have largely been descriptive, with little insight gained towards predicting driver yielding at such crosswalks. The rate of driver yielding to pedestrians at unsignalized crosswalks varies across locations (Rodegerdts, 2007), but in nearly all cases is less than 100%. A range of treatments exist that are intended to increase the rate of driver yielding (Fitzpatrick, 2006).

Findings of pedestrian operations at roundabouts by NCHRP report 572 (Rodegerdts, 2007) show that 43% of the drivers at two-lane approaches of the roundabout do not yield to pedestrians. The lack of yielding is only 17% for single-lane roundabouts. Lack of yielding is also higher at exit (54% not yielding) compared to 33% not yielding at the entry. Based on these findings, the number of lanes and crosswalk location (entry or exit) are the two design elements that affect pedestrian accessibility at roundabouts.

A study by Salamati et al. (2012) at six two-lane roundabout approaches across the country showed that the yielding rate varies from 0% to 85% at the exit and entry leg of roundabouts depending on pedestrian assertiveness to cross the street, pedestrian disability (blind or sighted), entry or exit leg of the roundabout and the study location.

In previous research, Sun et al (2002) collected data on driver yielding and pedestrian gap acceptance at an unsignalized midblock pedestrian crossing and compared the fit of different statistical models. The authors estimated yielding probabilities based on the discrete parameters of driver gender, driver age, type of vehicle, number of pedestrians and the presence of an opposing yield. They found that drivers are more likely to yield to a group of pedestrians and that older drivers were more likely to yield than younger drivers. Their results also showed that a logistics modeling approach outperformed a probit model for driver yielding, as well as for pedestrian gap acceptance. The authors collected 1.5 hours of data for each AM and PM peak over 5 days, for a total of 15 hours of data. The resulting samples included 687 accepted gaps, 938 rejected gaps and 1254 motorist yield data points, and were deemed sufficient to estimate statistically significant probit and logit models. However, the authors looked at only one crosswalk and did not analyze any pedestrian treatment effects.

The research findings above can be summarized in that the decision of a driver to yield is a function of both operational and behavioral parameters. In the first category, the yield decision is triggered by both the speed of the vehicle and the assertiveness of the pedestrian. In the behavioral category, drivers may be influenced by clothing and the number of pedestrians at the crosswalk. Similarly, it can be hypothesized that yielding is impacted by the presence of a conflict downstream of the crosswalk. There are also cases where a driver may be forced to yield, because of a pedestrian's decision to proceed with crossing in a too-short gap in traffic.

Summary

The literature review and synthesis reaffirms that there is a need to develop robust pedestrian gap acceptance and driver yield behavior models based on a broad set of data collected at various locations, and to gain a better understanding of the true dynamics of pedestrians and vehicles at unsignalized midblock locations. To fulfill this need, this study developed enhanced behavioral models based on empirical observations at midblock pedestrian crossings in Alabama, Florida, and North Carolina, as well as naturalistic driving data in Florida. Details on the study methodology and results follow.

CHAPTER 3: ANALYSIS FRAMEWORK

This chapter presents the analysis framework underlying this research. It discusses elements of pedestrian-vehicle interaction and the translation of these interaction parameters into algorithms suitable for integration in a microsimulation environment. The framework and algorithms concepts then form the basis for the data collection methodology described in Chapter 4.

Nature of Pedestrian-Vehicle Interaction

Pedestrian-vehicle interaction occurs whenever a desired vehicle movement and pedestrian movement conflict. This includes crosswalks at intersections, midblock crosswalks, and even jaywalking. The type of interaction varies based on geometry, legal restrictions, and driver and pedestrian judgment. At signalized intersections, pedestrians are given the right of way to cross the street during the pedestrian walk phase. However, it is more difficult to identify the pedestrian right-of-way at unsignalized midblock crossings. Frequently, state laws give pedestrians the right of way while crossing *within* the crosswalk, however, field observations indicate that these crossing events have mixed priority with less than 100% yield rate. At the same time, some drivers yield to pedestrians waiting *at* the crosswalk, even though they are not legally required to do so in most states. Ultimately, the yielding rate of drivers varies from one crosswalk to the other and across states and regions. Whenever drivers do not yield, pedestrians make gap judgments for crossing in-between successive vehicles.

This project focuses on identifying and modeling mixed-priority (some drivers yield to create crossing opportunities for pedestrians, other pedestrians use judgment to for accepting gaps) crossing events at unsignalized midblock locations. The crossing event must be clearly defined for both field observation and application of any models developed. The pedestrian-vehicle interaction event is defined by Schroeder and Rouphail (2011) as a pedestrian arriving at a crosswalk influence area (CIA) when a vehicle is approaching. Crosswalk influence area (CIA) is defined as the area in the proximity of the crosswalk that is within line of driver sight distance and pedestrians wait to cross the street. During the data collection, an observer records pedestrian and vehicle characteristics when the pedestrian has indicated intent to cross whether in a gap in traffic or in front of a yielding vehicle. The driver's reaction is then classified as either yielding by slowing and/or stopping to allow the pedestrian to cross, or not yielding. While it is assumed that pedestrians will utilize any crossing opportunity, certain pedestrian populations, such as visually impaired, may not be able to identify and therefore utilize some crossing opportunities. Therefore, yield or gap utilization rate by pedestrians is not typically 100%.

Pedestrian-Vehicle Interactions Modeling Approach

Driver yielding and pedestrian gap acceptance are choices which are discrete in nature. Discrete choice modeling of yielding has been used in prior research by Sun et al. (2003) and Schroeder and Rouphail (2011). A pedestrian's choice to accept a gap can be modeled as a binary choice of accepting or rejecting. In the case of pedestrians, a binary choice model can be considered to identify the probability of accepting a gap for a given set of conditions.

A driver's decision to yield can also be broken down into a binary choice of yielding or not yielding for a pedestrian, but the type of yield can also be separated by a complete stop (called hard yield) and a rolling stop (called soft yield). When considering the final three outcomes of no yield, soft yield, and hard yield there are many discrete choice model structures that can be considered. There are three major structures used to model discrete choices with more than two outcomes: multinomial logit, ordered logit, and nested logit. Each of the models are discussed in the following subsections.

Binary Logit and Probit Models

In binary logit models, decision maker q considers choices *i* and *j*. Decision makers select between the choices based on their utility U_{qi} and U_{qj}. Decision maker *q* has both a systematic utility and random utility as shown in Equation 4. The systematic utility for choice i can be estimated using Equation 5 using *n* variables. Variables that describe attributes of the two alternatives are used in V_{qj}, while variables that describe decision maker characteristics and the constant β_0 are not included in V_{qj}. The random component of utility ε_q is assumed to be distributed according to the standard logistic function. The probability of decision maker q making choice i is then shown in Equation 6. SAS uses maximum likelihood estimation to find values of β that best estimate the decisions observed in the dataset.

$$U_q = V_q + \varepsilon_q \qquad \textbf{Equation 4}$$

$$\boldsymbol{V_{qi} = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n} \qquad \textbf{Equation 5}$$

$$\boldsymbol{P_q(i) = \frac{e^{V_{qi}}}{e^{V_{qi}} + e^{V_{qj}}}} \qquad \textbf{Equation 6}$$

The binary probit model uses identical formulation with the assumption that the random component of utility ε_q is assumed to be distributed according to the standard normal distribution. This results in a probability of choosing alternative *i* $P_q(i) = \Phi(V_{qi})$

Multinomial Logit Model

In multinomial logit models (MNL), decision maker *q* considers *K* alternatives from the choice set $C_q = \{1, 2, \ldots, K\}$. Decision makers select between the choices based on their utility U_{qi} for each alternative *i*. Decision maker *q* has both a systematic utility and random utility as shown in Equation 7. The systematic utility for choice *i* can be estimated using Equation 8 using *n* variables. Variables that describe attributes of the alternatives are used in V_q for each alternative *i*, while variables that describe decision maker characteristics and the constant β0 are not included in the final utility function V_{qK}. The random component of utility ε_q is assumed to be distributed according to the standard logistic function. The probability of decision maker *q* making choice *i* is then shown in Equation 9. SAS uses maximum likelihood estimation to find values of β that best estimate the decisions observed in the dataset.

$$\boldsymbol{U_{qi} = V_{qi} + \varepsilon_{qi} \forall\, \mathrm{i} \in C_q} \qquad \textbf{Equation 7}$$

$$\boldsymbol{V_{qi} = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n} \qquad \textbf{Equation 8}$$

$$\boldsymbol{P_q(i) = \frac{e^{V_{qi}}}{\sum_{j \forall \mathrm{j} \in C_q} e^{V_{qj}}}} \qquad \textbf{Equation 9}$$

Ordered Probit Model

In the ordered probit formulation of the ordered response model, decision maker *q* considers *K* alternatives from the choice set $C_q = \{0, 1, \ldots, K\text{-}1\}$. Decision makers select choice *k* if and only if their propensity (in this case, propensity to yield) U_q is between boundaries ψ_{k-1} and ψ_k, where $\psi_{-1} = -\infty$ and $\psi_{K-1} = \infty$. Decision maker *q* has a systematic and unknown propensity as shown in Equation 10. ε_q is assumed to be distributed according to the standard normal distribution, while V_q is estimated for n parameters as shown in Equation 11. Equation 12 shows how the probability for decision maker *q* is calculated under this assumption, where $\Phi(x)$ represents the CDF of the standard normal distribution. SAS uses maximum likelihood estimation in order to find the values of ϕ and β that can best recreate the observed choices.

$$U_q = V_q + \varepsilon_q \qquad \textbf{Equation 10}$$

$$V_q = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \qquad \text{Equation 11}$$

$$P_q(k) = \Phi(\psi_k - V_q) - \Phi(\psi_{k-1} - V_q) \qquad \text{Equation 12}$$

Nested Logit Model

The nested logit model (NL) is a special case of the multinomial logit model. (Wen and Koppelman, 2001) In the multinomial logit model, all outcomes are considered independent, while the nested model incorporates correlated outcomes. Correlated outcomes are grouped into "nests", and each nest is considered independently. In the nested logit model, Equation 13 describes the decomposition of the probability of alternative n as the sum of the product of P_m, the probability of nest m, and $P_{n/m}$, the probability of choice n if nest m is selected. P_m and $P_{n/m}$ are described in Equation 14 and 15 respectively. In these equations, $\alpha_{n\prime m}$ characterizes the portion of alternative n that is assigned to nest m, and must be greater than zero and sum to one across all nests. $V_{n\prime}$ is the utility for alternative n, while μ_m is the logsum parameter for nest m, with a value between zero and one.

$$P_n = \sum_m P_{n/m} P_m \qquad \text{Equation 13}$$

$$P_m = \frac{\left(\sum_{n' \in N_m} \left(\alpha_{n'm} e^{V_{n'}}\right)^{1/\mu_m}\right)^{\mu_m}}{\sum_m \left(\sum_{n' \in N_m} \left(\alpha_{n'm} e^{V_{n'}}\right)^{1/\mu_m}\right)^{\mu_m}} \qquad \text{Equation 14}$$

$$P_{n/m} = \frac{\left(\alpha_{n'm} e^{V_{n'}}\right)^{1/\mu_m}}{\sum_{n' \in N_m} \left(\alpha_{n'm} e^{V_{n'}}\right)^{1/\mu_m}} \qquad \text{Equation 15}$$

Model Selection

Pedestrian gap acceptance will be modeled using binary logit or probit models and then simulated in a microsimulator. The pedestrian decision to begin crossing has two distinct outcomes (GO = 1, NoGO = 0). The driver yielding behavior is more complex and the desired model formulation must reflect the assumptions of the driver's available choices. In the MNL model, each alternative is assumed independent, but since hard yield and soft yield are both types of yield, this structure is not ideal. The ordered probit model works on the assumption of a natural order of alternatives, but the three choices for a driver do not have a clear order. The nested logit model provides for correlation between hard yield and soft yield without ordering choices, but the model parameters are somewhat difficult to interpret. Therefore a pseudo-nested logit model is used for driver yielding, where a binary logit model predicts the binary outcome (Yield = 1, Non-yield = 0) and a second binary logit model is used to predict the binary outcome (Hard Yield = 1, Soft Yield = 0 | Yield). SAS statistical software was used to determine the model parameters and significant effects by different selection methods. Variables can be added or dropped manually to arrive at a satisfactory model.

In order to determine which model most accurately represents the data, various test statistics can be examined. Parameter estimates, standard error of the estimate, p-value, odds ratio, and R-squared. P-value indicates the confidence level, where $p<0.05$ indicates a 95% confidence level. Slope parameters in exponential relationship can be interpreted through odds ratio of the parameter. For continuous variables, a one-unit increase in the variable results in an e^{β} increase in odds of the response variable. For binary explanatory variable the odds ratio is interpreted as increase in odds of response from when variable increases from levels 0 to 1. Odds ratio is increase in likelihood of response for variable assuming all other variables are fixed. R-square test statistic describes amount of variability in data that is explained by the model. Higher R^2 value indicates better model fit, but the statistic is inflated with addition of more variables. Adjusted R^2 penalizes the model for inclusion of additional variables and is a better measure for models with many independent variables. Max-rescaled R^2 statistic divides generalized R^2 by upper bound R^2 estimate to achieve statistic that ranges all the way from 0 to 1.

Simulation Modeling Approach

The pedestrian-vehicle interaction event sequence is depicted in Figure 1. In this chart and in subsequent charts in the document, the decisions which use basic physics or field observations are shown as diamonds, while models developed to describe decision making or reactions by drivers or pedestrians are represented as rectangles.

Figure 1 shows the process flow for vehicles approaching a midblock crosswalk, which begins once a vehicle is close enough to the crosswalk to identify a pedestrian waiting to cross, and ends once the vehicle passes through the crosswalk. In cases where the required deceleration rate (breaking) to stop at the crosswalk is larger than a maximum deceleration rate feasibly achieved by any car, future pedestrian arrivals are ignored (represented as “Required Decel > Max Decel”). The yielding decision described in the flowchart is limited to the lead vehicle in the platoon. The vehicles following a lead vehicle in the platoon would have no other choice but to stop once the lead vehicle in front of them has stopped to yield to a pedestrian. On the other hand, if the lead vehicle in the platoon decided not to yield to the pedestrians, field observations have shown that the propensity of a following vehicle yielding is low. It is important to note that once the yielding decision is made by the model (for the approaching driver) it is final and will not be revised or reconsidered in subsequent steps of the simulation.

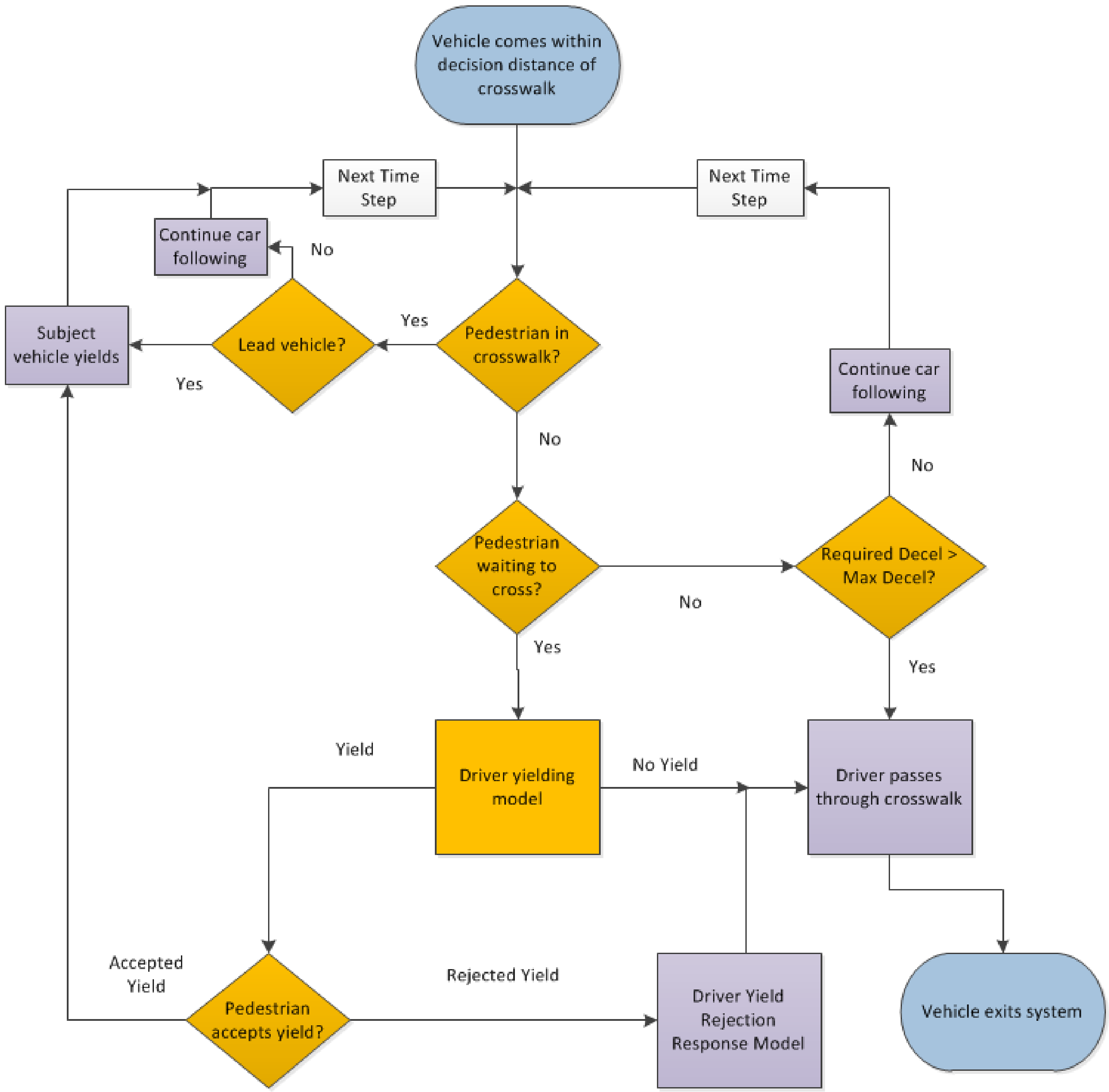


Figure 1: Vehicle Process Flow Approaching Midblock Crosswalk

Figure 2 shows the proposed process flow for a pedestrian crossing a single lane midblock crosswalk. Pedestrians either cross in a gap in traffic or in front of a yielding vehicle. If the pedestrian crosses in a gap, the gap or lag should be long enough to provide enough crossing time for the pedestrians (crossable lag or gap). A lag is defined as the time between the pedestrian arrival time at the crosswalk and the arrival of the first vehicle. A gap is defined as the time between two consecutive vehicles crossing the crosswalk.

Overall, two types of crossing models are required, the process of accepting a gap or lag and the process of accepting a yield. While the first model is described in detail in this report, it is assumed that pedestrians will accept all yields by vehicles.

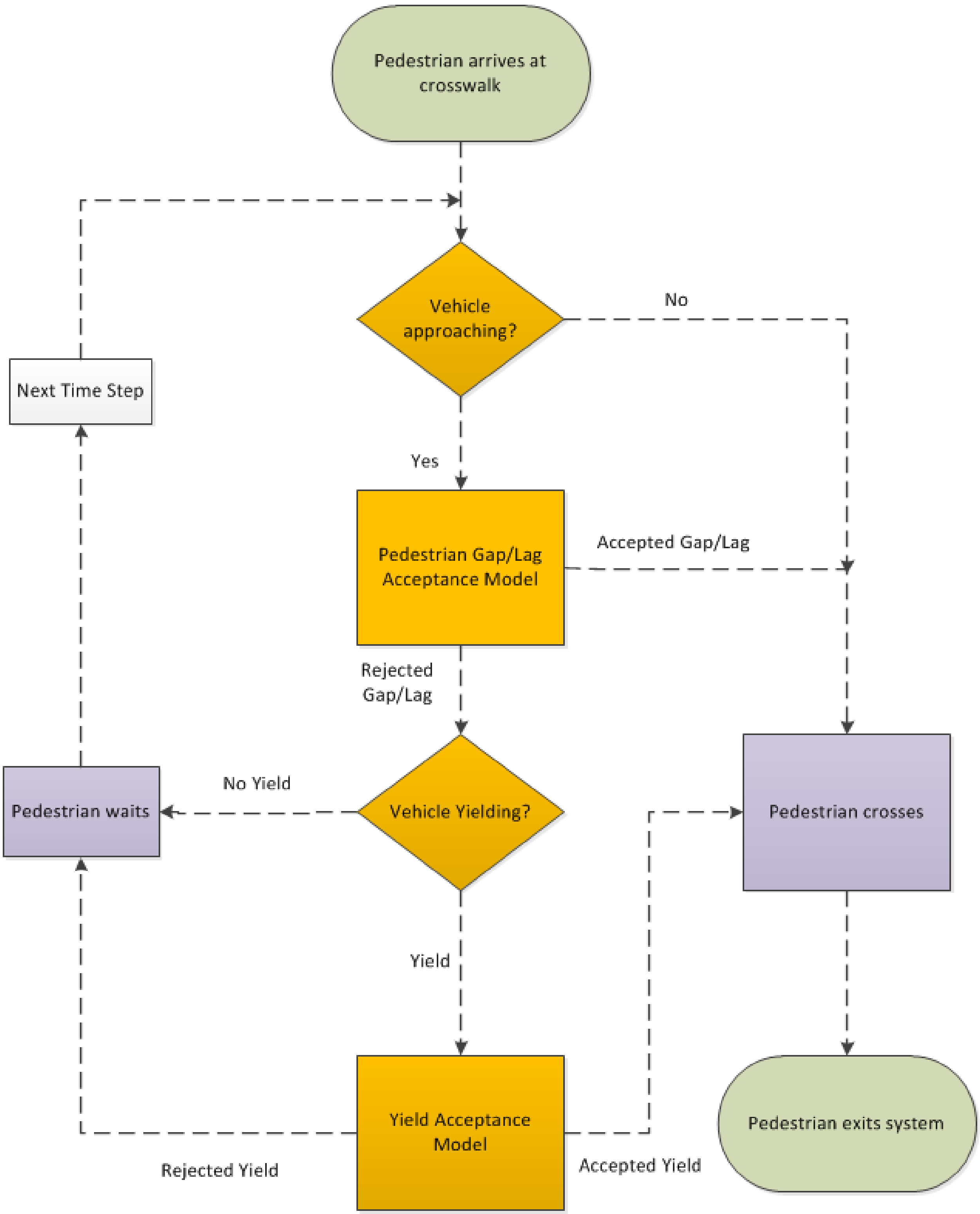


Figure 2: Pedestrian Process Flow for Single Lane Midblock Crossing

Due to multiple lanes of traffic, the pedestrian crossing decision making process is more complicated at multilane midblock crosswalks without a median pedestrian refuge, where pedestrians must identify a yield or acceptable gap in all lanes and both directions of traffic before leaving the sidewalk. With a refuge present, the pedestrian may cross in two stages and must identify a yield or acceptable gap in all lanes for a given direction before leaving the sidewalk or median. This project focuses on two-lane crossings with and without a pedestrian refuge in the median. For a crossing with refuge, it is assumed that the pedestrian makes the decision for crossing one direction of travel at a time. Figure 3 shows the pedestrian process flow for a two-lane midblock crossing without a pedestrian refuge in the median, while Figure 4 provides the process flow when a median refuge is present.

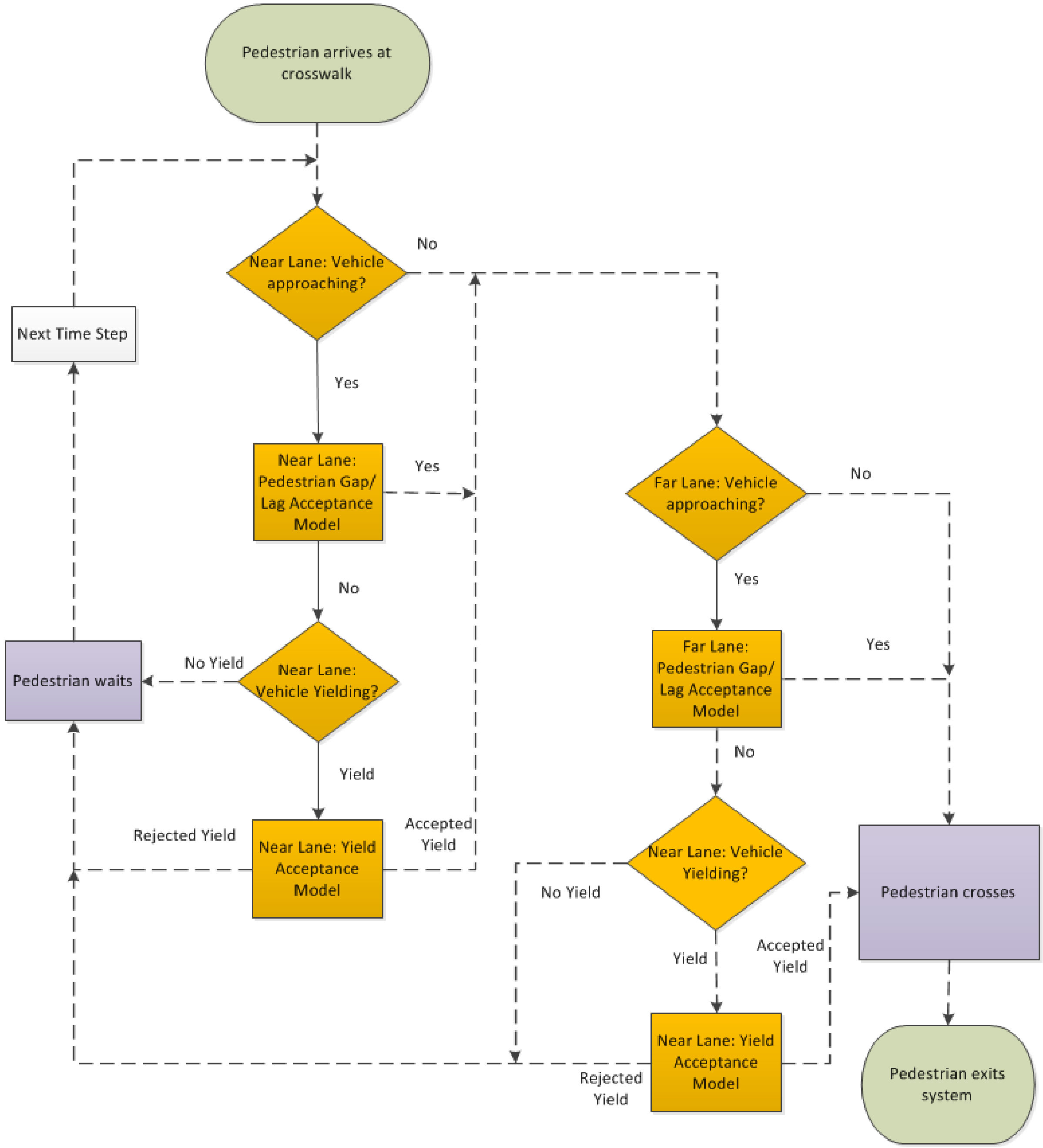


Figure 3: Pedestrian Process Flow for Two Lane Midblock Crossing without Median Refuge

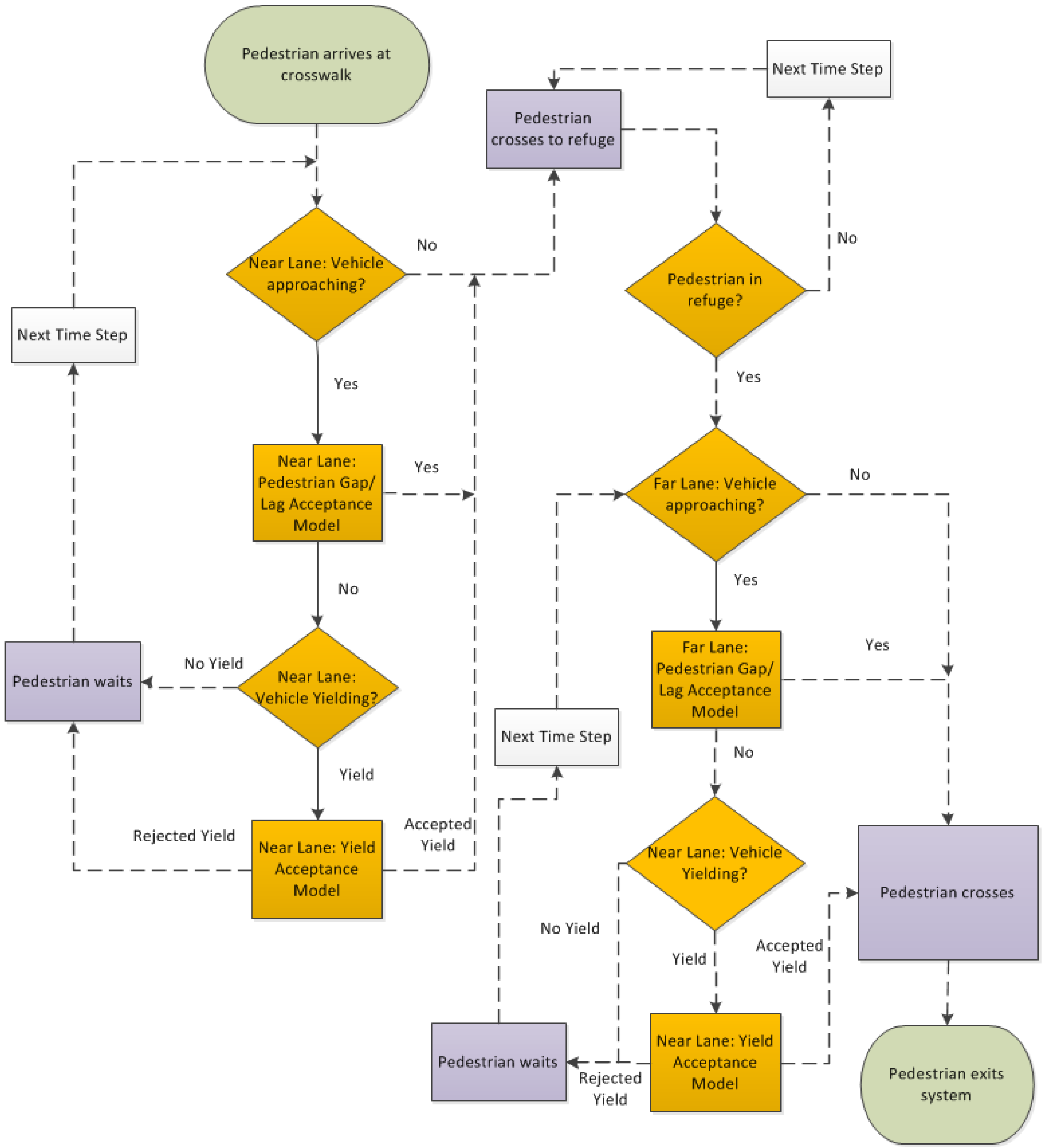


Figure 4: Pedestrian Process Flow for Two Lane Midblock Crossing with Median Refuge

The modeling approach developed is conveniently adaptable for many other geometric layouts and crossing types. More discussion on other implications of the models is provided in Chapter 7.

CHAPTER 4: DATA COLLECTION

A scan of relevant literature identified several research efforts that aimed at studying driver attitudes or pedestrian crossing behaviors. For obtaining data in such studies, three different data collection techniques have been adopted, namely observational, instrumented vehicle, and driving/pedestrian simulator approaches.

Observational studies are the most traditional method employed in the collection of empirical driving and pedestrian behavior data. They can be used to obtain data from attributes that are fixed (such as vehicle type, pedestrian characteristics, geometric characteristics, etc.), those that change dynamically (e.g., vehicle speeds, pedestrian speeds, distance headways, traffic signal indications, etc.) as well as to record qualitative observations (such as driver or pedestrian distraction). Observational data are typically obtained from trained observers with the help of tally sheets, count boards, video surveillance equipment, and radar detection devices.

Instrumented vehicles, on the other hand, permit quantitative assessment of driver performance in the field, under actual road conditions. Instrumentation in modern vehicles, including access to the on-board diagnostic (OBD) system, makes it possible to obtain information from the driver's own automobile, providing opportunities to study in depth driver strategy, vehicle usage, upkeep, drive lengths, route choices, and decision-making (Rizzo et al., 2002). The instrumentation enables researchers to record driver characteristics and vehicle operational parameters. Driver characteristics may include, among others, galvanic skin response, heart rate, and muscle activity. Previous research has used instrumented vehicle observations to develop lane changing models based on driver behavior (Sun and Elefteriadou 2010 & 2012), as well as freeway merging models (Kondyli and Elefteriadou 2012). Observations recorded include free flow speed, lane change frequency, and other driver actions. Examples of vehicle operating characteristics that can be gathered using an instrumented vehicle include steering motion, braking actions, speed, distance and triaxial accelerations (Helander and Hagvall, 1976).

Laboratory simulators can be also employed to assess behavior in response to synthetic reality. Driving simulators make it possible to observe driver behavior in controlled environments without the risk of driving on the road. They offer a cost-effective alternative to real world naturalistic studies and allow for independent variables to be systematically manipulated so that driver behavior can be measured precisely and safely (Rizzo et al., 2002). Since their introduction in the 1960s driving simulators have undergone many advances in terms of computing, visual display, and vehicle dynamics capabilities (Rudin-Brown et al., 2002). Even the lower fidelity simulators are able to collect vast amounts of data, which is one of their reported advantages over naturalistic investigative methods (Moroney and Lilienthal, 2009). Typical dependent measures of driving performance that are collected in driving simulation research studies include vehicle speed, acceleration, braking reaction time, and lane position. Similar to the driving simulators, pedestrian simulators also exist that can be used to study pedestrian behavior in controlled environments.

Additional details on the merits and shortcomings of each approach are available in Appendix A. The following sections discuss the details of an observational study, and an instrumented vehicle study that were performed in this project in order to supply data required for the study of vehicle-pedestrian interactions at midblock pedestrian crossings.

Observational Study

The observational study consisted of observation of actual pedestrian crossings and the respective driver actions. The benefit of conducting an observational study is that it allowed for direct observation of vehicle type, pedestrian type, gap size, pedestrian-vehicle conflicts, etc. An observational study is also suitable for gathering data to determine the percentage of driver yielding, average observed speeds, and pedestrian delay.

Several candidate unsignalized crosswalk locations were visited initially in order to select suitable data collection locations. Several midblock locations with varying pedestrian treatments, lane configurations, built-up environment, and travel activity from Alabama, Florida, and North Carolina were considered. Some sites had little naturally occurring pedestrian interactions with vehicular movements. For obtaining pedestrian vehicle interactions at such locations, staged crossings were performed by research members, while another member recorded detailed vehicular data. Staged crossings were distinguished from naturally occurring pedestrian crossings in the analysis – especially for the purpose of pedestrian gap acceptance, to avoid bias resulting from studying the experimenters. The nature of data collection required real time gathering of vehicle dynamics, with the majority of other variables extracted from post processing of video-records.

Data Collection Methodology

The research team collected a variety of empirical data on pedestrian-vehicle interaction at each study site. The data consisted of two kinds of attributes, namely attributes that were changing dynamically (e.g. vehicle speed, vehicle relative distance to the crosswalk), and others that were static descriptors of the pedestrian-vehicle interaction event (e.g. vehicle type, pedestrian characteristics). Members of the research team met early on, through conference calls and web meetings, in the data collection process to discuss these variables and any misunderstandings or disagreements. A data collection manual for consistent interpretation of these variables, events, and site locations was prepared and is provided in Appendix B.

The temporal beginning of an interaction event was defined as follows:

> *A pedestrian-driver interaction event commences as a pedestrian arrives in the crosswalk influence area (CIA) or waiting location while a driver is on the approach of the crosswalk.*

All interaction variables were coded relative to the above reference point in time, and the methodology assumed that the following statements are true:

- The pedestrian indicated his or her intention to cross at the facility (rather than continuing along the sidewalk).
- The pedestrian was aware of the approaching vehicle and decided whether or not he or she felt comfortable to cross the road.
- The driver was aware of the pedestrian’s intention and had to react in some fashion (make the decision to yield or continue through the crosswalk).
- The observer was aware that an event sequence (action-reaction) was about to take place (from video observation) and recorded the attributes of the interaction event.

The assumptions above were valid if both driver and pedestrian were consciously aware of each other’s presence. Clearly, there are cases where that was not true, as driver or pedestrian may be distracted or their intentions may not be clear and/or consistent. In an observational study the cognitive awareness of the involved parties is not discernible, but can only be presumed from erratic or unexpected behavior. For example, a pedestrian may step into the roadway and then retreat quickly realizing that he or she misjudged the position of the vehicle. Similarly, a driver may perform an emergency braking maneuver after belatedly recognizing the presence of the pedestrian. In the case of a pedestrian retreating, this event was coded as a separate event. Speed and other vehicle dynamics were recorded and a note such as “pedestrian pull-back” was made.

From the onset of a pedestrian-driver interaction event, there were three potential outcomes to the interaction of the two modes:

1. <u>Pedestrian GO Decision [GO]</u> – The pedestrian decides that there is sufficient time for a safe crossing and steps into the crosswalk.

2. Pedestrian NO-GO Decision [NOGO]/ Driver Non-Yield Decision [NY] – The pedestrian decides that the time until the expected vehicle arrival time to the crossing point is too short to safely cross the facility, i.e. he/she rejects the lag or gap. At the same time, the driver decides that it is either physically impossible to yield to the pedestrian, or he/she is unwilling to yield.
3. Driver Yield Decision [Y] – The approaching driver decelerates and creates a crossing opportunity for the pedestrian, which may occur with or without coming to a complete stop.

The team carried out a three-pronged data collection approach that combined real-time observations by a trained observer on a tally sheet, video recording of the crosswalk, and laser speed gun (LIDAR) speed measurements of approaching vehicles. The LIDAR was used to record the speed and distance of the approaching vehicle once the pedestrian (waiting at the curb or in the crosswalk) entered the view of the driver. A video camera was set up on a tripod to capture the pedestrian-vehicle interactions so that the researchers could gather additional data at a later point. Researchers made an effort to make themselves and the equipment inconspicuous by hiding behind trees, bushes, and poles, or beside parked vehicles, when these were available at the site. The researcher taking the speed and distance measurements would read these values out loud, so that the interaction could be quickly identified in the video. Figure 5 shows a schematic of the data collection set-up.

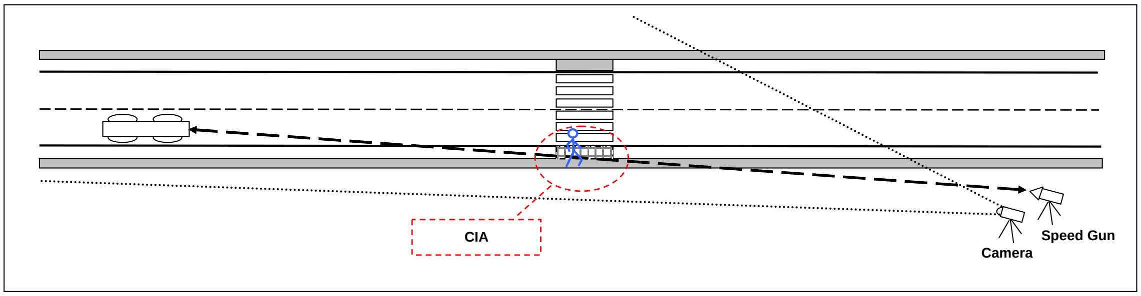


Figure 5: Field Data Collection Set-Up

In order to capture all relevant data, the video angle was adjusted to cover events concurrent to the interaction, such as the presence of an adjacent yield or multiple pedestrians. As shown in the diagram, the video camera angle was wide enough to cover the crosswalk influence area (CIA) or waiting location, and the approach to the crosswalk. The LIDAR was positioned so that speed recordings from Lidar were visible on the video camera and to the field observer, or alternatively were recorded audibly on the video. This experimental setup was retained for controlled experiments for sites with low pedestrian activity. A slight delay between trigger pull and response was observed, but this was very minor. Researchers accounted for this error by pulling the trigger right before the pedestrian entered the CIA.

The data collection sheet was used to record data of interest in a consistent manner across the study sites. The heading of the data collection sheet provided the date/time, observer(s), distance to crosswalk (dist. to CW), intersection, approach, and crossing distance. The entire distance had to be visible in the video, so that walking speed may later be calculated using the crossing distance and the crossing time (TIME on the data collection sheet). Average speed had to be calculated from unimpeded speeds at each location. This could be done by collecting speeds and distances for vehicles where the speed is not affected by pedestrians or platoons. These speeds and distances (for a sample of 30 vehicles) were reported on a separate sheet of paper or along the edge of the data collection sheet. Table 1 summarizes the data collection measurements obtained and used in the modeling effort.

Additional variables were needed to determine gap acceptance behavior and are summarized in Table 2. A lag event was defined as occurring between a pedestrian who has just reached the crosswalk and the next vehicle to arrive at the crosswalk. A gap event occurred between successive vehicles while a pedestrian waits at the crosswalk. These variables did not have values for all observations. While in soft and hard yields, the driver decided to yield to the pedestrian, for triggered yields, the pedestrian utilized what they saw as an acceptable gap size. Crossing events that were listed as gaps under the pedestrian variable CROSS have at least one observation.

Table 1: Data collection measures obtained and used in modeling

	Factor	Description	Value
First Vehicle Variables	SPD	The speed of the first vehicle (mph), at the time the pedestrian arrives at crosswalk influence area (waiting location), recorded from speed gun	Mph
	DIST	The distance from the first vehicle to the researcher recorded from the laser speed gun	Ft
	YIELD	Whether the first vehicle yielded and if it was a hard or soft yield	No=NY, Soft= SY, Hard=HY
	NEAR	Whether the vehicle for which speed and distance was recorded (first vehicle) was in the lane nearest or farthest from the pedestrian	Near=1, Far=0
	TRIG	If the first vehicle yielded, was it triggered (forced) by the pedestrian. In other words, if the yield happened before pedestrian stepping into the crosswalk (0) or after (1)	Triggered Yield, Yes=1, No=0
	STP	Whether the first vehicle had already stopped at the time that the pedestrian arrived	Stopped=1
	ADJ	Whether there was a yield on the other side of the road (opposite direction) or a yield in an adjacent lane (same direction)	Adj. Yield=1
	PLT	If the first vehicle was in a platoon or had a close follower	Platoon=1
	LSPLT	If the first vehicle was in a platoon or had a close follower and was travelling at a speed less than or equal to 15 mph	Low-speed Platoon=1
	HGV	First vehicle type: passenger car or heavy vehicle (bus or truck)	Heavy Vehicle=1
Pedestrian Variables	MUP	If there were other pedestrians present near the crosswalk; if any pedestrian is at either sides of the street or the splitter island and intends to cross	Multiple Pedestrian=1
	MED	Whether the pedestrian crossed from the median or the curb	Median=1, Curb=0
	CTRL	Whether the crossing pedestrian was controlled (researcher) or random (observational study)	Controlled=1, Random=0
	CROSS	Whether the pedestrian crossed in a gap or a yield	Gap/Yield (G/Y)
	IN_CW	Whether the pedestrian stopped in the crosswalk or at the curb. This variable shows the behavior of the pedestrian. A passive pedestrian is defined to wait at the curb for a crossing opportunity however an assertive pedestrian is defined to be waiting in the crosswalk or walking toward the crosswalk	Crosswalk=1, Curb=0

	AGE	Researcher's estimate of the pedestrian's age group	Young=1, Middle/Oldr=0
	DISTR	Whether the pedestrian was distracted by an outside source, such as a cell phone	Distracted, Yes=1, No=0
	BUSINESS	Researcher's observation of the pedestrian's attire or clothing	Business=1, Casual=0
	FEMALE	Pedestrian's gender	Female=1, Male=0
Site Variables	CAMPUS	This variable distinguishes sites on-campus (1) from those off-campus (0)	On-Campus=1
	FLORIDA	This variable distinguishes sites in the state of Florida (1) from those in the other two states (0)	Florida=1
	NCAROLINA	This variable distinguishes sites in the state of North Carolina (1) from those in the other two states (0)	North Carolina=1
	Distance to Crosswalk	The distance from the researcher using the laser speed gun to the middle of the crosswalk along the curb.	Ft
	Crossing Distance	The distance from the curb to a measured location, such as a specific white crosswalk marking, a center line, or the opposite side of the crosswalk.	Ft
Other	COUNT	If the first vehicle did not yield, how many vehicles went through before the pedestrian crossed	Number
Video	DIST_DEL	Delay between when the speed should have been taken (time pedestrian arrives at the waiting location) and actual measurement	Seconds
	ADJDIST	Vehicle position at the time of pedestrian arrival in crosswalk influence area measured in feet using a LIDAR speed measurement device; ADJDIST is calculated from measured distance, speed, distance delay and Distance to Crosswalk; ADJDIST=DIST+SPD*1.467*DIST_DEL-Distance to Crosswalk	Ft
	TTC	Time until vehicle would theoretically arrive at the crosswalk; TTC is calculated from the measured speed and distance at the time pedestrian arrives in the crosswalk influence area; TTC=ADJDIST/(SPD*1.467)	Seconds
	DECEL	Deceleration rate necessary to come to a full stop prior to crosswalk; DECEL is calculated from measured speed and adjusted distance; DECEL=(SPD*SPD)/(2*ADJDIST)	Ft/s^2

Table 2: Additional data collection elements used in determining the gap acceptance behavior

Factor	**Description**	**Value**
GO	Whether the pedestrian accepted (GO) or rejected (No-GO) a gap/lag event.	GO=1, No-GO=0

LAG	Whether the pedestrian event is a lag or gap	Lag=1, Gap=0
TIME	Time from the pedestrian stepping into the crosswalk to reaching a measured location (such as a specific white crosswalk marking, a center line, or the opposite side of the crosswalk). This value is used in calculating the pedestrian walking speed.	Seconds
W_SP	Pedestrian walking speed while crossing	Ft/s
OBS	Observed lag or gap time, measured with a stopwatch between pedestrian and vehicle arrival (lag) or successive vehicles (gap)	Seconds
DELAY	Time from the pedestrian arriving at the crosswalk influence area (waiting location) to stepping into the crosswalk to cross	Seconds

Study Site Selection

The driver yielding behavior and pedestrian gap acceptance may vary significantly at locations with varying land use and facility designs. In order to achieve greater heterogeneity in the study, various mid-block crossing locations with various lane configurations, and demographic elements were selected as study sites. Site visits and/or using Google Earth served as a starting point to collect basic information such as number of lanes, adjoining land use, average annual daily traffic, other facilities in the vicinity etc. Each mid-block location was visited at different times during the day to evaluate whether the site had adequate pedestrian and vehicle activity to warrant data collection. A few sites had sparse pedestrian and / or vehicle activity and thus were dropped from the initial candidate sites inventory. Sites with more than three lanes were avoided because the camera would have to be setup further from the crosswalk to capture the interaction. Drivers are forced to reduce their speed when approaching a raised crosswalk, so sites with raised crosswalks were not considered.

A total of 27 sites were selected for data collection and analysis (9 in Alabama, 10 in Florida, and 8 in North Carolina). The selected sites included mid-block locations with varying lane configurations, pedestrian and vehicular volumes were selected for data collection effort. 11 out of these 27 sites were university on-campus locations.

An example of a study site from North Carolina is depicted in Figure 6. Posted speeds at these locations ranged from 15 mph to 40 mph. Single lane and multilane configurations were considered in narrowing the final sites for data collection. Some of the sites also had bike lanes present, although bike traffic was negligible at these locations.

Figure 6: Example of study site and data collection setup

The complete inventories of data collection sites in Alabama, Florida and North Carolina are provided in Table 3, Table 4, and Table 5 respectively.

Table 3: Data collection sites in Alabama

No.	Crosswalk Location	City	No. of Lanes	Description	Type
1	University Blvd. and Hackberry Lane	Tuscaloosa	Two	Bike lanes	On
2	Richard Arrington Blvd. and 7th Ave N.	Birmingham	Three (one way)	Curbside Parking	Off
3	Greens Springs Hwy. and 24th Ave S.	Birmingham	Three (one way)	Transit stop near crosswalk	Off
4	10th St. S. and 10th Ave. S.	Birmingham	Four	Left turn lane in each direction	Off
5	7th Ave. and Campus Dr. (A)	Tuscaloosa	Two	Bike lanes	On
6	7th Ave. and Campus Dr. (B)	Tuscaloosa	Two	Bike lanes	On
7	University Blvd. and Colonial Dr.	Tuscaloosa	One (one way)	Bike lane alongside	On
8	Campus Dr. E. and 5th Ave.	Tuscaloosa	Two	Transit stop near crosswalk, Bike lanes	On
9	Ridge Rd. and Oxmoor Rd.	Homewood	Two	School near crosswalk	Off

Table 4: Data collection sites in Florida

No.	Crosswalk Location	City	No. of Lanes	Description	Type
1	Gale Lemerand Dr.	Gainesville	Two	Two stage, Flashing Ped sign, Bike lanes	On
2	Museum Rd. & Fraternity Row (EB)	Gainesville	Two	Bike lanes	On
3	Museum Rd .& Fraternity Row (WB)	Gainesville	Two	Bike lanes	On
4	Museum Rd. & Reitz Union Dr.	Gainesville	Two	Bike lanes, In-road sign	On
5	Hull Rd.	Gainesville	Two	Bike lanes	On
6	Museum Rd. & SW 13th St.	Gainesville	Two	Two stage, Ped sign, Bike lanes	On
7	SW 2nd Ave. & SW 8th St.	Gainesville	Two	Two stage, In-road sign, On-street parking, Bike lanes	Off
8	SW 2nd Ave. & SW 3rd St.	Gainesville	Two	Two stage, In-road sign, On-street parking, Bike lanes	Off
9	SE 2nd Ave. & SE 6th St.	Gainesville	Two	On-street parking, Bike lanes	Off
10	SW 2nd Ave. & SW 1st St.	Gainesville	Two	Two stage, In-road sign, On-street parking, Bike lanes	Off

Table 5: Data collection sites in North Carolina

No.	Crosswalk Location	City	No. of Lanes	Description	Type
1	Fayetteville at Hargett & Martin	Raleigh	Two	On-street parking	Off
2	Fayetteville at Morgan & Hargett	Raleigh	Two	On-street parking	Off
3	Wilmington at Hargett & Martin	Raleigh	Two (one way)	On-street parking	Off
4	S. Elm at Washington & February 1	Greensboro	Two	Two stage, On-street parking	Off
5	S. Elm at Washington & McGee	Greensboro	Two	Two stage, On-street parking	Off
6	Dan Allen at Thurman & Cates	Raleigh	Two	Access gate	On
7	Sullivan at Dan Allen & Varsity	Raleigh	Two	In-road sign, Bus stop	On
8	Main Campus at Research & Campus Shore	Raleigh	Two	On-street parking	On

In-Vehicle Study

Data Collection Overview

The instrumented vehicle (in-vehicle) data collection was geared at providing inputs for driver behavior and vehicle dynamics, which can be used for simulating pedestrian-driver interactions at crossing locations, as well as understanding driver attitudes in the interaction process. The data collection approach is based on previous studies of important factors that impact driver yielding and pedestrian crossing behavior, such as vehicle speed, pedestrian characteristics, and other related elements. A total of 15 participants were involved in this experiment and their behaviors in terms of pedestrian interaction at crosswalks were recorded to address the need. The detailed information on participant's demographic characteristics and driving attitude can be found in Appendix F.

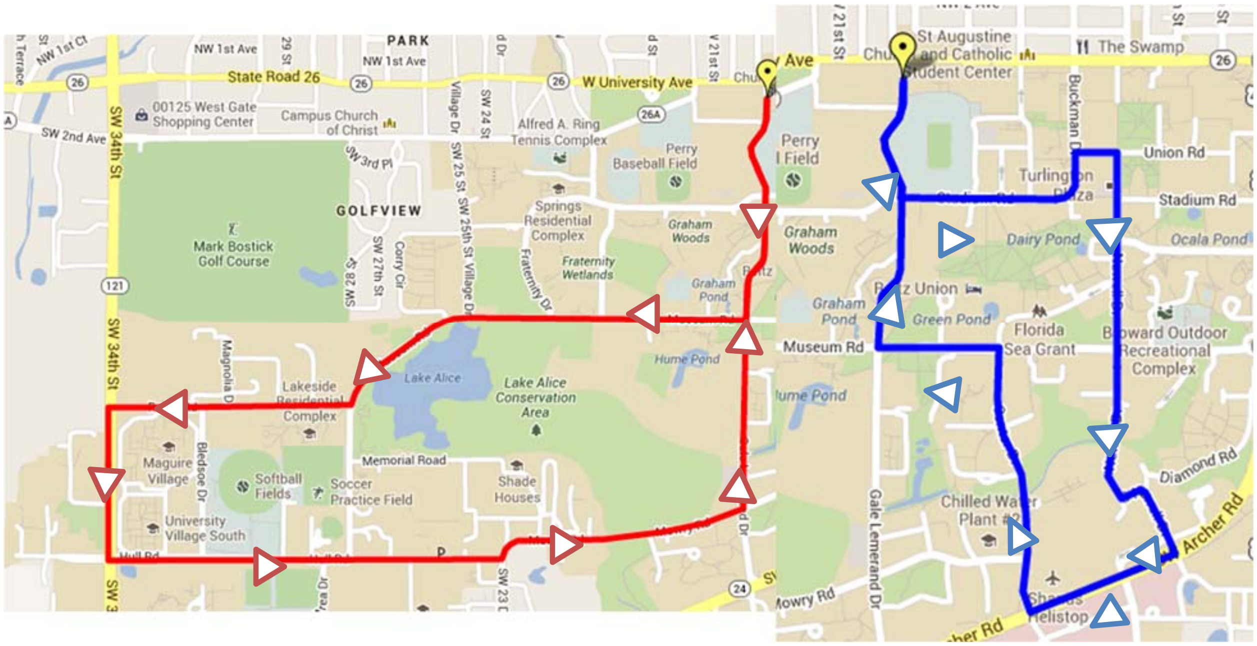
Participants and Driving Routes

In this study, 15 participants were recruited through advertisements publicized on the Craigslist website or posted around the Gainesville area, Florida (IRB form is attached in Appendix G). The participants were selected based on age, gender, driving experience, occupation, vehicle ownership through a prescreening questionnaire. Three criteria for participant recruitment were set: (1) must be a regular driver with driving experience no less than one year; (2) must be at least 25 years old; (3) must have a valid driver license. The criteria were set to make sure the subjects were adequately skilled drivers, and that the pool of subjects was reasonably diverse.

Two different routes on the campus of the University of Florida were selected for the in-vehicle study. These locations include approximately 18 midblock crossings each. Figure 7 provides the route maps for two routes, and information about each route is shown in Table 6 and Appendix C.

Table 6: Route Information for In-Vehicle Study

	Total Length	# Midblock Crossings	# Signal Crossings			Total Duration
			TR	LT	RT	
Route #1	4.7 mile	17	4	2	1	16 min
Route #2	2.8 mile	19	2	2	3	20 min



(a). Route #1 (b). Route #2

Figure 7: In-Vehicle Study Route Maps

Data Collection Process for Each Participant

Each participant was asked to meet our researchers at a pre-specified point and time. The researchers provided each subject with the informed consent form, pre-driving survey and the route maps. After a test drive, the subject began to drive along the two routes. After the completion of each route, he/she was asked to stop and complete a questionnaire regarding the driving actions throughout the route. Questions related to lane-changing, yielding, and actions around pedestrian walkways, bikeways, and transit vehicles. After the completion of all routes, a final questionnaire was used to summarize drivers' experiences during the entire experiment. The data collection was approximately 45-60 minutes for each participant, including a 5-minute discussion after the completion of each route.

Collected Data and Obtained Models

During the data collection, vehicle speed throughout the network, driver's car-following and lane changing behaviors, as well as the yielding and braking behaviors at crosswalks were all recorded by the cameras installed in the vehicle or by the researchers taking notes. Meanwhile, the driving attitudes were extracted from the questionnaires. The collected data were summarized as follows:

- Driver speed, acceleration, and vehicle trajectory throughout the experiment,
- Pedestrian characteristics, pedestrian presence at each crosswalk and the corresponding driver behavior,
- Traffic flow condition and the road environment, and
- Participants' demographic characteristics, general driving attitudes, urban campus issues and specific pedestrian interaction issues

Several sub-models used in simulation implementation were developed based on the obtained information from in-vehicle study, such as the models of driver decision point, and yield dynamics (results are presented in Chapter 5). Moreover, driver types and their aggressiveness, as well as their overall behavior along the test routes could be observed and further analyzed.

CHAPTER 5: EMPIRICAL MODELS

This chapter presents the modeling results after analyzing all collected data. The chapter first presents the yield modeling results, followed by gap acceptance data and model results. Finally, the chapter presents various other algorithms captured from the in-vehicle study.

Yield Models

Descriptive Statistics

Analysis of the data began with descriptive statistics of the variables, which led to a better understanding for the data trends and variability. The mean and standard deviation for all variables used in the process of yield modeling were calculated and provided in Table 7 and Table 8.

The values are provided for the overall data set, but are also broken into separate columns for different data sets based on location (on-campus or off-campus and Florida, Alabama, or North Carolina), as well as response type (yield or non-yield).

There were initially 1,178 observations to analyze, but certain observations were deleted when the decision to yield was made prior to the pedestrian arriving or the decision was forced by the pedestrian, which is defined as a “triggered yield”. Consequently, the sample size dropped to 975 observations for the modeling process.

For binary variables, the mean is equivalent to the rate at which this variable was observed. Therefore, for the response variables (HY, SY, and NY), the means are the observed yielding rates. The overall yielding rate was 53.3% (the sum of the HY and SY means). Higher yielding rates were seen at on-campus locations and at sites in Florida. Few events are associated with heavy vehicles (HGV, 5.4%), distracted pedestrians (DISTR, 1.2%), and pedestrians in business attire (BUSINESS, 5.1%). 8.6% of pedestrians were observed to “trigger” a yield (TRIG) by stepping into the roadway and 7.0% of drivers were stopped prior to the pedestrian arriving at the curb (STP). Observations associated with a triggered yield or a vehicle already stopped at the crosswalk were removed prior to any modeling efforts, since these drivers did not make the decision to yield for that particular observation.

The average observed speed (SPD) at the sites is 20.6 mph, with higher speeds being seen at off-campus locations, Alabama sites, and for non-yield events. Data showed that 63.2% of observations occurred when the pedestrian was on the near side of the vehicle (NEAR). 24.6% of the observed pedestrian crossings events occurred when pedestrians were traveling in a group of two or more people (MUP). 26.9% of yields occurred when the driver was part of a low-speed platoon of 15 MPH and lower (LSPLT). Non-yield events were associated with lower time-to-collision (TTC) and higher necessary deceleration rates (DECEL).

Table 7: Descriptive Statistics for the Yielding Study – Binary Variables

	All Data		CAMPUS				STATE						Response			
			On		Off		Florida		Alabama		N. Carolina		Yields		Non-Yields	
Variable	**Mean**	**StdDev**	**Mean**	**StdDev**	**Mean**	**StdDev**	**Mean**	**StdDev**	**Mean**	**StdDev**	**Mean**	**StdDev**	**Mean**	**StdDev**	**Mean**	**StdDev**
Sample Size*	1178		599		579		442		382		354		628		550	
Response Variables																
Hard Yield	0.237	0.425	0.367	0.482	0.102	0.303	0.473	0.500	0.126	0.332	0.062	0.242	0.444	0.497	0.000	0.000
Soft Yield	0.296	0.457	0.369	0.483	0.221	0.415	0.362	0.481	0.233	0.423	0.282	0.451	0.556	0.497	0.000	0.000
Non-Yield	0.467	0.499	0.264	0.441	0.677	0.468	0.165	0.372	0.641	0.480	0.655	0.476	0.000	0.000	1.000	0.000
Binary Factors																
NEAR	0.632	0.483	0.588	0.493	0.677	0.468	0.586	0.493	0.618	0.487	0.703	0.457	0.597	0.491	0.671	0.470
TRIG	0.086	0.280	0.080	0.272	0.092	0.289	0.093	0.290	0.076	0.265	0.088	0.283	0.156	0.363	0.005	0.074
STP	0.070	0.255	0.135	0.342	0.002	0.042	0.183	0.387	0.000	0.000	0.003	0.053	0.124	0.330	0.007	0.085
ADJ	0.121	0.326	0.177	0.382	0.062	0.242	0.213	0.410	0.055	0.228	0.076	0.266	0.189	0.392	0.042	0.200
PLT	0.383	0.486	0.412	0.493	0.352	0.478	0.502	0.501	0.267	0.443	0.359	0.480	0.451	0.498	0.305	0.461
LSPLT	0.163	0.370	0.244	0.430	0.079	0.271	0.315	0.465	0.018	0.134	0.130	0.337	0.269	0.444	0.042	0.200
HGV	0.054	0.225	0.083	0.277	0.022	0.148	0.059	0.236	0.018	0.134	0.085	0.279	0.059	0.236	0.047	0.212
MUP	0.246	0.431	0.372	0.484	0.116	0.320	0.398	0.490	0.128	0.335	0.184	0.388	0.371	0.483	0.104	0.305
MED	0.207	0.405	0.275	0.447	0.136	0.344	0.416	0.494	0.118	0.323	0.042	0.202	0.285	0.452	0.118	0.323
CTRL	0.547	0.498	0.367	0.482	0.732	0.443	0.507	0.501	0.623	0.485	0.514	0.501	0.441	0.497	0.667	0.472
AGE	0.809	0.393	0.928	0.258	0.685	0.465	0.998	0.048	0.534	0.499	0.870	0.337	0.895	0.307	0.711	0.454
DISTR	0.012	0.108	0.013	0.115	0.010	0.101	0.000	0.000	0.008	0.088	0.031	0.174	0.014	0.119	0.009	0.095
BUSINESS	0.051	0.220	0.020	0.140	0.083	0.276	0.000	0.000	0.037	0.188	0.130	0.337	0.040	0.196	0.064	0.244
FEMALE	0.423	0.494	0.367	0.482	0.481	0.500	0.397	0.490	0.202	0.402	0.695	0.461	0.455	0.498	0.387	0.488
CAMPUS	0.508	0.500	-	-	-	-	0.774	0.419	0.385	0.487	0.311	0.463	0.702	0.458	0.287	0.453
FLORIDA	0.375	0.484	0.571	0.495	0.173	0.378	-	-	-	-	-	-	0.588	0.493	0.133	0.340
ALABAMA	0.324	0.468	0.245	0.431	0.406	0.491	-	-	-	-	-	-	0.218	0.413	0.445	0.497
NCAROLINA	0.301	0.459	0.184	0.388	0.421	0.494	-	-	-	-	-	-	0.194	0.396	0.422	0.494

*All values found prior to removing observations with SPD=0, STP=1, TRIG=1

Table 8: Descriptive Statistics for the Yielding Study – Continuous Variables

	Variable	All Data	CAMPUS		STATE			Response	
			On	Off	Florida	Alabama	N. Carolina	Yields	Non-Yields
Sample Size		975	450	509	300	353	322	432	543
Continuous Factors									
SPD	Mean	20.585	18.398	22.461	18.367	24.748	18.090	17.580	22.977
	St. Dev.	7.115	6.192	7.322	6.536	7.629	4.482	5.923	7.082
	Min.	4	4	10	4	10	10	4	5
	Max.	51	40	51	36	51	30.2	44	51
ADJDIST	Mean	130.780	138.031	124.565	138.218	129.004	125.797	128.743	132.401
	St. Dev.	69.106	72.729	65.274	63.672	65.155	77.325	61.415	74.678
	Min.	3.370	3.370	3.900	3.370	23.700	3.900	3.370	5.900
	Max.	435.272	435.272	411.100	348.972	411.100	435.272	348.972	435.272
TTC	Mean	4.596	5.346	3.952	5.393	3.627	4.915	5.176	4.134
	St. Dev.	2.411	2.630	1.995	2.246	1.619	2.888	2.198	2.475
	Min.	0.230	0.230	0.233	0.230	0.740	0.233	0.230	0.277
	Max.	16.184	16.184	12.157	15.839	10.987	16.184	15.839	16.184
DECEL	Mean	2.829	2.547	3.070	3.356	2.891	2.270	2.332	3.224
	St. Dev.	2.832	3.094	2.565	3.520	1.779	2.957	2.259	3.163
	Min.	0.185	0.185	0.325	0.185	0.433	0.211	0.185	0.211
	Max.	38.015	38.015	21.521	38.015	16.118	21.521	31.911	38.015

*All values found AFTER removing certain observations (such as SPD=0, STP=1, TRIG=1)

Data Preparation

Hard Yield and Soft Yield

The behavioral models predict discrete decisions by the driver on whether to yield or not. In the case of a driver yielding to a pedestrian (or multiple pedestrians), the decision outcome can be a yield or a non-yield or the yield decision can be broken into a hard or a soft yield.

Validation

For each state, one site was removed from the dataset for model validation. There were 49 observations each for the validation site in North Carolina and Florida and 43 observations for the Alabama site.

In preparation for modeling, observations were also removed if the vehicle speed was 0 mph, the vehicle was already stopped when pedestrian approached crosswalk (STP=1), or if the yield event was triggered (TRIG=1). These were deleted because the driver did not make the decision for that observation, they either made the decision for a prior observation or the pedestrian forced the vehicle to yield. The total number of observations used in the modeling procedure was 975 (432 Yields and 543 Non-yields). Variables collected in the field or from video that did not apply to the driver decision to yield, such as pedestrian walking speed, were removed from the dataset. Florida was shown to have the most yielding while North Carolina had the least yielding. Based on this, it was decided that dummy variables would be created for these two states to capture the effects of these sites, while yielding in Alabama is captured in the intercept value.

Correlation tables were created throughout the modeling process when variables were added or removed from the data set. Many variables were shown to be significantly correlated with the dependent variable, as indicated by low p-values. The complete set of correlation tables can be found in Appendix D.

The following variables show a significant positive correlation with the dependent variable, suggesting an increase in yielding with an increase in the variable (or binary variable change from 0 to 1): ADJ, PLT, LSPLT, MUP, MED, TTC, AGE, FEMALE, CAMPUS, and FLORIDA.

The following variables show a significant negative or inverse correlation with the dependent variable, suggesting a decrease in yielding with an increase in the variable (or binary variable change from 0 to 1): SPD, NEAR, CTRL, DECEL, BUSINESS, and NCAROLINA. The high number of significant correlation coefficients shows that the yielding decision relies on several factors, rather than being the result of any single variable.

Only a few of the explanatory variables were intercorrelated. The variables that were intercorrelated are not surprising: NEAR to MED, PLT to LSPLT, and MUP to CTRL, for example. ADJDIST, TTC, and DECEL are also intercorrelated, which is expected since TTC and DECEL are calculated using the ADJDIST. Low sample sizes may show independent variables to be virtually perfect predictors of a certain response level. These variables included HGV, DISTR, and BUSINESS.

Logit Model Development

In order to create a model to predict the likelihood of a driver yielding to a pedestrian or pedestrians on a crossing event, SAS was used to analyze the data collected in Alabama, Florida, and North Carolina (except from the 3 sites that were reserved for validation purposes). Various types of models were tested, including binary logit, pseudo-nested logit, etc. The team decided to choose binary logit modeling, since it best describes the probability of yielding as a function of the variables collected. Binary logit models were created, taking into account microscopic traffic parameters related to the pedestrian crossing. In the first step of modeling, a correlation table was created to determine if any variables are intercorrelated, either with the response variable or with other independent variables. Correlation tables are also used to determine if certain variables are negatively or positively correlated with the dependent variable and how strong the relationship is. Table 9 below shows one of the correlation tables.

Table 9: Correlation Table for Yielding Data

Pearson Correlation Coefficients, N = 975
Prob > |r| under H0: Rho=0

	Y_NY	SPD	ADJDIST	NEAR	ADJ	PLT	LSPLT	HGV	MUP	MED	CTRL	TTC	DECEL	AGE	BUSINESS	FEMALE	CAMPUS	FLORIDA	NCAROLINA
Y_NY	1.00000	-0.37698 <.0001	-0.02631 0.4119	-0.06577 0.0400	0.19581 <.0001	0.09812 0.0022	0.22912 <.0001	0.01653 0.6061	0.24610 <.0001	0.21079 <.0001	-0.14696 <.0001	0.21460 <.0001	-0.15638 <.0001	0.23039 <.0001	-0.06696 0.0366	0.10427 0.0011	0.40015 <.0001	0.43875 <.0001	-0.21806 <.0001
SPD	-0.37698 <.0001	1.00000	0.36158 <.0001	0.05869 0.0670	-0.15163 <.0001	-0.09315 0.0036	-0.40952 <.0001	-0.14360 <.0001	-0.25649 <.0001	-0.02244 0.4840	0.31240 <.0001	-0.31806 <.0001	0.33970 <.0001	-0.24174 <.0001	-0.11683 0.0003	-0.25464 <.0001	-0.28481 <.0001	-0.20801 <.0001	-0.24644 <.0001
ADJDIST	-0.02631 0.4119	0.36158 <.0001	1.00000	-0.09417 0.0032	-0.02610 0.4156	0.02504 0.4347	-0.16035 <.0001	0.09996 0.0018	-0.09842 0.0021	0.09946 0.0019	0.01135 0.7233	0.70652 <.0001	-0.33511 <.0001	0.06402 0.0456	-0.02299 0.4733	-0.18821 <.0001	0.09719 0.0024	0.07179 0.0250	-0.05066 0.1139
NEAR	-0.06577 0.0400	0.05869 0.0670	-0.09417 0.0032	1.00000	-0.03318 0.3006	0.07319 0.0223	0.06364 0.0470	-0.02888 0.3677	-0.09352 0.0035	-0.54721 <.0001	0.30739 <.0001	-0.13092 <.0001	0.03724 0.2454	-0.01371 0.6690	-0.06026 0.0600	-0.02080 0.5165	-0.08528 0.0077	-0.10318 0.0013	0.11205 0.0005
ADJ	0.19581 <.0001	-0.15163 <.0001	-0.02610 0.4156	-0.03318 0.3006	1.00000	0.06436 0.0445	0.10513 0.0010	-0.03734 0.2441	0.09436 0.0032	0.12620 <.0001	-0.00967 0.7629	0.08887 0.0055	-0.04453 0.1647	0.05329 0.0963	-0.00673 0.8338	0.08597 0.0072	0.10380 0.0012	0.18445 <.0001	-0.04154 0.1950
PLT	0.09812 0.0022	-0.09315 0.0036	0.02504 0.4347	0.07319 0.0223	0.06436 0.0445	1.00000	0.45414 <.0001	0.09491 0.0030	0.03982 0.2141	0.04096 0.2013	0.03208 0.3170	0.09995 0.0018	-0.10581 0.0009	0.00049 0.9878	0.00803 0.8022	0.02182 0.4962	-0.04166 0.1937	0.07451 0.0200	0.02004 0.5320
LSPLT	0.22912 <.0001	-0.40952 <.0001	-0.16035 <.0001	0.06364 0.0470	0.10513 0.0010	0.45414 <.0001	1.00000	0.13305 <.0001	0.19131 <.0001	0.02624 0.4132	-0.14722 <.0001	0.17986 <.0001	-0.17452 <.0001	0.06698 0.0365	0.03337 0.2979	0.07596 0.0177	0.05657 0.0775	0.12566 <.0001	0.07814 0.0147
HGV	0.01653 0.6061	-0.14360 <.0001	0.09996 0.0018	-0.02888 0.3677	-0.03734 0.2441	0.09491 0.0030	0.13305 <.0001	1.00000	0.12388 0.0001	-0.01339 0.6762	-0.10871 0.0007	0.21782 <.0001	-0.08961 0.0051	0.08209 0.0103	0.01157 0.7182	0.01172 0.7148	0.13167 <.0001	0.01264 0.6934	0.11237 0.0004
MUP	0.24610 <.0001	-0.25649 <.0001	-0.09842 0.0021	-0.09352 0.0035	0.09436 0.0032	0.03982 0.2141	0.19131 <.0001	0.12388 0.0001	1.00000	0.01919 0.5495	-0.48996 <.0001	0.10353 0.0012	-0.09046 0.0047	0.03030 0.3447	0.03319 0.3005	0.03250 0.3106	0.22231 <.0001	0.18108 <.0001	-0.03623 0.2584
MED	0.21079 <.0001	-0.02244 0.4840	0.09946 0.0019	-0.54721 <.0001	0.12620 <.0001	0.04096 0.2013	0.02624 0.4132	-0.01339 0.6762	0.01919 0.5495	1.00000	0.05617 0.0796	0.11924 0.0002	0.03762 0.2405	0.22320 <.0001	-0.10065 0.0017	-0.02150 0.5026	0.14490 <.0001	0.41416 <.0001	-0.27307 <.0001
CTRL	-0.14696 <.0001	0.31240 <.0001	0.01135 0.7233	0.30739 <.0001	-0.00967 0.7629	0.03208 0.3170	-0.14722 <.0001	-0.10871 0.0007	-0.48996 <.0001	0.05617 0.0796	1.00000	-0.21989 <.0001	0.12273 0.0001	0.11919 0.0002	-0.28967 <.0001	-0.02401 0.4539	-0.29794 <.0001	0.00700 0.8271	-0.09311 0.0036
TTC	0.21460 <.0001	-0.31806 <.0001	0.70652 <.0001	-0.13092 <.0001	0.08887 0.0055	0.09995 0.0018	0.17986 <.0001	0.21782 <.0001	0.10353 0.0012	0.11924 0.0002	-0.21989 <.0001	1.00000	-0.55213 <.0001	0.19072 <.0001	0.05629 0.0790	-0.03702 0.2481	0.28831 <.0001	0.22051 <.0001	0.09311 0.0036
DECEL	-0.15638 <.0001	0.33970 <.0001	-0.33511 <.0001	0.03724 0.2454	-0.04453 0.1647	-0.10581 0.0009	-0.17452 <.0001	-0.08961 0.0051	-0.09046 0.0047	0.03762 0.2405	0.12273 0.0001	-0.55213 <.0001	1.00000	-0.03817 0.2338	-0.04083 0.2027	0.03580 0.2640	-0.09215 0.0040	0.12411 0.0001	-0.13859 <.0001
AGE	0.23039 <.0001	-0.24174 <.0001	0.06402 0.0456	-0.01371 0.6690	0.05329 0.0963	0.00049 0.9878	0.06698 0.0365	0.08209 0.0103	0.03030 0.3447	0.22320 <.0001	0.11919 0.0002	0.19072 <.0001	-0.03817 0.2338	1.00000	-0.23479 <.0001	0.22034 <.0001	0.28402 <.0001	0.34292 <.0001	0.15748 <.0001
BUSINESS	-0.06696 0.0366	-0.11683 0.0003	-0.02299 0.4733	-0.06026 0.0600	-0.00673 0.8338	0.00803 0.8022	0.03337 0.2979	0.01157 0.7182	0.03319 0.3005	-0.10065 0.0017	-0.28967 <.0001	0.05629 0.0790	-0.04083 0.2027	-0.23479 <.0001	1.00000	-0.03417 0.2865	-0.12197 0.0001	-0.15500 <.0001	0.24210 <.0001
FEMALE	0.10427 0.0011	-0.25464 <.0001	-0.18821 <.0001	-0.02080 0.5165	0.08597 0.0072	0.02182 0.4962	0.07596 0.0177	0.01172 0.7148	0.03250 0.3106	-0.02150 0.5026	-0.02401 0.4539	-0.03702 0.2481	0.03580 0.2640	0.22034 <.0001	-0.03417 0.2865	1.00000	-0.09455 0.0031	-0.01725 0.5906	0.40772 <.0001
CAMPUS	0.40015 <.0001	-0.28481 <.0001	0.09719 0.0024	-0.08528 0.0077	0.10380 0.0012	-0.04166 0.1937	0.05657 0.0775	0.13167 <.0001	0.22231 <.0001	0.14490 <.0001	-0.29794 <.0001	0.28831 <.0001	-0.09215 0.0040	0.28402 <.0001	-0.12197 0.0001	-0.09455 0.0031	1.00000	0.39022 <.0001	-0.24329 <.0001
FLORIDA	0.43875 <.0001	-0.20801 <.0001	0.07179 0.0250	-0.10318 0.0013	0.18445 <.0001	0.07451 0.0200	0.12566 <.0001	0.01264 0.6934	0.18108 <.0001	0.41416 <.0001	0.00700 0.8271	0.22051 <.0001	0.12411 0.0001	0.34292 <.0001	-0.15500 <.0001	-0.01725 0.5906	0.39022 <.0001	1.00000	-0.46814 <.0001
NCAROLINA	-0.21806 <.0001	-0.24644 <.0001	-0.05066 0.1139	0.11205 0.0005	-0.04154 0.1950	0.02004 0.5320	0.07814 0.0147	0.11237 0.0004	-0.03623 0.2584	-0.27307 <.0001	-0.09311 0.0036	0.09311 0.0036	-0.13859 <.0001	0.15748 <.0001	0.24210 <.0001	0.40772 <.0001	-0.24329 <.0001	-0.46814 <.0001	1.00000

All of the variables that are shown to be intercorrelated met experimenter expectations. Adjusted distance (ADJDIST), time to contact (TTC), and necessary deceleration rate (DECEL) are intercorrelated, which is expected since adjusted distance is used to calculate time to contact and necessary deceleration rate. A value of "1" for the low speed platoon variable (LSPLT) was determined from whether the vehicle was part of a platoon (PLT = 1) and if the vehicle speed (SPD) was less than or equal to 15 MPH, so it was expected that low speed platoon would be correlated with these variables. Vehicles in the lane near the pedestrian (NEAR = 1) are negatively correlated with pedestrians crossing from the median (MED = 1). This is due to the way the equipment was setup at the sites. Controlled crossings (CTRL = 1) are negatively correlated with the presence of multiple pedestrians (MUP = 1), which is reasonable since controlled crossings only have multiple pedestrians if an additional pedestrian showed up while the staged pedestrian was preparing to cross.

Most of the variables considered for the models showed significant correlation ($p<0.05$) to the dependent variable. Positive correlation suggests that an increase in yielding results from an increase in the value of the variable (or a change from 0 to 1 for binary variables). Negative correlation suggests a decrease in yielding from an increase in the value of the variable. ADJ, PLT, LSPLT, MUP, MED, TTC, AGE, FEMALE, CAMPUS, and FLORIDA showed positive correlation, while SPD, NEAR, CTRL, DECEL, BUSINESS, and NCAROLINA showed negative correlation.

In the next step, binary logit models were created to predict the likelihood of yielding. Y_NY is a binary variable that was created where a yield is represented by "1" and non-yield by "0." Models for the likelihood of yielding will be presented below based on different variable selection processes:

- Full Model – uses all independent variables regardless of their p-value.
- Forward selection – successively adds variables to the model at a $p<0.05$ inclusion threshold. This is an automated function in SAS.
- Backward elimination – starts with a full model and then removes variables starting with the highest p-value, until all remaining variables are at $p<0.05$.
- Manual selection – a custom model that is informed by the first three modeling results, but considers practical significance and feasibility of implementing variables in simulation (as opposed to just being motivated by statistical fit).

In addition to these binary models, additional models were created to show the likelihood of hard yielding rather than soft yielding. These models may be interpreted as a step-wise nested binary logit, where the first level (Y vs. NY) is predicted through one of the models above, and the second level (HY vs. SY) predicted from this model. A total of 432 observations (152 hard yields and 280 soft yields) were used for the hard yielding model.

Several different modeling approaches were applied that differ in number and type of parameters that are considered. Only 10 of the 975 observations involved a distracted pedestrian (DISTR = 1), so this variable was eliminated due to small sample size. Triggered yields, previously stopped vehicles, heavy vehicles, and pedestrians in business attire also showed small sample sizes, but were all greater than 5% of the data. The full model was used to identify the variables which were significant and how significant those variables were when compared to the other variables. Models formed through forward selection and backward elimination processes are expected to show the same variables to be significant and insignificant. Sometimes these processes produce slightly different results – especially when independent variables are intercorrelated. The table below shows a brief description of the models that were created in an effort to select a final model and a universal model (final model without state variables). Detailed modeling results are shown in Appendix D.

Table 10: Yield Model Summary

	Selection	Notes	R^2	Max-Resc. R^2
Y-1	Full	Excluding campus and state variables	0.2591	0.3470
Y-2	Forward*	Excluding campus and state variables	0.2553	0.3418
Y-3	Manual	All but TTC variable removed	0.2525	0.3382
Y-4	Manual	All but SPD and ADJDIST removed	0.1925	0.2578
Y-5	Forward*	North Carolina data only	0.2272	0.3248
Y-6	Forward*	Florida data only	0.1834	0.2779
Y-7	Forward*	Alabama data only	0.2916	0.4117
Y-8	Forward	On-campus data only	0.2246	0.3105
Y-8B	Backward	On-campus data only	0.2474	0.3420
Y-9	Forward	Off-campus data only	0.1424	0.2090
Y-9B	Backward	Off-campus data only	0.1543	0.2264
Y-10	Full	Campus and state variables added (FL and AL)	0.3648	0.4885
Y-11	Forward	Campus and state variables added (FL and AL)	0.3582	0.4797
Y-11B	Backward	Campus and state variables added (FL and AL)	0.3586	0.4802
Y-12	Full	LSPLT variable added	0.3691	0.4943
Y-13	Forward	LSPLT variable added	0.3613	0.4838
Y-14	Full	State variables changed to AL and NC	0.3691	0.4943
Y-15	Forward	State variables changed to AL and NC	0.3613	0.4838
Y-16	Manual	Y-2 with age switched out for campus	0.2967	0.3973
Y-17	Full	State variables changed to FL and NC	0.3691	0.4943
Y-18	Forward	State variables changed to FL and NC	0.3613	0.4838
Y-19	Manual	Final Driver Yielding Model	0.3582	0.4797
Y-20	Manual	Universal Model	0.2680	0.3589
HY-1	Forward	Excluding campus and state variables	0.2361	0.3249
HY-2	Forward	Campus and state variables added (FL and AL)	0.2577	0.3519
HY-3	Forward	LSPLT variable added (was not significant)	0.2577	0.3519
HY-4	Forward	State variables changed to AL and NC	0.2507	0.3449
HY-5	Forward	State variables changed to FL and NC	0.2507	0.3449

*Backward elimination showed same results as forward selection

For manual selection, variables were selected manually for use in the models. The variable selection was motivated by the ability to implement this model into a microsimulation environment, which is the primary objective of this project. The team therefore explored different model combinations that could more readily be implemented. In the first attempt at manual selection (Y-3), only the SPD and ADJDIST variables were used, and the interaction variable TTC was excluded. The team removed SPD and ADJDIST and instead only used the TTC variable for the next manual selection model (Y-4). Models Y-5, Y-6, and Y-7 were created from the datasets for each state to determine if behavior is different for each state. Since the models showed different variables to be significant, it was determined that dummy variables should be created for two of the states and have the third state represented in the intercept value.

For the final models, dummy variables were created for Florida and North Carolina, while Alabama was represented in the intercept value. Alabama was chosen to be represented in the intercept because it showed the middle level of yielding for the three states, with Florida having greater yielding and North Carolina having less yielding. A campus variable was included in model development since the on-campus data and off-campus data showed different variables to be significant in models (Y-8 and Y-9). A low-speed platoon variable was added (Y-12) to separate the effect of queued vehicles and other slower vehicle platoons from high-speed platoons. In Y-16 it was seen that CAMPUS had a more significant effect on driver yielding than AGE. Due to the low sample size of older pedestrians, it was decided that AGE should not be used in further model development.

Model Y-19, the final driver yielding model, uses nine explanatory variables and has R^2 value 0.3582 (max-rescaled R^2 is 0.4797). The variables chosen for this model were all significant at $p<0.05$. Increased speed (SPD) was seen to reduce the likelihood of yielding, as did increased required deceleration rate (DECEL). Presence of adjacent yields (ADJ), low speed platoons (LSPLT), presence of multiple pedestrians (MUP), and female pedestrians (FEMALE) were seen to increase the likelihood of yielding. Drivers are more likely to yield to pedestrians on-campus (CAMPUS). The coefficient estimates for Florida and North Carolina show that drivers are more likely to yield in Florida and less likely to yield in North Carolina than drivers in Alabama. The model is shown below in Equation 16.

$$\begin{aligned}&\mathrm{logit}[\mathrm{P}(\mathrm{Y}=1)] = 0.8344 - 0.0894\mathrm{SPD} + 0.9448\mathrm{ADJ} + \\ &0.9833\mathrm{LSPLT} + 0.7348\mathrm{MUP} - 0.1369\mathrm{DECEL} + 0.8247\mathrm{FEMALE} + \\ &1.0476\mathrm{CAMPUS} + 1.4245\mathrm{FLORIDA} - 1.2034\mathrm{NCAROLINA}\end{aligned}$$

Equation 16

Model Y-20, the universal driver yielding model, uses seven explanatory variables (same variables as Y-19) and has R^2 value 0.2680 with max-rescaled R^2 value 0.3589. The state variables were removed from the model to provide a model that can apply outside of the three states where data was collected. All variables were significant at $p<0.25$. The only variable that was not significant at $p<0.05$ were DECEL. This variable was kept so that the two models would be consistent. The model is shown below in Equation 17.

$$\begin{aligned}&\mathrm{logit}[\mathrm{P}(\mathrm{Y}=1)] = 0.1765 - 0.0758\mathrm{SPD} + 1.1365\mathrm{ADJ} + \\ &0.9066\mathrm{LSPLT} + 0.7171\mathrm{MUP} - 0.0328\mathrm{DECEL} + 0.3765\mathrm{FEMALE} + \\ &1.532\mathrm{CAMPUS}\end{aligned}$$

Equation 17

In addition to the yield model prediction, nested logit models were used to predict the likelihood of hard yielding given that a yield occurred. These models represent two-stage binary logit models with the first stage consistent with the initial binary logit model. The second level predicts likelihood that driver performs hard yield versus baseline soft yield, given that the first level of the nested logit predicts a yield. The overall probability of a driver hard yielding can be calculated by multiplying the two probability functions. Soft yielding probability correspondingly is likelihood of a yield multiplied by one minus probability of hard yield.

Hard yielding models are created from the subset of observations where YIELD = 1. HY-5, the final hard yield model, uses six explanatory variables and has R^2 value 0.2507 (max-rescaled R^2 is 0.3449). All variables used in this model were significant at $p<0.05$. Longer distances from the crosswalk (ADJDIST) and being in the lane closest to the pedestrian (NEAR) decreases the chance of a driver deciding to hard yield. This is likely because drivers who are further away from the pedestrian have time to react with a soft yield. Drivers are more likely to hard yield at Florida sites than North Carolina sites. Adjacent yields (ADJ), presence of multiple pedestrians (MUP), and higher required deceleration rates (DECEL) increase the chance that a driver will hard yield. If the deceleration rate required to yield is higher, then it is reasonable that the chance of hard yielding is increased. The model is shown below in Equation 18

$$\begin{aligned}&\mathrm{logit}[\mathrm{P}\,(\mathrm{HY}=1\,) = 0.3699 - 0.0134\mathrm{ADJDIST} - 0.7114\mathrm{NEAR} + 0.8860\mathrm{ADJ} + \\ &1.3643\mathrm{MUP} + 0.2865\mathrm{DECEL} - 0.9443\mathrm{NCAROLINA}\end{aligned}$$

Equation 18

Comparison between Candidate Yield Models

It is seen that the R^2 and max-rescaled R^2 values drop when going from the final driver yielding model with state variables (Y-19) to the universal model (Y-20). R^2 drops from 0.3582 to 0.2680, while max-rescaled R^2 value drops from 0.4797 to 0.3589. The ADJ, LSPLT, and MUP variables are more significant in Y-20. This is likely due to the removal of the other site-specific variables of FLORIDA and NCAROLINA. A few variables drop in significance, but they remain significant at $p<0.25$. The driver hard yield decision model (HY-5) has an R^2 value of 0.2507 and a max-rescaled R^2 value of 0.3449. These values are similar to those for the universal driver yielding model (Y-20). All models are fairly comparable based on these model fit statistics. Many of the variables that were shown to be significant in the driver yielding decision were also shown to be significant in the decision to hard yield or soft yield. The R^2 values may seem low, but practical significance for model selection is more important than the overall model fit. The final model results are summarized in Table 11.

Table 11: Results of Logistic Regression for Driver Yielding

	Y-19		Y-20		HY-5	
	Parameter	Pr > Chi Sq	Parameter	Pr > Chi Sq	Parameter	Pr > Chi Sq
Intercept	0.8344	0.0501	0.1765	0.623	0.3699	0.3957
SPD	-0.0894	< 0.0001	-0.0758	< 0.0001	---	---
ADJDIST	---	---	---	---	-0.0134	< 0.0001
NEAR	---	---	---	---	-0.7114	0.0032
ADJ	0.9448	0.0041	1.1365	0.0001	0.8860	0.0064
PLT	---	---	---	---	---	---
LSPLT	0.9833	0.0036	0.9066	0.0033	---	---
HGV	---	---	---	---	---	---
MUP	0.7348	0.0013	0.7171	0.0006	1.3643	< 0.0001
MED	---	---	---	---	---	---
CTRL	---	---	---	---	---	---
TTC	---	---	---	---	---	---
DECEL	-0.1369	0.0005	-0.0328	0.2371	0.2865	0.0005
AGE	---	---	---	---	---	---
BUSINESS	---	---	---	---	---	---
FEMALE	0.8247	< 0.0001	0.3765	0.0206	---	---
CAMPUS	1.0476	< 0.0001	1.532	< 0.0001	---	---
FLORIDA	1.4245	< 0.0001	N/A	N/A	---	---
NCAROLINA	-1.2034	< 0.0001	N/A	N/A	-0.9443	0.0054
R^2	0.3582		0.2680		0.2507	
Max Rescaled R^2	0.4797		0.3589		0.3449	

An increase of 1 mph in speed reduces the odds of yielding 0.91 times for the first model and 0.93 times for the universal model. The estimated odds of a yield occurring if there is an adjacent yield is 2.57 times the odds of an event without an adjacent yield for Y-19 and 3.12 for Y-20. The effect on the decision to hard yield is similar, with an increase of 2.42 times. The odds of yielding in relation to each variable was similar for Y-19 and Y-20 for most of the variables. The largest difference was seen for CAMPUS, with an increase in odds of 2.85 for Y-19 and 4.63 for Y-20. The presence of multiple pedestrians has a much greater effect on the decision to hard yield (3.91 times) than it does on the overall decision to yield (2.09 for Y-19 and 2.05 for Y-20). Odds in yielding are reduced 0.87 times for Y-19 and 0.97 times for Y-20 for each additional 1 ft/sec^2 in necessary deceleration rate, while the odds of hard yielding are increased 1.33 times in HY-5. The odds of a driver yielding are increased 2.85 times for Y-19 and 4.63 times for Y-20. Figure 8 below shows that the chance of a driver yielding in Y-20, the universal model, is greatest when there is an adjacent yield and multiple pedestrians are present (all other variables set to zero) and that chance of yielding decreases as speed increases. The same is true for Y-19.

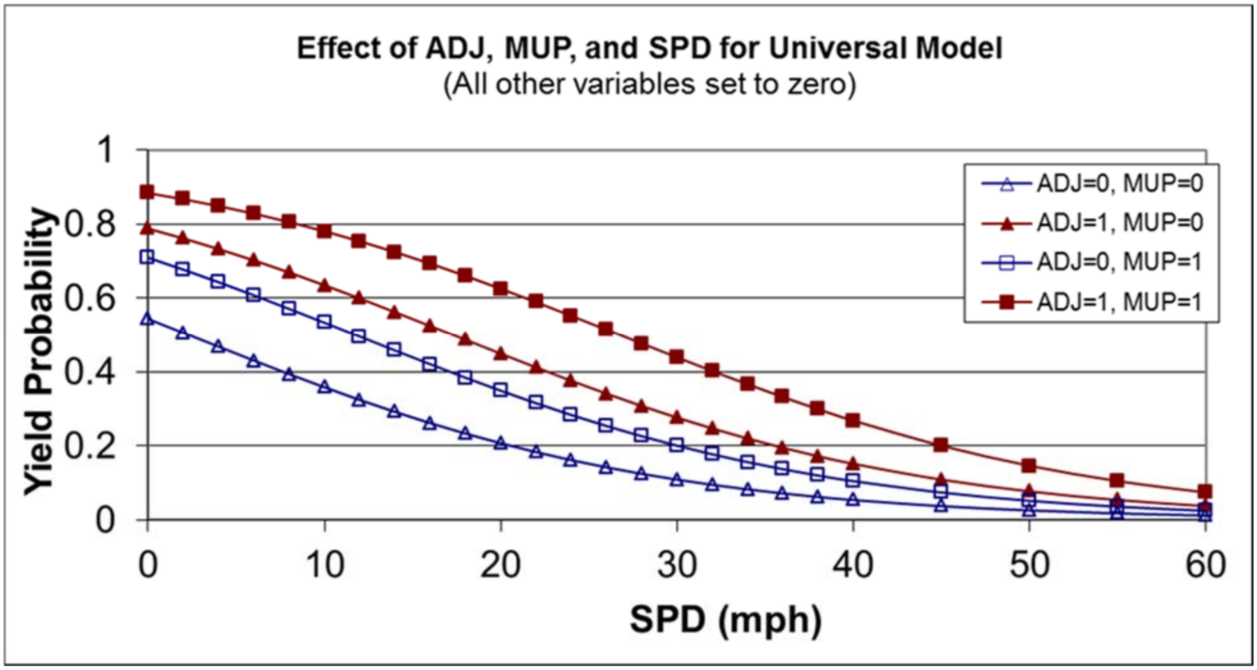


Figure 8: Effect of ADJ, MUP, and SPD on Yield Probability

Validation

The validation dataset was composed from one site per state that was not used in model development. After removing observations where the driver was previously stopped or the yield was triggered by the pedestrian, 127 observations remained in the validation dataset. Yield (and hard yield) probabilities were calculated for each observation from the models and plotted against the observed driver decision. For the hard-yielding model validation, non-yields were given the value of 0. The predicted probability of a hard yield is the probability of yielding multiplied by the probability of hard yielding. The plots for both yielding models (Y-19 and Y-20) are provided in Figure 9 below. Plots for the nested hard yielding model are shown in Figure 10 below using both driver yielding models. Observations are color-coded by state, with Florida represented by green, Alabama by red, and North Carolina by blue.

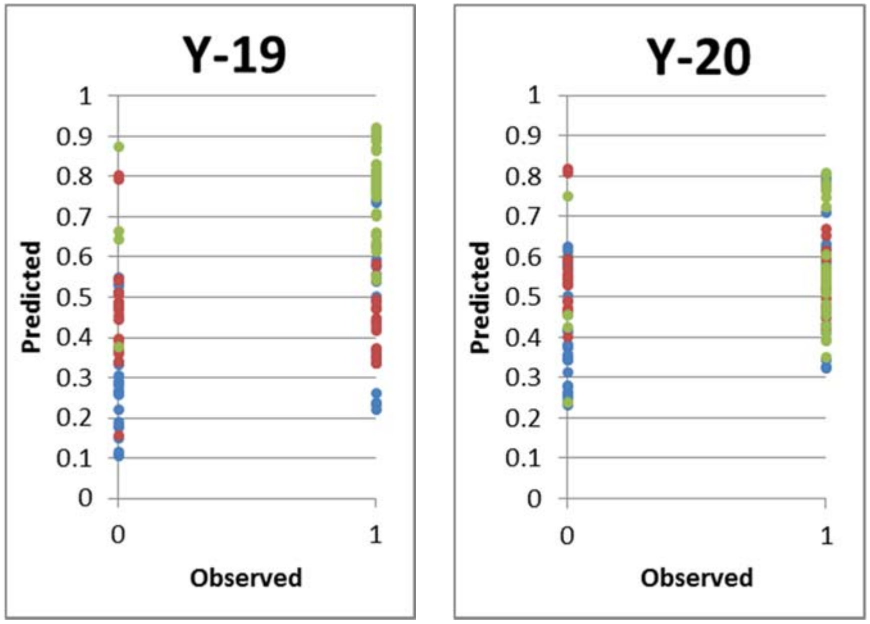


Figure 9: Validation Plots for Yield Models

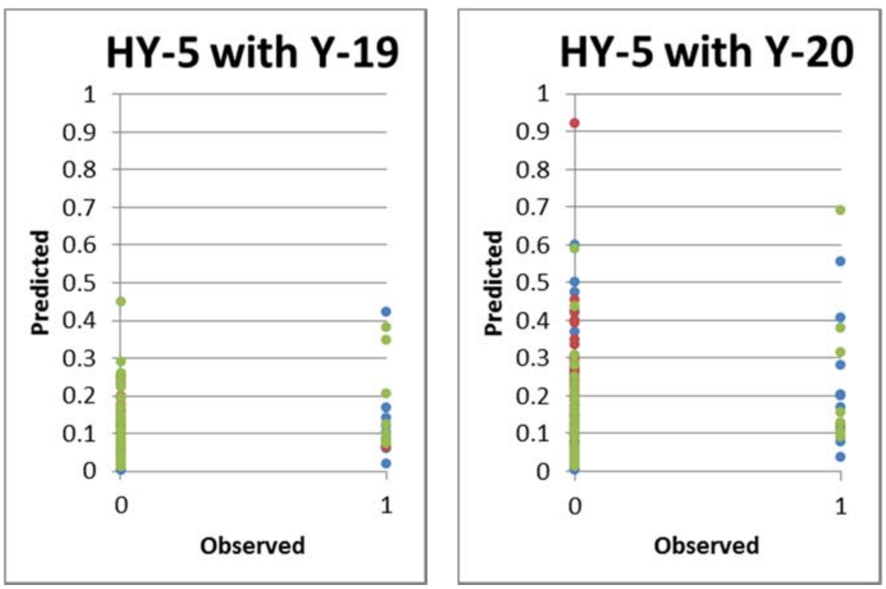


Figure 10: Hard Yielding Validation Plots

Y-19 showed 53.3% of drivers yielding and Y-20 showed 52.0% of drivers yielding. The data showed that 59.8% of drivers decided to yield. The results for HY-5 with Y-19 showed 10.9% of drivers deciding to hard yield, while HY-5 with Y-20 showed 19.1% and the data showed 14.2% of drivers deciding to hard yield. Data showed 20.9% of observations at the North Carolina validation site to be hard yields, while only 2.4% of the observations at the Alabama site were hard yields. The model was not as accurate at predicting hard yield probabilities for individual states, but this may have been due to the large variation in the datasets. Simulation will also be used to validate these models.

Gap Models

Descriptive Statistics

A total of 1,246 pedestrian crossings in gaps were recorded at 27 study sites spanning over the three participating states (Alabama, Florida, and North Carolina). Of these events, 339 gaps were accepted by the crossing pedestrian and the remaining 907 gaps were rejected. An individual site from each state (UF2, AL7, and NC2) was set aside for validation of analytical models obtained from rest of observed data.

The study builds on the empirical behavior modeling framework for mixed-priority pedestrian vehicle interaction developed by Schroeder et al. (2011). Vehicle arrival events were distinguished into gaps and lags. As mentioned earlier, gap describes the time elapsed between consecutive vehicle events with reference to start / end of crosswalk. Similarly, a lag corresponds to the time elapsed between pedestrian's arrival at the crosswalk and next vehicle event.

Owing to the varying vehicular and pedestrian activity across different study sites, the number of gap events recorded for each one of the individual sites was different in number. Table 12 below indicates the number of gap and lag events, and the accepted and rejected gap / lag by each site.

Table 12: Summary of Gap Data

SITE	Lags and Gaps by Individual Site			Accepted and Rejected Gaps by Individual Site		Accepted and Rejected Lags by Individual Site	
	Gap	Lag	Total	Rejected	Accepted	Rejected	Accepted
AL1	1	7	8	0	1	0	7
AL2	104	79	183	74	30	60	19
AL3	51	11	62	41	10	10	1
AL4	172	44	216	151	21	44	0
AL5	1	21	22	1	0	17	4
AL6	8	22	30	4	4	16	6
AL8	19	26	45	14	5	21	5
AL9	82	66	148	81	1	59	7
FL1	7	29	36	5	0	12	5
FL3	6	9	15	64	19	51	9
FL4	3	2	5	20	12	51	1
FL5	2	1	3	2	3	9	31
FL6	9	7	16	0	1	5	9
FL7	2	8	10	1	1	3	13
FL8	9	9	18	1	6	5	24
FL9	13	16	29	4	12	14	5
FL10	16	18	34	15	10	7	7
NC1	5	17	22	4	2	3	6
NC3	25	14	39	1	2	2	0
NC4	83	60	143	1	1	1	0
NC5	32	52	84	2	7	7	0

NC6	5	40	45	1	1	1	7
NC7	1	14	15	6	3	3	6
NC8	2	16	18	0	13	13	3
Total	658	588	1246	493	165	414	175

The study sites provide heterogeneity in assessing pedestrian gap outcomes in relation to different land use (i.e., off-campus versus on-campus), facility configuration (number of lanes, presence of refuge island), and pedestrian characteristics. A few sites witnessed sparse vehicular and/or pedestrian activity. To augment observational data in such sites staged crossings were undertaken. Of the data used in the analysis, 31.6% are based on non-controlled (natural) pedestrian crossings and the remaining 68.4% are from staged crossings.

Close observation of the database revealed that the data comprised of over-representation of gap events for greater crosswalk widths. Hence, to overcome this limitation, the observed gap length (OBS) was normalized for crosswalk width (CR_WIDTH) and the normalized gap length variable (N_GL) was introduced in the gap dataset.

For ease of integration of behavioral gap acceptance models with the simulation framework in later stages of the study, a dataset comprising of single lane crossing events was separated. This distinction was based on the existing simulation framework used in the study. Furthermore, non-controlled and staged crossings were separated to achieve greater inter-pedestrian independence of events.

Table 13 provides details on the sample size obtained for detailed analysis datasets.

Table 13: Gap Data Stratification

Dataset	**Sample Size**
Single Lane (Non-Controlled)	153
Single Lane (Staged)	213
All Sites (Non-Controlled)	394
All Sites (Staged)	852

As shown in Table 14, normalized gap lengths for above different segments were grouped into discrete percentile bins to compare acceptance and rejection of gaps.

Table 14: Gap Acceptance Distribution for Study Datasets

Percentile Bin	**Single Lane (Non-Controlled)**		**Single Lane (Staged)**		**All Sites (Non-Controlled)**		**All Sites (Staged)**	
	Rejected	**Accepted**	**Rejected**	**Accepted**	**Rejected**	**Accepted**	**Rejected**	**Accepted**
0	0	0	0	0	0	0	0	0
10	16	0	20	0	33	0	77	0
25	21	1	74	0	61	0	124	2
50	23	16	77	5	82	18	193	18
75	2	36	5	16	30	70	178	34
90	2	21	1	3	5	55	101	24
95	1	7	1	5	2	18	16	27
100	1	6	0	2	2	17	3	39

From Table 14 it can be observed that for normalized gap lengths below the 50^{th} percentile, the number of gaps rejected is greater in number. Conversely, for values greater than the 50^{th} percentile, gaps accepted are greater in number barring staged events for all sites combined.

Gap Acceptance Modeling

The outcome of a gap acceptance event is binary (Go vs No-Go) in nature. The observed data were used to construct probabilistic models, which can aid in predicting gap selection. The logit based binary choice model formulation was attempted first as described in Chapter 3. However, due to the nature of available data, non-convergence of likelihood maximization was observed. Consequently incorrect maximum likelihood estimates were obtained. For overcoming this limitation, the probit model formulation was used instead of the logit model with great success.

The PROC LOGISTIC function in SAS was used to develop model estimates for gap acceptance behavior (SAS, 2007). Forward selection and backward elimination strategies were used to assess different models for the given data. For same set of data, forward selection and backward elimination strategies yielded almost similar results.

Comparison

Probit models were developed for all four datasets separately. For datasets comparatively larger in size (All sites combined) variables like PLT (lead vehicle not being in platoon), LAG (Pedestrian faces a gap or a lag), NEAR (Lead vehicle traveling in a far lane) were found relevant. Vehicle dynamics (Speed, Distance from Crosswalk, and Time to Connect) were not included in the model, due to the fact that these variables were captured mainly for the first vehicle in the platoon. Similarly, variables such as HGV (Heavy Vehicle), MUP (Multiple Pedestrians), and DIST (Distracted) were under-represented in the field data and did not make it to final set of models. The final model results are summarized in Table 15.

Table 15: Results of Logistic Regression for Pedestrian Gap Acceptance

	Single Lane (Non-Controlled)		Single Lane (Staged)		All Sites (Non-Controlled)		All Sites (Staged)	
	Parameter	**Pr>Chi**	**Parameter**	**Pr>Chi**	**Parameter**	**Pr>Chi**	**Parameter**	**Pr>Chi**
Intercept	-1.8904	<.0001	-3.3283	<.0001	-2.0424	<.0001	-4.0525	<.0001
N_GL	5.0483	<.0001	5.1946	<.0001	5.3112	<.0001	6.2670	<.0001
PLT	-	-	-	-	0.5543	0.0058	0.5371	<.0001
LAG (0)	-0.7688	0.0081	-	-	-	-	0.2865	0.0337
GENDER (0)	-	-	-	-	-	-	1.5585	<.0001
NEAR (0)	-	-	1.2189	.0143	-	-	-	-
R^2	0.5177		0.3814		0.4484		0.3303	
Max Rescaled R^2	0.6946		0.6526		0.5595		0.5044	

Table 15 presents a set of models along with their explanatory power (expressed by R^2, Max Rescaled R^2). All models show acceptable results, with the single lane models yielding higher R^2, Max Rescaled R^2). It can be observed that the predicted probability of accepting a gap (GO =1) increases with increasing normalized gap length (N_GL). This is consistent across all models fitted. For all the models listed above, normalized gap length estimate seems consistent in value. For example, when other variables being constant, every unit increase in normalized gap length leads to 5.04 times increase in Z score for (GO =1).

For models comprising of all sites (Non-Controlled and Staged), the variable representing vehicle not being in platoon (PLT = 0) affects the acceptance of gap to a lesser extent. The estimate appears consistent to both models.

For single lane (Non-Controlled) events, the model shows that a pedestrian, when faced to accept a gap as opposed to lag, is less likely to accept the gap, considering that other things being constant.

The absolute fit of these different models can be compared using the max-rescaled R^2 criterion. This statistic explains the overall variability in the data by the model and it serves as counterpart similar to R^2 in linear regression. Generally, the greater the max rescaled R^2, the better the explanatory power of the model in terms of explaining overall variability. The single lane model drawn from Non-Controlled dataset has a max rescaled R^2 of 0.69 and shows the best model fit to describe gap-acceptance behavior. Besides, it best meets the needs of the simulation modeling platform, thus it is recommended for adoption in the next steps of the modeling process.

The predicted likelihood of a pedestrian accepting a gap (GO =1) can be plotted graphically by applying different set of models. For illustration, the probability of accepting a gap for crossing a single lane facility based on non-controlled dataset in equation form is:

$$\Phi^{-1}(\Pr(Go = 1)) = -1.8904 + 5.0483N\_GL - 0.7688LAG(0)$$ **Equation 19**

Equation 19 given above highlights that for every unit increase in normalized gap length, the likelihood of accepting gap given by Z-score increases almost five times. If a vehicle is closely followed by another vehicle or is a part of platoon, such a gap reduces the likelihood of gap acceptance as opposed to accepting a lag. The negative sign for LAG (0) estimate is indicative of this fact.

As shown in the Figure 11, using single lane model for a given normalized gap length shows greater likelihood of accepting a lag than gap. For example, for N_GL of 0.5, the probability of accepting a gap is 44% while the probability of accepting a lag is nearly 74%.

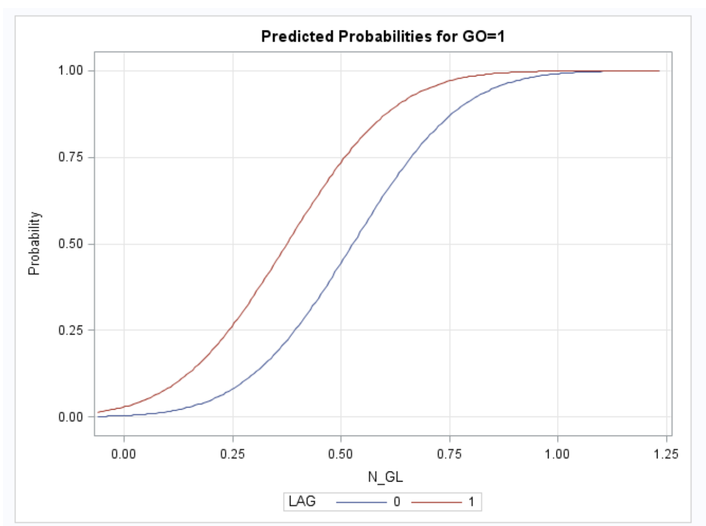


Figure 11: Model Probability for Gap Acceptance (Model: Single Lane, Non-Controlled)

Validation

An individual study site from each state was set aside for validating the statistical gap models. The validation data set comprised of 39 gap events (3 from UF2 and 36 from NC2). AL7 was the validation site from Alabama, but all crossing events were yield events. Final set of four probit models discussed in preceding sections were used to predict gap acceptance probability. In validating predictive ability of model certain simple conditions were tested: a) Strong Prediction b) Weak Prediction. These conditions are summarized in the Table 16 below.

Table 16: Validation Criteria for Predicted Outcomes

If (GO = 1) and Predicted Probability >= 0.5 , STRONG PREDICTION Denoted By 1
if (GO = 0) and Predicted Probability <=0.5, STRONG PREDICTION Denoted By 1
if (GO = 1) and Predicted Probability <=0.5, WEAK PREDICTION Denoted By 0
if (GO = 0) and Predicted Probability >=0.5, WEAK PREDICTION Denoted By 0

A gap acceptance predicted strongly would result in probabilities in higher order of 0.5 (>0.5). Similarly, rejected gaps would be predicted with probabilities lesser of 0.5 (<0.5). The predicted probabilities from the probit model were compared to observed gap outcomes from the validation set events. It was found that gap acceptance models I through IV exhibited accuracy of 82%, 77%, 77% and 93%, respectively. Figure 12 through Figure 15 display the predicted probabilities from gap acceptance models I through IV using the study validation data set.

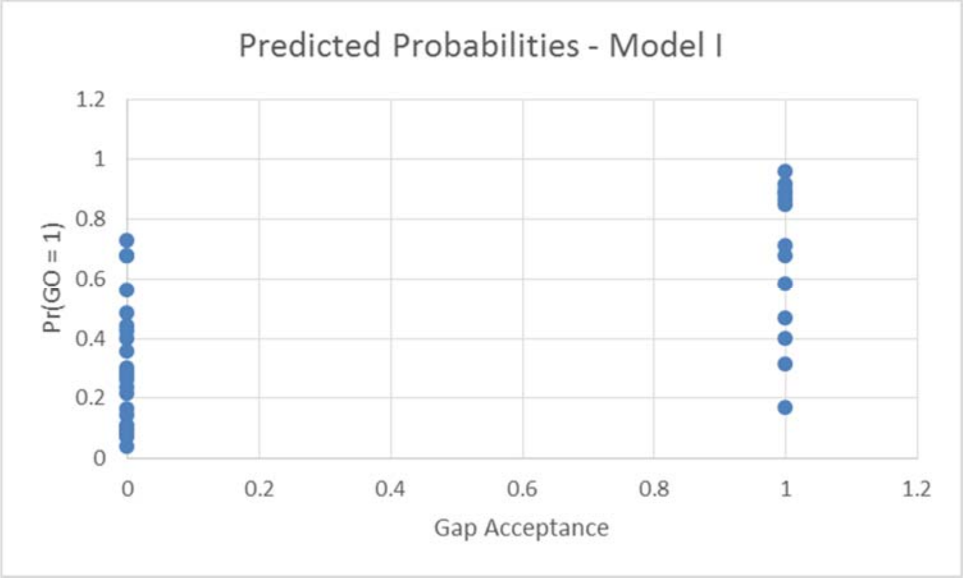


Figure 12: Predicted Probabilities; Validation Data Set; Gap Acceptance Model I

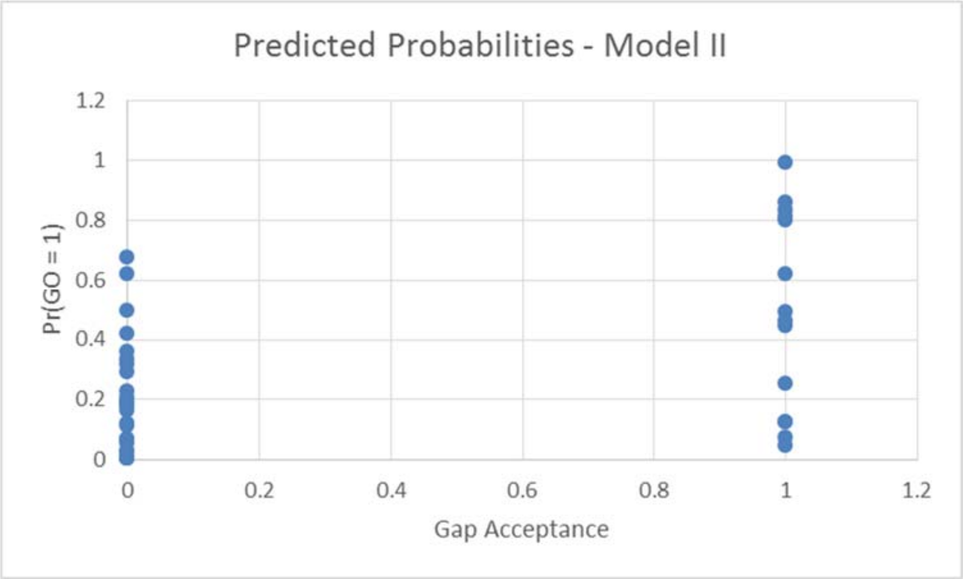


Figure 13: Predicted Probabilities; Validation Data Set; Gap Acceptance Model II

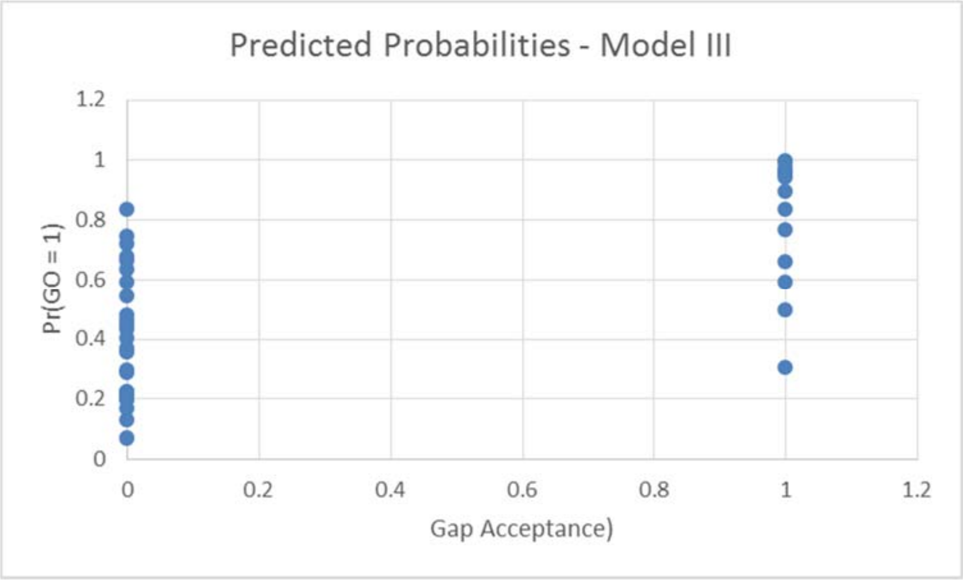


Figure 14: Predicted Probabilities; Validation Data Set; Gap Acceptance Model III

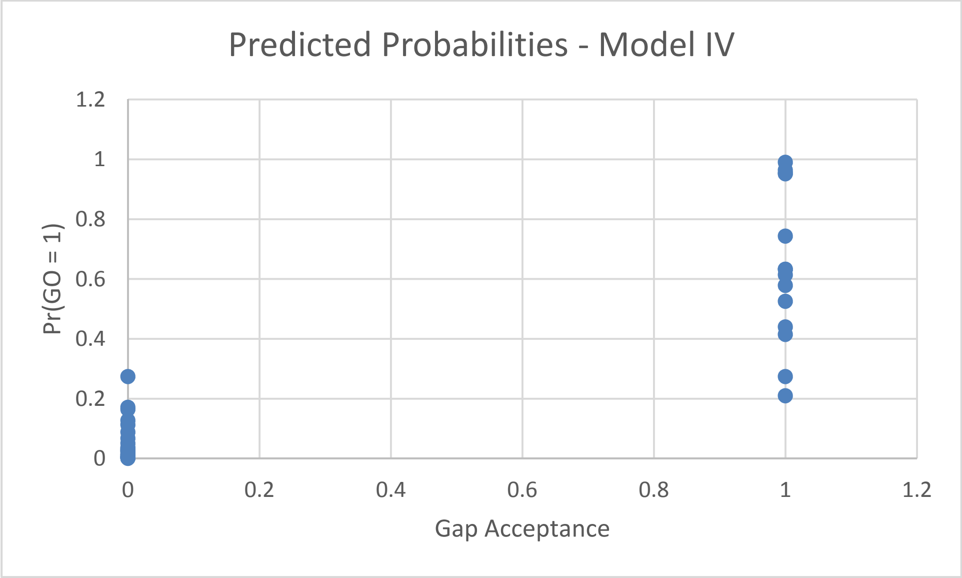


Figure 15: Predicted Probabilities; Validation Data Set; Gap Acceptance Model IV

A closer look at prediction accuracy charts below (Figure 16 through Figure 19), highlight the clustering of weak predictions (indicated by "w" on the vertical scale). For Model I, weak predictions are spaced approximately at equal length for probabilities 0.18 to 0.75. Similarly small cluster of three weak predictions can be observed for probability range 0-0.15 and 0.4-0.55 for Model II. For Model III, weak predictions lie in the range of 0.5 to 0.8. Model IV has the least number of weak predictions, all of them lying below probability mark of 0.5.

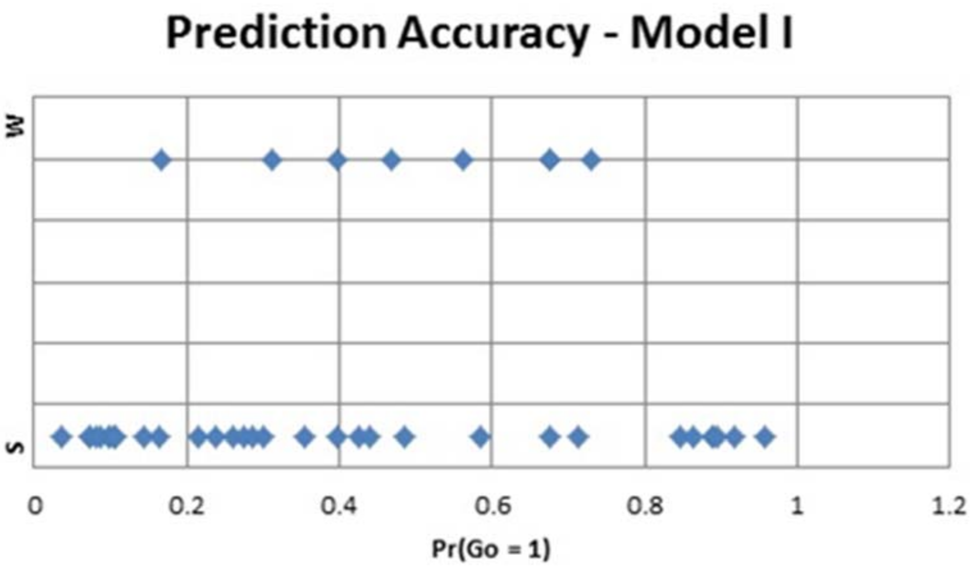


Figure 16: Validation chart for Gap Acceptance Model I

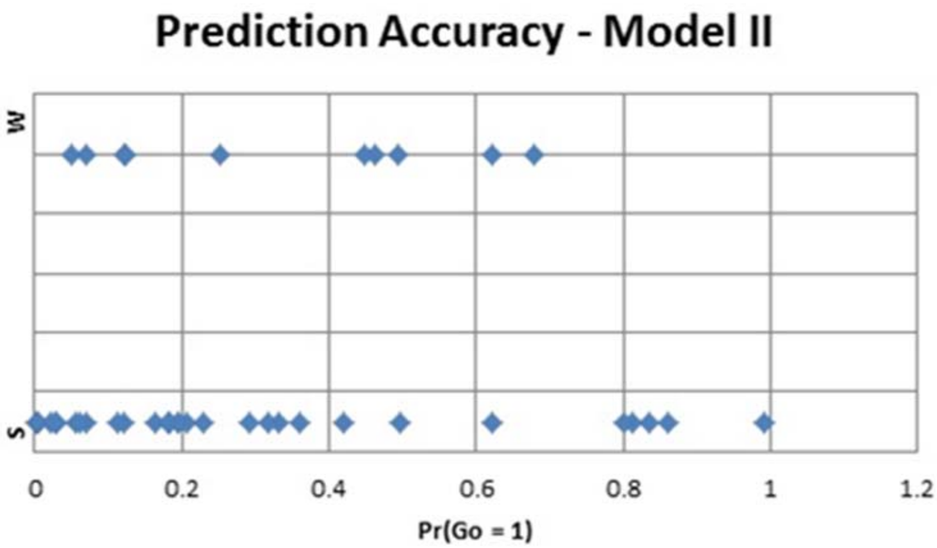


Figure 17: Validation chart for Gap Acceptance Model II

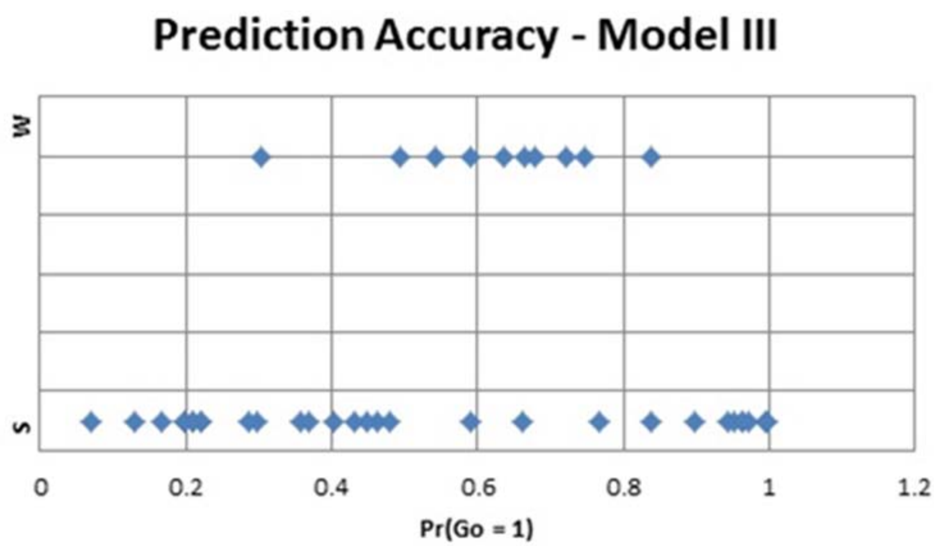


Figure 18: Validation chart for Gap Acceptance Model III

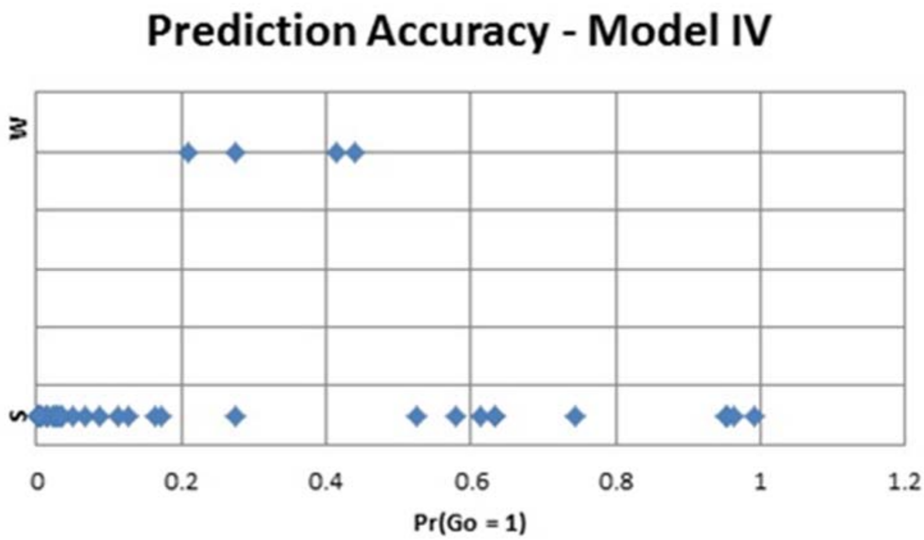


Figure 19: Validation Chart for Gap Acceptance Model IV

Other Submodels used in Simulation

Additional measures were obtained and compiled from the pedestrian observation study and the instrumented vehicle study, to provide quantitative information related to vehicle movement and driver choices. Several sub-models were developed for use in simulating vehicle and pedestrian operations at crosswalks.

Decision Distance Model

For the purposes of simulation, as each vehicle approaches a crosswalk, they are assumed to have a unique decision point. The research team developed a model using data from the observational study, where this decision point is the distance between the vehicle and the crosswalk at the time when the observational data were collected; in that data collection this point was assumed to be the location where the driver begins to react to the presence of the pedestrians. The driver's decision point model is developed to explain at which location to the crosswalks the driver makes a decision to Yield/No-Yield to waiting pedestrians at the curb.

Figure 20 summarizes the decision distances grouped by the vehicle's Free-Flow-Speed (FFS).As shown, the driver's decision point is a function of their FFS, and also its slope is different for low FFS (less than 30 mph) vs high FFS (over 30 mph) For low FFS the vehicle distance remains relatively flat, while for higher FFS it increases steadily. Thus these two conditions are discussed separately below.

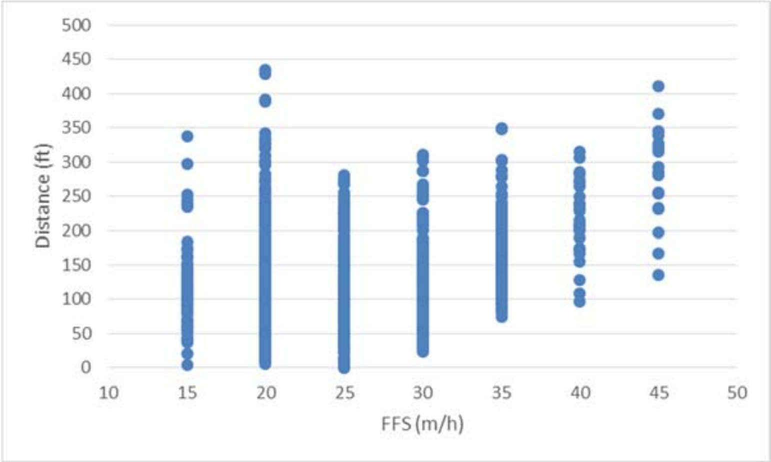


Figure 20: Driver Decision Distance to Crosswalks grouped by FFS

Low-Speed Group

Figure 21 shows the decision distance to crosswalks for the low FFS group. A statistical analysis (one-way Anova with single factor) was conducted to analyze the correlations between FFS groups and within groups (<30 mph) as shown in Table 1. Since F < F critical, the group means of these three groups are not significantly different, thus they can be combined in one category (low FFS).

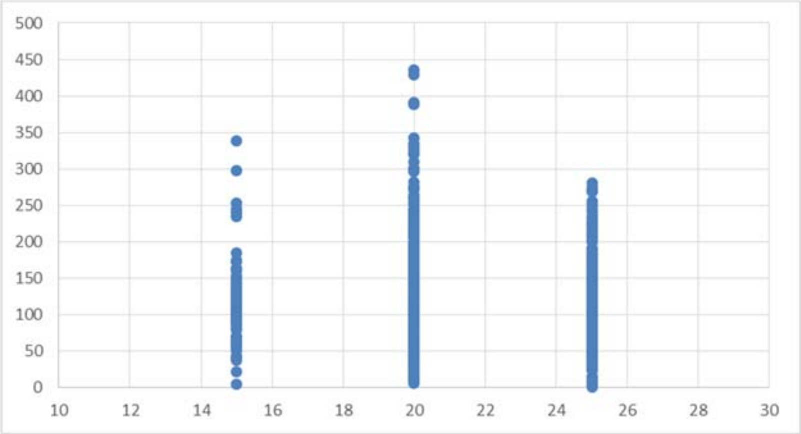


Figure 21: Driver Decision Distance to Crosswalks (FFS < 30 mph)

Table 17: Anova of Driver Decision Distance (FFS < 30 mph)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	16169.16	2	8084.58	1.869152	0.154962	3.007556
Within Groups	3291527	761	4325.266			
Total	3307696	763				

Based on the findings from Table 17, a combinational dataset of the low-speed group was built and the probability density function is shown in Figure 22. The decision points result in a good fit using the Burr

Distribution (Log Likelihood of -4197.18.) The cumulative distribution function of Burr Distribution is:

$$\boldsymbol{F(x) = 1 - \left(1 + \left(\frac{x}{a}\right)^{c}\right)^{-k}}$$ **Equation 20**

Where,

a is the scale parameter, c and k are the shape parameters. These are estimated to be 214.533, 3.643, and 2.241 in this study.

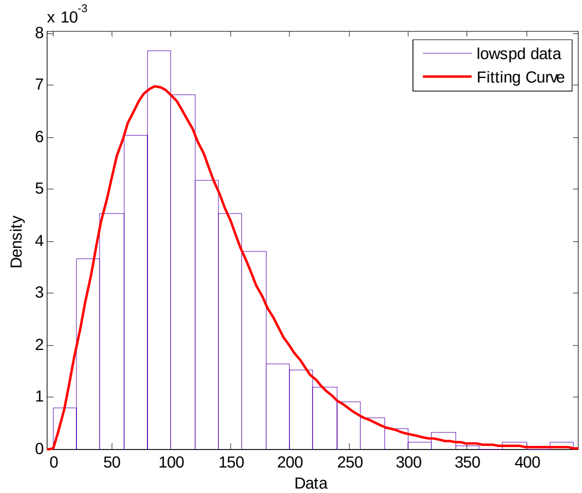


Figure 22: PDF of Driver Decision Points (FFS < 30 mph)

Thus, the model predicting the driver decision point for low FFS (FFS < 30 mph) is as follows:

$$\boldsymbol{Y = 214.533\left((1 - prob)^{\frac{1}{-3.643}} - 1\right)^{\frac{1}{2.241}}}$$ **Equation 21**

Where,

$prob$ is a random number from 0 to 1 representing the cumulative probability of the decision point.

High-Speed Group

Another one-way Anova with single factor analysis was conducted to analyze the correlations between FFS groups and within groups for the high FFS group (>=30 mph). As shown in Table 18, the group means of these three groups are not all the same. A regression analysis is conducted and shown in Figure 23 with the dotted line representing the regression model. The R^2 is 0.2883, which is low but was considered acceptable. Thus, the mean value of the decision point for the high FFS group is determined by the following equation:

$$\boldsymbol{y = 9.4297x - 159.17}$$ **Equation 22**

Table 18: Anova of Driver Decision Distance (FFS >= 30 mph)

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	490946.5	3	163648.8	46.94657	2.07E-25	2.631175
Within Groups	1185190	340	3485.853			
Total	1676136	343				

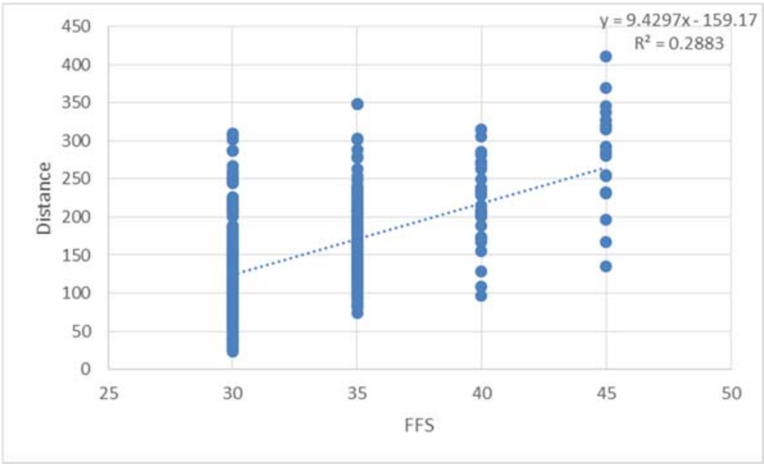


Figure 23: Driver Decision Distance to Crosswalks (FFS >= 30 mph)

Next, it is necessary to check the distribution of each of the high FFS groups to be able to better replicate the variability within each group. Figure 24 provides the cumulative density function of each FFS group, and all of them fit well the Logistic Distribution:

$$F(x) = \frac{1}{1+e^{-\frac{x-u}{s}}} \quad \textbf{Equation 23}$$

Where,

s is scale parameter and u is the location parameters. s is estimated as 34.363, and u is estimated by the regression line in Figure 23.

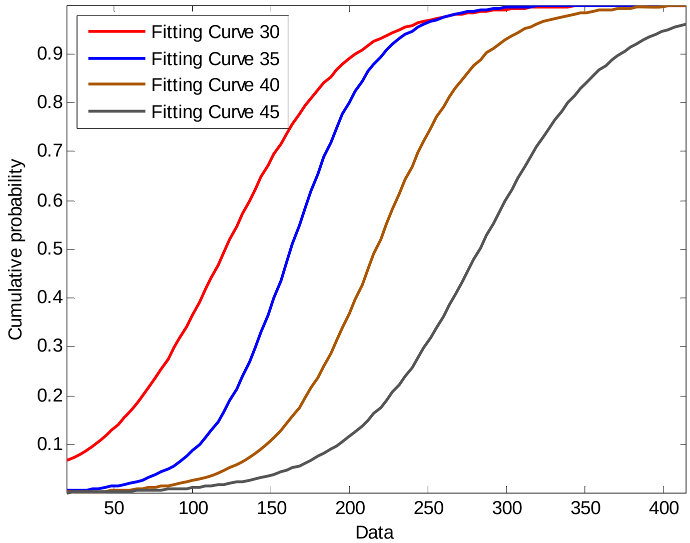


Figure 24: CDF of Driver Decision Points (FFS >= 30 mph)

Thus, the model of driver decision point (FFS >= 30 mph) for each FFS group is an inverse cumulative distribution function (quantile function) of the Logistic Distribution with $s = 34.363$ and u a linear function of FFS:

$$\boldsymbol{Y = 34.363\left(-\ln\left(\frac{1}{prob} - 1\right)\right) + 9.43 * FFS - 159.17} \qquad \textbf{Equation 24}$$

Where,

$prob$ is a random number from 0 to 1 representing the cumulative probability of decision point.

Soft-Yield Dynamics Model

Drivers have the choice to Yield or No-Yield when they approach a crosswalk with pedestrian(s) waiting at the curb. After they decide to yield, they may choose to do Hard-Yield if they are too close to the crosswalk, or choose to perform a Soft-Yield to avoid a complete stop. Regarding Hard-Yields, vehicle behavior and deceleration rates are similar to approaching a stop sign or a red indication at a signalized intersection. However, regarding Soft Yields, there are currently no models available that provide vehicle trajectories. In this study, we assume that Hard-Yields are modeled identically to decelerating at a stop sign, and only a Soft-Yield model is developed based on data from the instrumented vehicle study.

Using GPS data collected from the instrumented vehicle study, vehicle speed profiles were obtained for each vehicle-pedestrian interaction observed. In the simulation implementation, the total travel time during the vehicle-pedestrian interaction process is assumed to be the time for the pedestrian to cross the entire crosswalk plus a safety buffer corresponding to a pedestrian's body width of 3.5 ft. at the pedestrian's crossing speed.

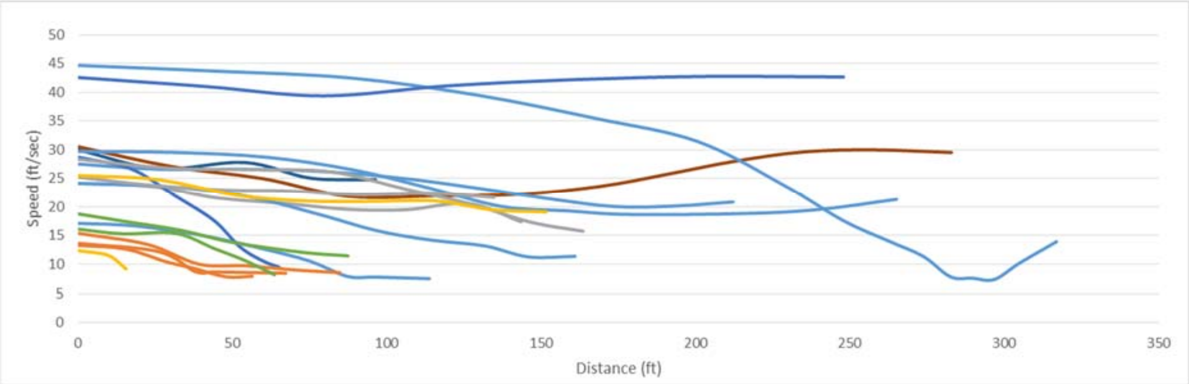
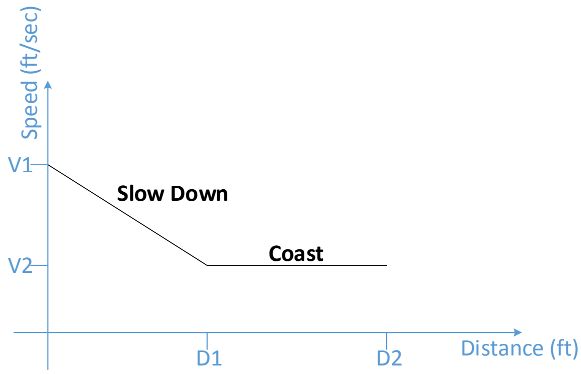
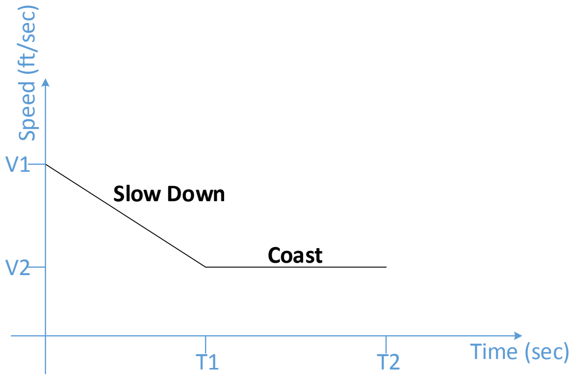


Figure 25: Distance-Speed Profile for Vehicles Approaching Crosswalk with Pedestrians (Soft-Yield)

Figure 26 provides the generalized vehicle distance-speed and time-speed models, developed using instrumented vehicle data. In Figure 26, V1 is the vehicle speed at the point of decision making, and V2 is the constant coast speed. D1 is the location where vehicle turns to coast to crosswalk, and D2 is the total distance vehicle travels (between the point of decision making and the crosswalk). The difference of T1 and T2 is the vehicle coast time.



(a) Distance-Speed Profile



(b) Time-Speed Profile

Figure 26: Generalized Vehicle Distance-Speed and Time-Speed Profiles

Regression analysis was conducted to generate the model of Soft-Yield deceleration rate using data from the in-vehicle study. The results indicate that this vehicle deceleration rate is a function of speed and distance at the decision point ($R^2 = 0.3388$):

$$\boldsymbol{Decel = 0.050815 + (-0.13986) * Speed + (0.010115) * Distance}$$ **Equation 25**

According to the generalized distance-speed/time-speed profiles, the deceleration time (T1), coast time (T2) and coast speed (V2) are as follows:

$$\boldsymbol{PedTime = \frac{PedCrossLength + PedBodyWidth}{PedSpeed}}$$ **Equation 26**

$$\boldsymbol{VehTime = PedTime}$$ **Equation 27**

$$\boldsymbol{DecelTime = VehTime - \sqrt{(VehTime)^2 - \frac{2*(Distance - Speed*VehTime)}{Decel}}}$$ **Equation 28**

$$\boldsymbol{CoastTime = \sqrt{(VehTime)^2 - \frac{2*(Distance - Speed*VehTime)}{Decel}}}$$ **Equation 29**

$$\boldsymbol{CoastSpeed = Speed + Decel * DecelTime}$$ **Equation 30**

The following chapter describes in more detail how these models were integrated into a micro simulator for illustration and testing of the new algorithms.

CHAPTER 6: SIMULATION IMPLEMENTATION

Development of Simulation

The pedestrian-vehicle interaction simulation was implemented in a new version microsimulation tool that is currently under development at the University of Florida under the guidance of Dr. Scott Washburn. This project adds pedestrian movement and interaction with vehicles to the existing model that is programmed in the C# programming language to demonstrate an application of the statistical interaction models in microsimulation. The simulation developed for this project is a proof of concept, as the models of interaction can be applied in any microsimulation that includes vehicles and pedestrians where it is possible to modify their behavior, such as VAP in VISSIM.

Simplifying Assumptions

Several simplifying assumptions were made in order to create a realistic simulation under the constraints of a research application rather than a full commercial simulator. The geometry simulated is limited to three configurations that are similar to the majority of data collection sites in the observational study. The first configuration is a single one-way lane with a midblock crosswalk in the middle of a 1,600 ft segment. The second configuration is a two-lane segment with one lane in both directions and no median between lanes, also with a midblock crosswalk in the middle of a 1,600 ft segment. The third configuration is identical to the second, but includes an 8-foot median between lanes, allowing for a two-stage crossing.

In the simulation, vehicle behavior during potential vehicle-pedestrian interactions is controlled by a crosswalk control point. In other words, vehicles do not directly "observe" the pedestrians, but they receive instructions on upcoming actions from the crosswalk control point (yielding, braking, etc.). This control point includes information about the nearest vehicle in each lane as well as whether pedestrians are waiting to cross and/or currently crossing. In the simulation, driver decisions on yielding are made once the driver encounters a waiting pedestrian when they are at or beyond the decision point. Driver decisions are permanent unless a safety concern is identified. Additionally, the simulation provides the option for pedestrians to reject a yield; however, the default probability for this case is zero. In the case of a rejected yield, drivers would react within one time step and continue car-following behavior. When no yielding is active the vehicle remains in normal car following behavior.

Pedestrians are limited to linear movement without interaction between pedestrians. Pedestrians enter with a desired walking speed and when able to safely cross, continue walking at their desired speed. This desired speed is used when the simulation estimates how long the crosswalk will be occupied in order to determine vehicle stop time or soft yield dynamics. Pedestrians that are first in queue at a yield will always accept a yield if all lanes are safe to cross. To determine the movement of other queued pedestrians the simulation uses the gap acceptance model if the yield type is a soft yield, while all queued pedestrians accept hard yields. This prevents constant re-evaluation of soft yield dynamics for the yielding vehicle so that the yield choice model is used once per crosswalk approach.

Limitations

Due to the assumptions made, the pedestrian delay only includes waiting time at the edge of the crosswalk and not any pedestrian friction effects from increased density in the crosswalk. Friction effects as well as pedestrian pooling/queueing could be analyzed if the framework was implemented in a simulation environment that included modeling of more detailed pedestrian movement. Distributions of pedestrian characteristics are taken from the data collected in the observational study and are not necessarily representative of any one location.

Additionally, in the simulation, vehicles make yielding decisions only when they are the closest vehicle in their lane to the crosswalk. Following vehicles maintain car following behavior until their leader passes through the crosswalk, so queued and platooned vehicles will make yield decisions very close to the crosswalk. This approach was used because the observational study only collected information about the lead vehicle in platoons and queues. Therefore, the following driver's ability to decide whether to yield prior to their leader passing through the crosswalk cannot be accurately simulated.

Finally, all of the models and submodels were developed, calibrated, and validated based on our study sites. Conditions with different vehicle and pedestrian flows and geometry cannot be accurately simulated. A sensitivity analysis is included in this chapter to identify trends especially around boundary conditions.

Simulation Flow Chart

The overall simulation flow chart is shown in Figure 27. The simulation is time step-based with a current time step of 0.1 seconds. Once the simulation is initialized, all vehicle movements and decisions are made prior to pedestrian movements and decisions at each time step. Major submodels developed for the simulation are shown in blue, with discussion on the implementation of each submodel in the following section.

In addition to the submodels shown in the flow chart, there are many constraints that are checked continuously to ensure that the simulation is realistically recreating the vehicle-pedestrian interaction. In order to prevent large vehicle queue spillback in high pedestrian volume situations, a maximum waiting time is used. After this time the vehicles are forced to pass through the crosswalk.

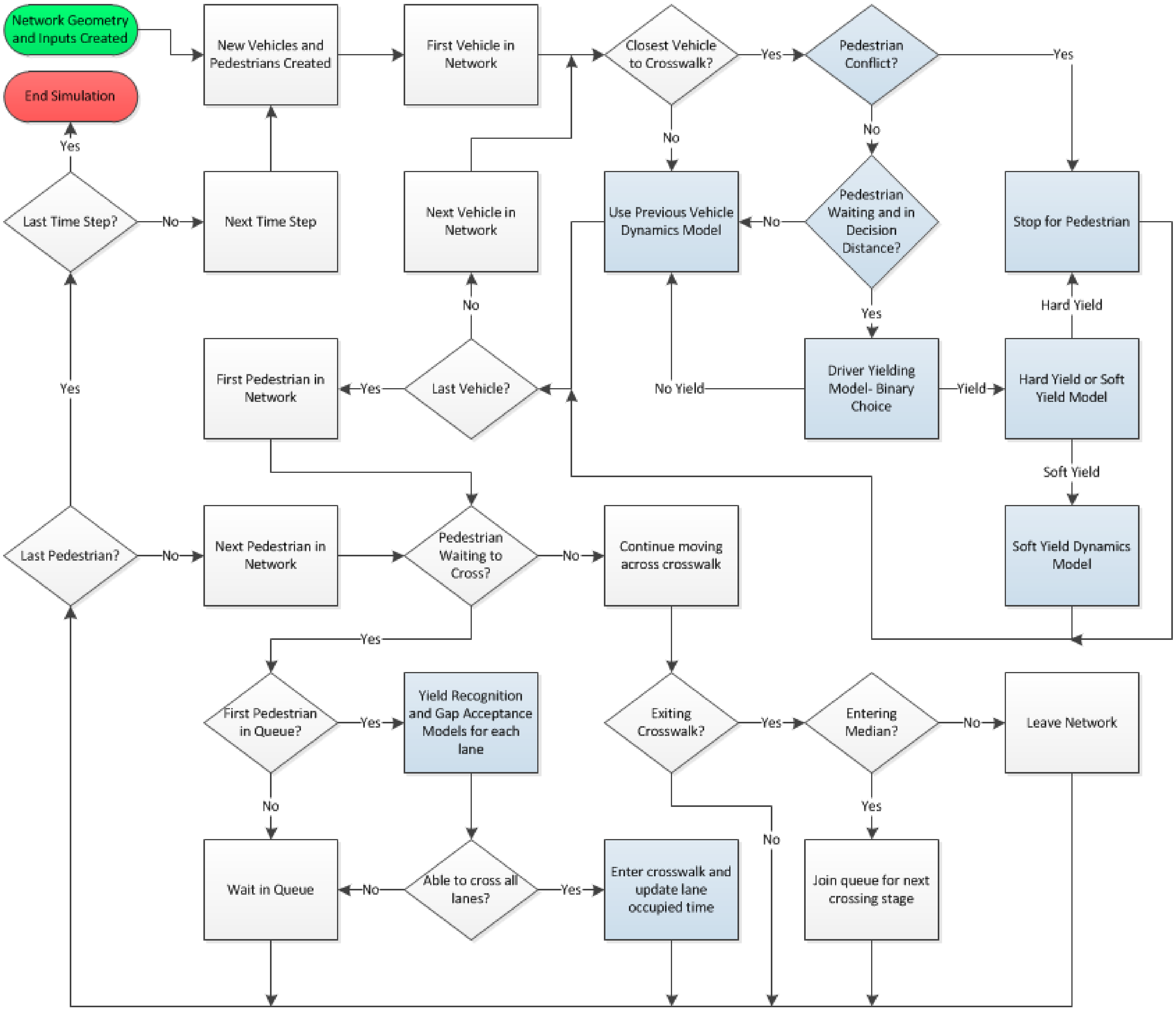


Figure 27: Simulation Flow Chart

Submodel Descriptions

Pedestrian-Vehicle Conflict Identification

As each vehicle approaches a crosswalk, the simulation checks at each time step for conflicts with currently crossing pedestrians. If the vehicle arrival time to the crosswalk is lower than the time that the pedestrian will occupy their lane in the crosswalk, the vehicle is forced to yield in order to avoid the conflict. This yield will take precedence over any other vehicle decision in order to model safe vehicle movement.

Driver Decision Distance

As each vehicle approaches a crosswalk, they are assigned a unique decision point. The decision point is associated with each crosswalk encountered and is unique to each vehicle/crosswalk combination. Two models were developed, one for low speeds and one for higher speeds, as discussed in the previous chapter.

Low-Speed Group:

Based on the inverse cumulative distribution function of Burr Distribution, we randomly generate a double number which is between 0 and 1. That number serves as the random position in the data distribution. The driver decision point for the low FFS group is estimated as:

$$\boldsymbol{Y = 214.533\left((1 - rand)^{\frac{1}{-3.643}} - 1\right)^{\frac{1}{2.241}}} \quad \textbf{Equation 31}$$

High-Speed Group:

Similarly, a random number is generated in the simulation and the driver decision point for the high FFS group is estimated as:

$$\boldsymbol{Y = 34.363\left(-\ln\left(\frac{1}{prob} - 1\right)\right) + 9.43 * DesiredSpeed - 159.17} \quad \textbf{Equation 32}$$

Driver Yield Decision

The driver yielding model only deals with the yield vs. no-yield option. If a vehicle decides to yield to waiting pedestrians, the driver yield check model is used to determine the type of yield between Hard-Yield and Soft-Yield. The binary choice model described in the previous chapter is implemented to determine the probability of a yield based on the vehicle and pedestrian characteristics. A random number between 0 and 1 is drawn and if the value is less than or equal to the probability of yielding then a yield is modeled, otherwise the vehicle continues in car-following mode.

Driver Yield Type Check

After the decision to yield is made, a binary choice model (described in the previous chapter) is used to select between a hard yield and a soft yield. A random number between 0 and 1 is drawn and if the value is less than or equal to the probability of soft yield, then a soft yield is modeled, otherwise the driver performs a hard yield.

Soft Yield Dynamics

Once a vehicle decides to soft yield, the trajectory of the soft yield is determined using the model described in the previous chapter. This trajectory is used as long as a conflict is not identified.

Hard Yield Dynamics

In the simulation, hard yields are modeled identically to stop control. Other models were considered but the data collected in the instrumented vehicle study did not indicate that hard yields differed significantly from full stops at signals.

Driver Yield Rejection Response

Currently the simulation allows for rejected yields for pedestrians, but the default in the simulator is that this has a zero probability of occurrence. If this default changes, vehicles would react to a rejected yield and return to car following behavior after one time step.

Driver Wait Time

In the case of extremely high pedestrian demand, presumably the driver would only remain stopped for a finite amount of time. The literature and our data collection did not provide any reasonable estimates for this maximum driver wait time. Therefore, in the simulation we used a default value of 60 sec (600 * 0.1 sec/time step), which was a bit longer than the longest wait we observed in the field. This parameter could be used to calibrate queue length for high pedestrian demand scenarios.

Pedestrian Yield Recognition

In the simulation, pedestrians “recognize” both hard yields and soft yields when the decision is made for the subject pedestrian. Any subsequent pedestrians will accept hard yields and step out. However they will not accept soft yields, as the vehicle trajectory has already been determined based on the first pedestrian.

Pedestrian Yield Rejection

Simulated pedestrians will accept all yields that they recognize under the previously described model. One potential extension for this model that may be added in the future is the ability to allow for blind pedestrians to only accept hard yields.

Pedestrian Gap Acceptance

Pedestrians “observe” gaps at each time step to determine the probability of accepting a gap. The gap acceptance model is described in the previous chapter. A random number between 0 and 1 is drawn and if the value is less than or equal to the probability of accepting a gap then the lane is safe to cross, otherwise the lane is not safe and the pedestrian will not cross. A safe crossing may be identified by either the yield recognition model or gap acceptance model in order for the lane to be safe to cross.

Simulation Output

Comparison of Simulation Output to Observational Data Set

A set of test runs of the simulation were made to compare simulation outputs to a dataset from the observational study to identify how well the operations were recreated. Test site UF-5 was selected with one lane of traffic in each direction and no median refuge for pedestrians. The site was modeled during the PM peak hour with vehicle flows of 550 veh/hr in the westbound direction and 250 veh/hr in the eastbound direction. Pedestrian flows are 250 ped/hr for both approaches of the midblock crosswalk. The average vehicle free flow speed was set to 25 mph, and 800 ft vehicle links were simulated before and after the crosswalk.

A total of 10 runs were made with these flow rates for a 15-minute duration in order to identify the average delay to vehicles due to the midblock crosswalk. Data on vehicle travel time and pedestrian travel time were compared to free flow travel time for each mode to determine the delay for each mode. Other information collected included pedestrian queue lengths, pedestrian characteristics, and vehicle acceleration modes. All other vehicle and pedestrian properties were set to default values, thus information such as pedestrian walking speed and gender were not site-specific values, but global values from the observational study dataset.

Vehicle Properties

Vehicle desired speeds are based on driver types and are distributed around the input link desired speed. Additionally, two vehicle classes are modeled, heavy vehicles and passenger cars. Table 19 shows the vehicle properties averaged from the 10 test runs as well as the proportion of vehicles that came to a complete stop for pedestrians as a leader.

Table 19: Vehicle Properties

	Westbound	Eastbound
% Heavy Vehicles	8.9%	7.4%
Desired Speed (mph)	24.85	24.88
% Stopped for Peds	28.2%	33.9%

Pedestrian Properties

Pedestrian characteristics were modeled based on default global distributions from the observational study dataset. Table 20 compares the simulated pedestrian properties to the observed properties. Overall, pedestrian characteristics from the simulation match the distribution of characteristics seen in the observational study dataset for all sites, but Site UF-5 does not match perfectly. This location is the crossing from a parking lot to a university gym, which explains the strong presence of younger pedestrians in casual attire.

Table 20: Pedestrian Properties

	Simulation	Site UF-5	All Sites
> 30 Years Old	19.7%	0%	21.5%
Casual Attire	88.0%	100%	89.6%
Male	60.3%	87.8%	60.1%

Operational Results

Vehicle and pedestrian delay were calculated for the simulation by subtracting free flow travel time from actual travel time. Comparisons for delays by approach are also appropriate as the different vehicle volumes will create different queuing dynamics. Vehicle delays could not be identified during the observational study, but pedestrian delays were recorded and are compared to simulated pedestrian delay.

Figure 28 shows the distribution of vehicle delays for the scenario sorted by approach direction. On average, the difference in delay is 7.38 seconds, with an average of 7.51 seconds for the westbound approach and 7.25 seconds for the eastbound approach. Many of the smallest delays may be attributed to vehicles following a leader with a smaller desired speed.

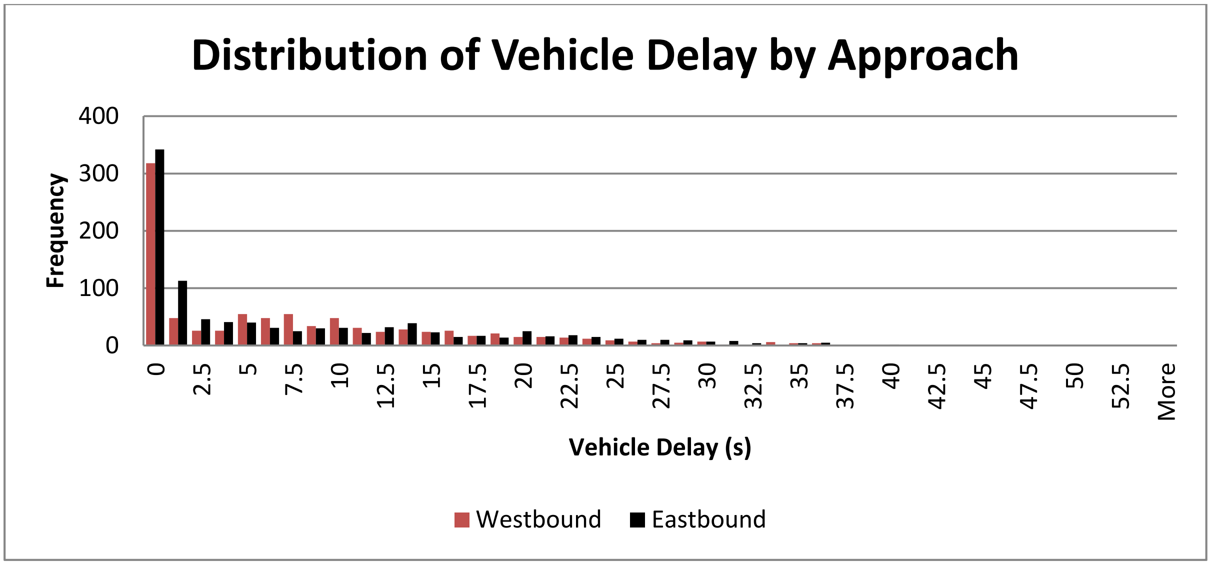


Figure 28: Vehicle Delays

Figure 29 shows the distribution of pedestrian delays in the observational dataset and in the simulation. Overall, there are a higher proportion of delays in the simulation than the full observational dataset and the observations at site UF-5. The average delay for the simulation is 0.383 seconds, while it is 1.987 seconds for the full observational dataset and 0.257 seconds for site UF-5. This result indicates that using the global default pedestrian parameters results in similar rates of no delay, but the average delay depends much more on site-specific flow rates. One major difference between the simulation and both observed datasets is the occurrence of very short delays, i.e., less than one second. The data collection procedure for the observational study used stopwatches to count pedestrian waiting time and was not as accurate as the simulation can be. Overall, the simulation recreates reasonable values of delay for a given set of site-specific flow rates, but further work is needed to replicate the larger delays observed.

Sensitivity Analysis

A preliminary sensitivity analysis was run to identify trends in delay for vehicles and pedestrians based on vehicle and pedestrian volumes using the geometry shown in the previous section. A total of 9 scenarios were run with identical vehicle flow rates for each approach and identical pedestrian volumes for each crossing direction. Each combination of 75, 125, and 200 pedestrians per hour per approach and 250, 500, and 750 vehicles per hour per approach was modeled for 15 minutes.

Figure 30 shows the average vehicle delay for each of the 9 scenarios. A clear pattern of increased delay with increased pedestrian volumes is seen, but the differences in vehicle volumes do not have a similar trend. This may be due to the fact that the range of vehicle volumes does not reach close to capacity, while a crosswalk with 400 total pedestrians per hour (200 ped/hr each way) is very crowded. Unlike vehicles, no pattern can be found in the average pedestrian delay for the 9 scenarios shown in Figure 31. Similarly to vehicle delays, this may be explained by the low vehicle volumes, with pedestrian volumes not having a large effect due to the limitations of the pedestrian model not including pedestrian to pedestrian interaction. Further study with a much larger set of scenarios may identify trends in both vehicle and pedestrian delays.

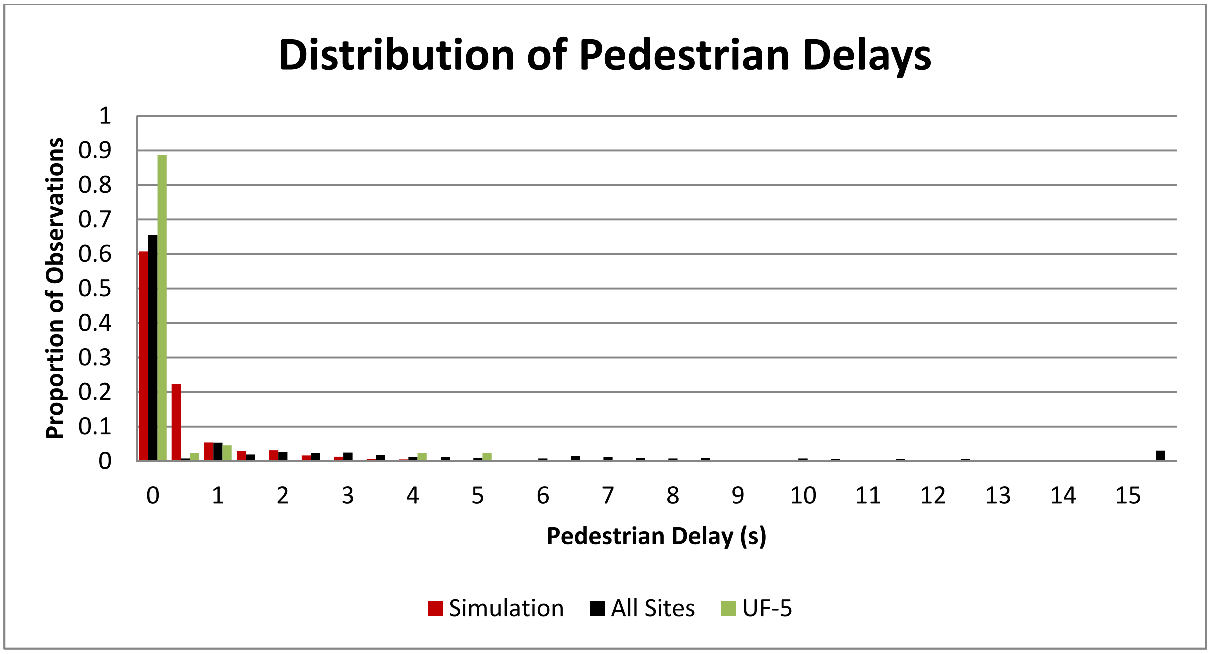


Figure 29: Pedestrian Delays

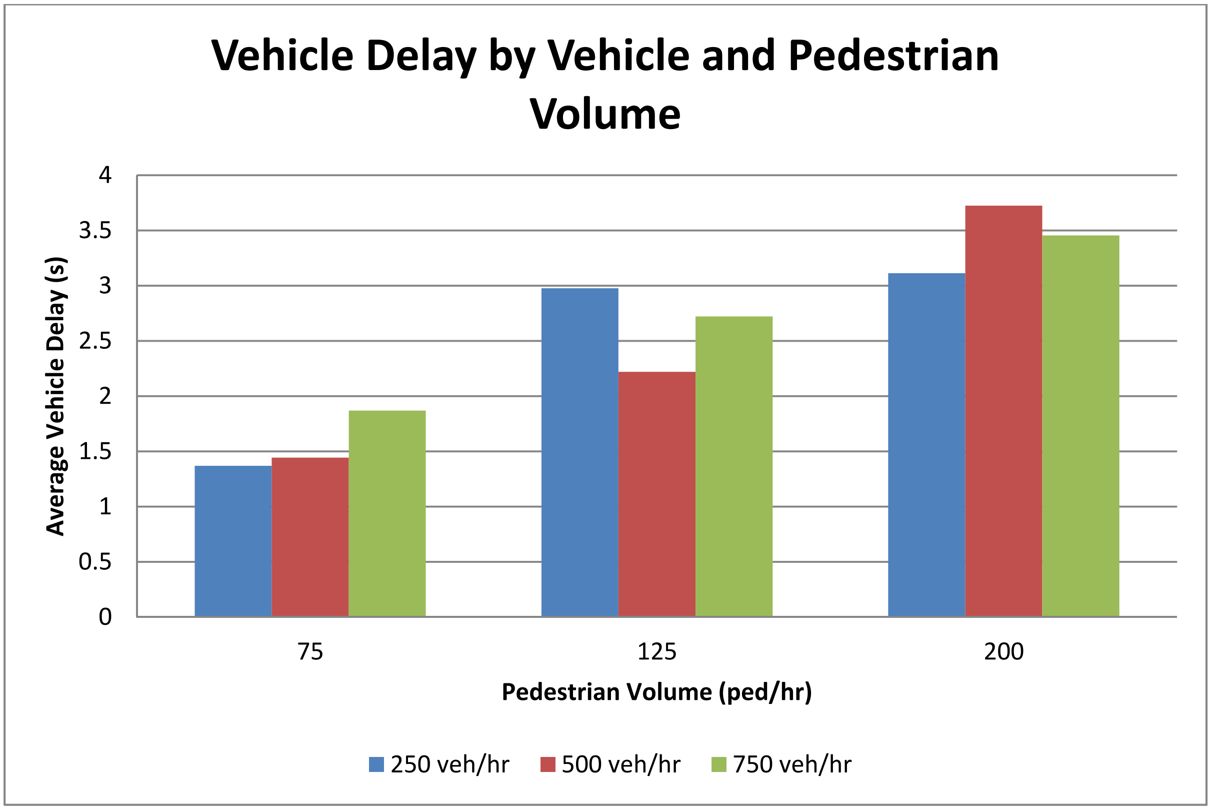


Figure 30: Average Vehicle Delay by Vehicle and Pedestrian Volume

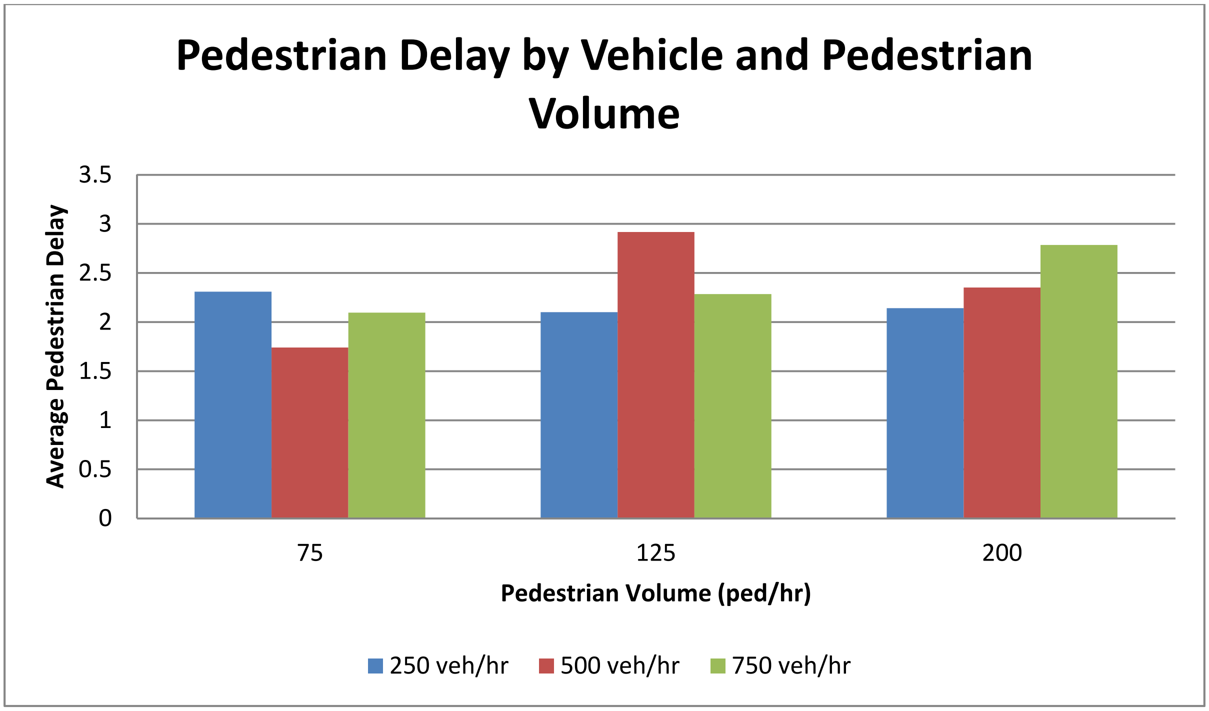


Figure 31: Average Pedestrian Delay by Vehicle and Pedestrian Volume

- This page intentionally left blank -

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

The objective of this research was to provide an improved understanding of pedestrian-vehicle interaction at mid-block pedestrian crossings and develop methods that can be used in traffic operational analysis and microsimulation packages. Models describing driver yielding and pedestrian gap acceptance behavior were developed from field data collected at 27 mid-block pedestrian crossings in three states (Alabama, Florida, and North Carolina), encompassing two different types of land use: university campuses and downtown areas. The study included an in-vehicle driver behavior study with 15 drivers performed in Florida. This project implemented algorithms describing vehicle-pedestrian interactions in a micro simulator and developed educational modules for dissemination of the research results to students and professionals in the southeast and nationwide. Specific outcomes for this research include: (a) a standalone model of pedestrian gap acceptance behavior at unsignalized crossings, (b) a driver yielding behavioral model, (c) models describing vehicle dynamics and driver behavior in advance of the crosswalk, and (d) prototype algorithms incorporated and tested in a micro simulator. Key deliverables include the prototype algorithms implemented in simulation, a final report summarizing the research and findings, and educational modules on the research results that can be incorporated into university curricula, or serve as material for standalone professional development courses.

This chapter summarizes the results of the research on yielding and gap acceptance models and implementation of these models into microsimulation. The chapter consists of separate sections for summary and conclusion for driver yielding models, pedestrian gap acceptance models, and the results of the instrumented vehicle experiment. The chapter then presents limitations of the research and recommendations for future efforts.

Conclusions

This section presents the research conclusions for the yielding models, gap acceptance models, and the instrumented vehicle experiment.

Yielding Model Conclusions

The yielding models were developed as a binary logit model after testing various types of other model forms. Binary logit models best describe the probability of yielding as a function of microscopic traffic parameters related to the interaction of pedestrians and vehicles at the crosswalk. In the binary logit model the likelihood of yielding is represented as a value "1" and non-yielding as "0". Various selection methods, including a full model with all variables, forward selection, and backward elimination were tested to explore effects of various independent variables. Eventually, a custom model was developed informed by these results.

The proposed yield model takes into account practical significance of model terms, and feasibility of implementing the variables into simulation. The team developed two final models, one is calibrated for the data collection sites and includes the effect of each of the three states. The other is a universal model that is independent of behavior in a particular state. The two final models have R^2 values of 0.3582 and 0.2680, respectively. These results speak to the fact that a lot of variability in yielding behavior was not explained by the models. The team therefor believes that much of the variability in yielding is a function of driver preferences and awareness of the situation at the crosswalk, neither of which can be evaluated with the observational study performed.

In the final models, increased speed (SPD) was seen to reduce the likelihood of yielding, as did an increased required deceleration rate (DECEL). Presence of adjacent yields (ADJ), low speed platoons (LSPLT), presence of multiple pedestrians (MUP), and female pedestrians (FEMALE) were seen to increase the likelihood of yielding. Drivers are also more likely to yield to pedestrians on-campus

(CAMPUS) than off-campus. The coefficient estimates for Florida and North Carolina show that drivers are more likely to yield in Florida and less likely to yield in North Carolina, both relative to drivers in Alabama.

In addition to the yield model prediction, nested logit models were used to predict the likelihood of a hard yield given that a yield occurred. A hard yield is defined as drivers coming to a full stop, as opposed to a rolling or soft yield, where vehicles are still in motion during the yield event. These models represent two-stage binary logit models, with the first stage consistent with the initial binary logit model. The second level predicts the likelihood that a driver performs a hard yield, given that the first level of the nested logit predicts a yield. The overall probability of a driver hard yielding can be calculated by multiplying the two probability functions. The soft yielding probability correspondingly is defined as the likelihood of a yield multiplied by one minus the probability of a hard yield.

The final model to predict the likelihood of a hard yield has an R^2 value 0.2507. Longer distances from the crosswalk (ADJDIST) and being in the lane closest to the pedestrian (NEAR) decrease the chance of a driver deciding to hard yield. This is likely because drivers that are further away from the pedestrian have time to react with a soft yield. Drivers are more likely to hard yield at Florida sites than North Carolina sites. Adjacent yields (ADJ), presence of multiple pedestrians (MUP), and higher necessary deceleration rates (DECEL) increase the chance that a driver will hard yield. If the deceleration rate required to yield is higher, then it is reasonable that the chance of hard yielding is increased.

Gap Model Conclusions

Using a Probit model formulation, gap acceptance models were fitted to data from the crosswalks in three states. Four models were explored, including (a) single lane non-controlled crossings; (b) single lane staged crossings; (c) all sites combined with non-controlled crossings and (d) all sites combined with staged crossings. The single lane models in (a) and (b) focused on sites with a single lane of traffic in each direction; approaches (c) and (d) also included data from a few three-lane and four-lane sites, with an adjustment variable to control for varying crossing widths. Non-controlled crossings refer to observational data from naturally-occurring pedestrian traffic at the crosswalks, while staged crossings refer to crossing decisions made by members of the research team.

Single lane gap acceptance models, though simple in nature, were observed to have greater explanatory power, while also offering the advantage of relative ease of integration in a simulation environment. Goodness of fit tests showed that the single lane gap acceptance model drawn from the non-controlled dataset yielded the best model (Max Re-scaled R^2=0.69). Thus this model was recommended for adoption in the simulation process, since it is developed from an internally-consistent data set (all single-lane decisions), and is derived from naturally occurring pedestrian events, thus avoiding potential bias by including crossing decisions made by members of the research team.

The recommended model uses only two parameters, the size of the gap length in seconds, and a binary variable distinguishing between gaps and lag events (first arriving vehicle, without a prior lead vehicle to “open” the gap). An increase in gap length is associated with an increased probability of pedestrians crossing. A lag event has a negative coefficient, meaning that a pedestrian is less likely to accept a lag than a gap given the same length in seconds. This effect may be explained because the pedestrian may require some of the lag time to evaluate the available time to cross, after first arriving at the crosswalk. For a gap event, these “screening time” and decision making likely takes place before the gap “opens”.

In the present study, every gap event was considered as an independent event. However, there are occasions where pedestrians tend to accept gaps on rolling basis (lane by lane basis). These lane-by-lane accepted gaps can be thought as a set of sequential discrete choices. Further efforts are recommended to develop models for multilane rolling gap acceptance.

Instrumented Vehicle Conclusions

The instrumented vehicle experiment was conducted to provide insight on other models and variables that define the interaction between a vehicle and pedestrian(s) at the crosswalk. The models are implemented in the simulation to further supplement the yield and gap acceptance models described above. Overall, nine supplemental models were developed, for a total of eleven predictive models:

1. Driver Yield Decision (from observational study),
2. Pedestrian Gap Acceptance (from observational study),
3. Pedestrian-Vehicle Conflict Identification,
4. Driver Decision Distance,
5. Driver Yield Type Check,
6. Soft Yield Dynamics,
7. Hard Yield Dynamics,
8. Driver Yield Rejection Response,
9. Driver Wait Time,
10. Pedestrian Yield Recognition, and
11. Pedestrian Yield Rejection.

One of the models developed based on instrumented vehicle experiment is the driver's decision point model. This model is developed to explain at which location relative to the crosswalks the driver makes a decision of Yield/No-Yield to the waiting pedestrians at curb. Based on the data collected from the instrumented vehicle experiment two models are developed, one for low speeds and one for higher speeds.

Soft-Yield Dynamics model is developed based on the GPS data from in-vehicle study. It showed that vehicles tend to slow down (with a lower rate than stopping behaviors) and then coast (with a constant speed) until they pass the crosswalk. The results indicate that the vehicle deceleration rate of Soft-Yield behavior is a function of speed and distance at decision point with an R square of 0.3388. The total travel time during that Vehicle-Pedestrian Interaction process is the sum of crosswalk occupied time by pedestrians plus 0.25 sec of safety buffer.

Project Deliverables

The objective of this research was to develop new and improved algorithms for describing pedestrian and vehicle interaction at unsignalized midblock pedestrian crossings and to implement them in a traffic simulation environment. The algorithms developed address pedestrian and driver behavior at mid-block crosswalks, based on targeted empirical observations of naturally occurring and staged crossings. The models describe pedestrian gap selection and driver yielding behavior, and are compatible in form with algorithms used in microsimulation tools.

The eleven resulting algorithms were implemented in a microsimulation model to assure successful technology transfer. The documentation of the behavioral models will allow for implementation of the algorithms in other simulation software packages as well. The behavioral models are based on empirical observations, and were derived from field observations of naturally occurring and staged pedestrian crossings.

Results from this research may be used to support policy-level recommendations on multimodal transportation infrastructure design that considers pedestrian access. Specifically, the research produced:

Specific outcomes for this research include: (a), (b), (c) models describing vehicle dynamics and driver behavior in advance of the crosswalk, (d) prototype algorithms incorporated and tested in a micro simulator, and (e) educational modules for dissemination of the research results.

1. A standalone model of pedestrian gap acceptance behavior at unsignalized crossings, sensitive to the available gap length and distinguishing gap and lag events, and developed from data collected at 24 unsignalized crosswalks in three states;
2. A driver yielding behavioral model developed from 27 unsignalized crosswalks in three states, sensitive to vehicle dynamics, pedestrian attributes, traffic condition, and geographical area (North Carolina, Florida, and Alabama);
3. Nine supplemental model describing various attributes of pedestrian-vehicle interaction, derived from an in-vehicle driver study with 15 participants conducted in Florida;
4. Implementation of these algorithms in a simulation environment, which for the first time incorporates pedestrian-vehicle interactions at unsignalized crossings in a micro simulator; and
5. Educational modules for dissemination of the research results to students in the southeast and nationally, supported by seamless technology transfer through the available simulation modules in a readily available simulator.

Implications for Practice

This research has broad impact on the state of the practice of pedestrian analysis in the Southeast region, and likely beyond. The field of pedestrian analysis and modeling has documented gaps and limitations, and this research aims to make significant improvements to the ability to model pedestrian traffic. In an age of increasing focus on accommodation of non-motorized road users in our transportation systems, engineers need the tools to evaluate the impacts of different intersection treatments on both pedestrians and the conflicting vehicle stream. Oftentimes, engineering analyses include the use of microsimulation tools, which to this point had not been specifically calibrated for pedestrian-vehicle interaction behavior. The behavioral models resulting from this research will assist in evolving these microsimulation tools to the point where analysts can predict the operational characteristics of unsignalized pedestrian crossings.

This research delivered an improved understanding of pedestrian and driver behavior at unsignalized midblock crossing points and provides practitioners with enhanced tools for considering pedestrian presence. This goal is being achieved by developing algorithms for microsimulation tools to model the interaction between pedestrians and drivers. The pedestrian-vehicle interaction simulation was implemented in a microsimulator. This project adds pedestrian movement and interaction with vehicles to the existing model that is programmed in the C# programming language to demonstrate an application of the statistical interaction models in microsimulation. The simulation developed for this project is a proof of concept, as the models of interaction can be applied in any microsimulation that includes vehicles and pedestrians. The simulation is time step based and once the simulation is initialized, all vehicle movements and decisions are made prior to pedestrian movements and decisions, at each time step. Major sub models developed for the simulation are the results of various data collection and modeling efforts from this research and are presented in detail in the previous chapters.

Limitations

Data Collection Limitations

The data used for this research was collected at multiple midblock crosswalks at on and off of campus locations in the three states of North Carolina, Florida and Alabama. While these states all belong to the south-east region of the United States, they have shown different driver behavior, with state indicator variables being significant in the yielding models. This suggests that other states may have yet different yielding behavior. Therefore, one of the limitations of the model is that the yielding model calibration variable is only estimated for these three states.

While doing the data collection all variables related to the interaction of the pedestrian and vehicle and any additional variables related to either the pedestrian or vehicle were recorded. However, not all of these variables were considered for modeling due to small sample size. For example, only 3% of the observations included an interaction between a pedestrian and a heavy vehicle, and therefore that variable was eliminated for final modeling since it wasn't being fairly represented in the dataset.

For the staged pedestrian crossings in North Carolina only female pedestrians were tested and in the Alabama only male participants were used. Naturally occurring pedestrians are a mix of females and males pedestrians. This limitation was overcome by limiting the recommended gap acceptance model to only naturally-occurring pedestrians, which included both male and female pedestrians in all three states.

Model Limitations

Data collection limitations impose similar model limitations. The models predict the likelihood of an action taken by the pedestrian or the driver. In case of a yield, given certain conditions, such as speed, proximity to crosswalk or other factors, the likelihood of yielding increases or decreases. However, the probability of yielding is only being calculated based on the variables used in the model. In a real condition other factors might play in yielding. These factors include driver courtesy, whether the driver is rushed to get to work for example, and other behavioral factors that cannot be captured in an observational study. The same is true for the likelihood of accepting a gap in traffic by a pedestrian, where pedestrian attitude, risk preference, or even visual and cognitive abilities cannot be observed. Previous research has shown that the likelihood of accepting a shorter gap increases as pedestrian wait time increases at the crosswalk. This factor is not reflected in the gap acceptance model.

Simulation Implementation Limitations

The simulation Implementation in Chapter 6 discusses the simplifying assumptions and limitations of the methodology in detail. A summary of the limitations is presented here. Other than the models developed as part of this research, simulation requires comprehensive details regarding the interaction of the pedestrians and vehicles at the crosswalk. Therefore the simulation implementation includes limitations related to both pedestrian behavior and driver/vehicle behavior. All of the models and sub models were developed, calibrated, and validated based on our study sites. Conditions with different vehicle and pedestrian flows and geometry cannot be accurately simulated using our tool.

Vehicles make yielding decisions only when they are the closest vehicle in their lane to the crosswalk. Following vehicles maintain car following behavior until their leader passes through the crosswalk, so queued and platooned vehicles will make yield decisions very close to the crosswalk. This approach was used because the observational study only collected information about the lead vehicle in platoons and queues. Therefore, the following driver's ability to decide whether to yield prior to their leader passing through the crosswalk cannot be accurately simulated using our data.

The pedestrian delay in this simulation approach only includes waiting time at the edge of the crosswalk and not any pedestrian friction effects from increased density in the crosswalk. Distributions of pedestrian characteristics are taken from the data collected in the observational study and are not necessarily representative of any one location.

Future Research

This research focused on studying pedestrian crossings at unsignalized midblock crosswalks in three states, North Carolina, Alabama and Florida. Although a universal model is developed for yielding and gap acceptance models, future research is needed to develop calibration factors for other states across the United States. The models include two types of land-use, university campus areas and down-town crosswalks. Future research should consider other types of land use, such as residential, recreational, or suburban and rural areas, as well as sites in other states. The research does not include any mid-block crosswalk with any type of treatment such as a speed table or a pedestrian beacon. Additional research is needed to test the driver compliance with such treatments and quantify the effect of these treatments on yield and gap acceptance.

Additional modeling approaches are recommended as future research. Multinomial regression and logit models, as well as ordered probit models, could be used to predict the probabilities of hard and soft yielding. Macroscopic characteristics could be examined in future modeling efforts. These characteristics could include number of lanes, lane width, and crossing or striping type. A continuous decision model will be considered as opposed to event-based modeling. This may better represent the true interaction between drivers and pedestrians in decision-making, since decisions are likely not made at a single point but change over time.

REFERENCES

Agresti, A. (2007), An Introduction to Categorical Data Analysis, Second Edition. John Wiley & Sons Inc., Hoboken, New Jersey.

Alhajyaseen, W. K. M., Asano, M., Nakamura, H. (2012), Estimation of left-turning vehicle maneuvers for the assessment of pedestrian safety at intersections, IATSS Research, Volume 36, Issue 1, July 2012, Pages 66-74.

Ashmead, D., Guth, D., Wall, R., Long, R., & Ponchillia, P. (2005), *Street crossing by Sighted and blind pedestrians at a modern roundabout*. In ASCE Journal of Transportation Engineering, Vol. 131, No. 11, November 1, 2005, 812-821.

Avineri, E., Shinar, D., Susilo, Y. O. (2012), Pedestrians' Behavior in Cross Walks: The Effects of Fear of Falling and Age. Accident Analysis and Prevention 44, 30–34.

Ben-Akiva, M. and S. Lerman (1985), *Discrete Choice Analysis*, MIT Press, Cambridge, MA. Referenced in Cambridge Systematics, *NGSIM Task E.1-1: Core Algorithm Assessment.* Prepared for Federal Highway Administration. Oakland, California. February 2004.

Bennett, S., A. Felton, and R. Akcelic (2001), *Pedestrian Movement Characteristics at Signalized Intersections,* 23rd Conference of Australian Institutes of Transport Research (CAITR 2001), Melbourne, Australia.

Blue, V. J. and Adler, J. L. (2001), Cellular automata microsimulation for modeling bi-directional pedestrian walkways, Transportation Research Part B: Methodological, Volume 35, Issue 3, March 2001, Pages 293-312.

Blue, V. J. and Adler, J. L. (2000), *Cellular Automata Model of Emergent Collective Bi-Directional Pedestrian Dynamics.* In Artificial Life VII Bedau, M.A., McCaskill, J.S. Packard, N.H. and Rasmussen, S. MIT Press, 2000.

Bromberg, S., Oraon-Gilad, T., Ronen, A., Borowsky, A., Parmet, Y. (2012),The perception of pedestrians from the perspective of elderly experienced and experienced drivers, Accident Analysis and Prevention 44, pp. 48-55.

Boyce, T.E and Geller, S.E. (2012), An instrumented vehicle assessment of problem behavior and driving style: Do younger males really take more risk. Accident Analysis and Prevention 34, 51-64.

Cambridge Systematics (2004), *NGSIM Task E.1-1: Core Algorithm Assessment.* Prepared for Federal Highway Administration. Oakland, California. February 2004.

Cassidy, M. et al. (1995), Unsignalized Intersection Capacity and Levels of Service: Revisiting Critical Gap, Transportation Research Record 1484, pp. 16-23. Referenced in Cambridge Systematics, NGSIM Task E.1-1: Core Algorithm Assessment. Prepared for Federal Highway Administration. Oakland, California. February 2004.

Charron, C., Aurelie, F., Gueguen, N. (2012), Do child pedestrians deliberately take risks when they are in a hurry? An experimental study on a simulator. Transportation Research Part F 15, pp. 635-643.

Chraibi, M., Seyfried, A., Schadschneider, A., Mackens, W. "Quantitative Description of Pedestrian Dynamics with a Force-Based Model," Web Intelligence and Intelligent Agent Technologies (2009), WI-IAT '09. IEEE/WIC/ACM International Joint Conferences on, vol.3, no., pp.583-586, 15-18 Sept. 2009.

Chu, X., Guttenplan, M., and Baltes M. (2002), "Why People Cross Where They Do." TRB Paper No. 03-3078, TRB, Washington, DC, September 2002.

Cooper, J., Yager, C., Chrysler, S. T. (2011), An Investigation of the effects of reading and writing text-based messages while driving" Technical Report: SWUTC/11/476660-00024-1.

Davis, G. and Inman, V. (2007), *Pedestrian Access to Roundabouts: Closed Course Test of Yielding Vehicle Detection System*. http://www.accessmanagement.gov/AM2004/AM0411p_Roundabouts.pdf. Last Accessed February 8, 2007.

Dewar, R (2002), Human Factors in Traffic Safety, "Chapter 18: Pedestrians and Bicyclists." Lawyers and Judges Publishing Company, Tucson, AZ, 2002.

Dunn, R., and R. Petty (1984), *Mid-Block Pedestrian Crossings – An Examination of Delay.* 12th Annual Australian Road Research Board Conference Proceedings, Hobart, Tasmania, Australia. August 1984.

Edquist, J., Rudin-Brown, C. M., Michael, G. L. (2012), The effects of on street parking and road environment visual complexity on travel speed and reaction time. Accident Analysis and Prevention 45, pp. 759-765.

Fellendorf, M., Schönauer, R., Huang, W. (2012), Social Force based Vehicle Model for Two-Dimensional Spaces, Presented at the Transportation Research Boards 91st Annual Meeting, Washington, D.C.

FHWA (2000), *Roundabouts: An Informational Guide*. Federal Highway Administration. Turner Fairbank Highway Research Center. FHWA-RD-00-067. McLean, VA. 2000.

FHWA (2003), *Manual on Uniform Traffic Control Devices for Streets and Highways (MUTCD)*. US DOT, FHWA, Washington, DC. 2003. Available online at http://mutcd.fhwa.dot.gov, accessed January 2007.

FHWA (2004), *NGSIM Task E.1: Core Algorithm Assessment.* Publication No. #FHWA-HOP-06-0009. Department of Transportation. Federal Highway Administration. Washington, DC. 2004.

FHWA (2009), Manual on Uniform Traffic Control Devices (MUTCD), Federal Highway Administration, Washington, DC., 2009.

Fisher, D., Garay-Vega, L. (2012), Advance yield markings and driver's performance in response to multiple-threat scenarios at midblock crosswalks. Accident Analysis and Prevention, 44 pp. 35-41.

Fitzpatrick, K., et al. (2006), *Improving Pedestrian Safety at Unsignalized Intersections*. TCRP Report 112/NCHRP Report 562. Transportation Research Board. http://onlinepubs.trb.org Accessed February 2008. Washington, D.C. 2006.

Florida Pedestrian Planning and Design Handbook. Florida DOT, Tallahassee, FL, April 1999. Available online at: http://www11.myflorida.com/safety/ped_bike/ped_bike_standards.htm#Florida%20 Ped%20Handbook, accessed April 14, 2003.

Geruschat, D. R., and. Hassan, S. E. (2005), Driver Behavior in Yielding to Sighted and Blind pedestrians at Roundabouts. Journal of Visual Impairment & Blindness 99(5):286-312.

Guo, X., Dunne, M.C., Black, J.A. (2004), "Modeling of Pedestrian Delays with Pulsed Vehicular Traffic Flow." Transportation Science, Vol. 38, No. 1, February 2004, pp. 86-96.

Guth D., Ashmead D., Long R., Wall R., Ponchillia P. (2005), Blind and Sighted Pedestrians' Judgments of Gaps in Traffic at Roundabouts. *Human Factors*, 47, 314–331.

Harkey, D. L. and Zegeer, C. V. (2004), *PEDSAFE: Pedestrian Safety Guide and Countermeasure Selection System*. Federal Highway Administration. FHWA-SA-04-003, Washington, DC. September 2004.

Hatfield, J. and Murphy, S. (2007), The effects of mobile phone use on pedestrian crossing behavior at signalized and un-signalized intersections. Accident Analysis and Prevention 39, 197-205.

Helander, M. B. and Hagvall, B. (1976), An Instrumented Vehicle for studies of driver behavior. Accident Analysis and Prevention 8, 271-277.

Helbing, D. and Molnár, P. (1995), Social force model for pedestrian dynamics, Phys. Rev. E 51, 4282–4286.

Holden, R. and A. Cangelosi (2003), *Cellular Automata Models of Human Traffic.* Submitted to: First Conference of the European Social Simulation Association, Groningen, The Netherlands.

Indiana Legislative Services Agency; http://www.in.gov/legislative/ic/code/title9/ar21/ch17.html Accessed June 12, 2012.

Institute of Transportation Engineers, ITE (2010), *Manual of Transportation Engineering Studies – Second Edition*. Washington, DC.

Ishaque, M.M., R.B. Noland (2007), "Behavioral Issues in Pedestrian Speed Choice and Street Crossing Behavior A Review." Transport Reviews, Vol. 28, No. 1, January 2008, pp.61-85.

Kaptein, N.A., Theeuwes, J., and van der Horst, R. (1996), Driving simulator validity: some considerations". Transportation Research Record 1550: 30-36.

Kita, H. (1993), *Effects of Merging Lane Length on Merging Behavior at Expressway On-Ramps,* Transportation and Traffic Theory, pp. 37-51. Referenced in Cambridge Systematics, *NGSIM Task E.1-1: Core Algorithm Assessment.* Prepared for Federal Highway Administration. Oakland, California. February 2004.

Kittelson and Associates, Roundabout / Traffic Circle Inventory Database. Available online at: http://roundabouts.kittelson.com/InvMain.asp. Accessed December 31, 2007.

Kondyli, A. and L. Elefteriadou (2012), "Driver Behavior at Freeway-Ramp Merging Areas Based on Instrumented Vehicle Observations", Transportation Letters, Volume 4, Issue 3, July 2012, pp. 129-141.

Lu, L., Ren, G., Wang, W., Yu, C., Yong, Z. (2012), Modeling Bi-direction Pedestrian Flow by Cellular Automata and Complex Networks Theories. Presented at the Transportation Research Boards 91st Annual Meeting, Washington, D.C.

Madanat, S.M et al. (1994), *A Probabilistic Model of Queuing Delay at Stop Controlled Intersection Approaches,* ASCE Journal of Transportation Engineering, Volume 120 (1), pp. 21-36. Referenced in Cambridge Systematics, *NGSIM Task E.1-1: Core Algorithm Assessment.* Prepared for Federal Highway Administration. Oakland, California. February 2004.

Mahmassani, H. and Y. Sheffi (1981), *Using Gap Sequences to Estimate Gap Acceptance Functions,* Transportation Research 15B, pp. 143-148. Referenced in Cambridge Systematics, *NGSIM Task E.1-1: Core Algorithm Assessment.* Prepared for Federal Highway Administration. Oakland, California. February 2004.

Maryland Department of Transportation (MDOT), State Highway Administration; http://www.sha.maryland.gov/Index.aspx?PageId=376 Accessed June 12, 2012.

Moroney, W.F. and Lilenthal, M.G. (2009). Human Factors in Simulation and Training: An Overview. In: P. Hancock, D. Vicenzi, J. Wise and M. Mouloua, Hyman Factors in Simulation and Training. Boca Raton: Traylor & Francis Group, LLC, 3-38.

National Cooperative Highway Research Program (NCHRP), Report 672 (2010), Roundabouts: An Informational Guide – Second Edition.

National Cooperative Highway Research Program (NCHRP), Report 674 (2011), Crossing Solutions at Roundabouts and Channelized Turn Lanes for Pedestrians with Vision Disabilities.

NHTSA (2003), *Pedestrian Roadway Fatalities*, Publication No. # DOT HS 809 456 Technical Report, U.S. Department of Transportation, National Highway traffic Safety Administration, 2003.

NHTSA (2014), *Traffic Safety Facts 2012 Data*, Publication No. # DOT HS 811 888, U.S. Department of Transportation, National Highway Traffic Safety Administration, 2014.

North Carolina Department of Transportation (NCDOT), Laws and Policies; Available at: http://www.ncdot.gov/bikeped/lawspolicies/laws/ Accessed June 12, 2012.

Ottomanelli, M., Iannucci, G., Sassanelli, D., A Simplified Pedestrian-Vehicle Interaction Model at Road Crossings based on Discrete Events System (2012). Presented at the Transportation Research Boards 91st Annual Meeting, Washington, D.C.

Pradhan, A.K., Hammel, K.R., DeRamus, R., Pollatsek, A., Noyce, D.A., Fisher, D.L. (2005), The use of eye movements to evaluate the effects of driver age on risk perception in an advanced driving simulator, Human Factors 47, 840-852.

Rodegerdts, L.; et al, (2007), Roundabouts in the United States. NCHRP Report, Issue 572, 123p.

Rouphail, N. and Eads, B. (1997), Pedestrian Impedance of Turning-Movement Saturation Flow Rates: Comparison of Simulation, Analytical, and Field Observations, Transportation Research Record: Journal of the Transportation Research Board, Volume 1578, pp. 56-63.

Rouphail, N., Hughes, R. and Chae, K. (2005), *Exploratory Simulation of Pedestrian Crossings at Roundabouts.* ASCE Journal of Transportation Engineering. March 2005. pp. 211-218.

Rizzo, M. J., Jermeland, J., Severson, J. (2002), Instrumented Vehicles and Driving Simulators, Gerontechnology; 1(4): 291 - 296.

Rudin-Brown C., Williamson, A., Lenne, M. G. (2009), Can Driving Simulation be Used to Predict Changes in Real-World Crash Risk?, Monash University Accident Research Center, Report No. 299.

SAS Institute Inc. (2007), The SAS System for Windows, Release 9.02. Copyright 1999-2001 by SAS Institute Inc. Cary, NC

Salamati, K., Schroeder, B. J., Geruschat D. R., Rouphail, N., M. (2013), Event-Modeling of Driver Yielding Behavior to Pedestrians at Two-Lane Roundabout Approaches, Submitted for presentation at the 92th Annual Meeting of the Transportation Research Board, January 2013, Washington DC.

Schadschneider, A. (2002), Cellular automaton approach to pedestrian dynamics - theory. In: Schreckenberg, M., Sharma, S.D. (eds.) Pedestrian and Evacuation Dynamics, pp. 75–85. Springer, New York.

Schaefer, L. A., Mackulak, G. T., Cochran, J. K., Cherilla, J. L. (2006), Application of a General Particle System Model to Movement of Pedestrians and Vehicles. WSC, Vol. 2, pp.1155-1160. 1998 Winter Simulation Conference (WSC'98), 1998.Schroeder, B. J., Rouphail, N. M. and Emerson, R. W. (2006), *Exploratory Analysis of Crossing Difficulties for Blind and Sighed Pedestrians at Channelized Turn Lanes*. Journal of the Transportation Research Board, No. 1956. October 2006.

Schroeder, B J., N. M. Rouphail and R. Wall (2006), Exploratory Analysis of Crossing Difficulties for Blind and Sighed Pedestrians at Channelized Turn Lanes. Transportation Research Record: *Journal of the Transportation Research Board*. No. 1956. pp. 94-102. Washington, D.C.

Schroeder, B. J., Rouphail, N. M. (2007), A Framework for Evaluating Pedestrian-Vehicle Interactions at Unsignalized Crossing Facilities in a Microscopic Modeling Environment. Presented at the 86th Annual Meeting of the Transportation Research Board, Washington, DC. 2007.

Schroeder, B. J. (2008), *A Behavior-Based Methodology for Evaluating Pedestrian-Vehicle Interaction* *at Crosswalk*. Doctoral Dissertation in Civil Engineering, North Carolina State University. May 2008.

Schroeder, B. J., Rouphail N. M., Lehan, B. A. (2009), *Observational Study of Pedestrian Behavior Along a Signalized Urban Corridor*. Transportation Research Board 88th Annual Meeting, 23p.

Schroeder, B. J., Rouphail, N. M. (2010), *Mixed-Priority Pedestrian Delay Models at Single-Lane Roundabouts*. Transportation Research Record: Journal of the Transportation Research Board, Issue 2182, 2010, pp. 129-138.

Schroeder, B. J., Rouphail, N. M. (2011), Empirical Behavioral Models to Support Alternative Tools for the Analysis of Mixed-Priority Pedestrian-Vehicle Interaction in a Highway Capacity Context. Procedia-Social and Behavioral Sciences, Volume 16, 2011, pp. 653-663.

Schroeder, B. and N. Rouphail (2011), *Event-Based Modeling of Driver Yielding Behavior at Unsignalized Crosswalks.* ASCE Journal of Transportation Engineering. Volume 137. Issue 7. American Society of Civil Engineers.

Schroeder, B. J.; et al (2011), Crossing Solutions at Roundabouts and Channelized Turn Lanes for Pedestrians with Vision Disabilities. NCHRP Report, Issue 674, 150p.

Schwebel, D.C., Stavrinos, D., Byington, K.W., Davis, T., O'Neal, E.E., de Jong, D., (2012), Distraction and pedestrian safety: How talking on the phone, texting, and listening to music impact crossing the street, Accident Analysis and Prevention, 45 pp.266-271.

Simpson, G., Johnston, L., Richardson, M. (2003), An investigation of road crossing in a virtual environment, Accident Analysis and Prevention 35, pp. 787-796.

Sisiopiku, V.P., D. Akin (2003), "Pedestrian Behaviors at and Perceptions towards Various Pedestrian Facilities: an Examination Based on Observation and Survey Data." Transportation Research Part F 6 (2003) 249-274.

Sun et al. (2002), *Modeling of Motorist-Pedestrian Interaction at Uncontrolled Mid-Block Crosswalks.* Submitted for Publication in Transportation Research Record. TRB 2003 Annual Meeting CD-ROM. Washington, DC.

Sun, D., L. Elefteriadou (2010), "Research and Implementation of Lane Changing Model Based on Driver Behavior" Transportation Research Record: Journal of the Transportation Research Board, No 2161, Transportation Research Board of the National Academies, Washington DC, pp. 1-10.

Sun, D., L. Elefteriadou (2012), "Lane Changing Behavior on Urban Streets: An In-Vehicle Field Experiment Based Study", Computer-Aided Civil and Infrastructure Engineering, Volume 27, Issue 7, August 2012, pp. 525, 542.

Tarawneh, M.S. (2001), "Evaluation of Pedestrian Speeds in Jordan with Investigation of some Contributing Factors." Journal of Safety Research 32, 2001, pp. 229-236.

Teknomo, K. (2006) Application of microscopic pedestrian simulation model, Transportation Research Part F: Traffic Psychology and Behaviour, Volume 9, Issue 1, January 2006, Pages 15-27.

Tennessee Traffic Safety Resource Service; http://www.tntrafficsafety.org/htm/Laws/ped134.htm Accessed June 12, 2012.

Tiwari, G., Bangdiwala S., Saraswat, A. and Garurav S., (2007), *Survival Analysis: Pedestrian Risk Exposure at Signalized Intersection.* Transportation Research Part F 10 (2007) 77–89.

Toyama, M.C., Bazzan, A.L.C., da Silva, R. (2006), An agent-based simulation of pedestrian dynamics: from lane formation to auditorium evacuation. In: AAMAS '06, pp. 108–110. ACM Press, New York.

TRB (1985), National Cooperative Highway Research Program (NCHRP) Report 279. Intersection Channelization Guide. *Transportation Research Board (TRB)*. Washington, DC. November 1985.

TRB (2000), *Highway Capacity Manual (HCM),* Transportation Research Board (TRB), Washington, DC 2000.

TRB (2003), National Cooperative Highway Research Program (NCHRP) Project 3-72. Synthesis on Channelized Right Turns at Intersections on Urban and Suburban Arterials. *Midwest Research Institute*. Kansas City, Missouri.

TRB (2010), *Highway Capacity Manual (HCM),* Transportation Research Board (TRB), Washington, DC 2010.

Troutbeck, Rod J. and Werner Brilon (2002), *Chapter 8: Unsignalized Intersection Theory*, Revised Monograph on Traffic Flow Theory. Turner-Fairbank Highway Research Center. Available at: http://www.tfhrc.gov/its/tft/tft.htm. 2002

Wen, C., Koppelman, F.S. (2001) "The Generalized Nested Logit Model." Transportation Research Part B 35, 627-641.

Yagil, D. (2000), "Beliefs, Motives and Situational Factors Related to Pedestrians' Self-reported Behavior at Signal-controlled Crossings." Transportation Research Part F 3 (2000) 1-13.

Yang, J., Deng, W., Wang, J., Li, Q., Wang, Z. (2006), "Modeling Pedestrians' Road Crossing Behavior in Traffic System Micro-Simulation in China." Transportation Research Part A 40 (2006) 280-290.

Zeedyk, S.M. and Kelly, L. (2003) Behavioral observations of adult-child pairs at pedestrian crossings, Accident Analysis and Prevention 35, 771-776

APPENDIX A: DETAILED LITERATURE REVIEW

Introduction

Pedestrian behavior at unsignalized crossings is a subject that needs to be further researched. The facility type as well as the user type can affect how the pedestrian behaves. Behavioral attributes include pedestrian-vehicle interaction (gap acceptance and driver yielding) and pedestrian crossing traits (path choice, walking speed, and compliance). Current analysis methods for pedestrian behavior include HCM methods and microsimulation. There are also several forms of data collection approaches that should be examined. This paper will be a resource document for generating the models and analytical frameworks for the methodologies to be developed in subsequent research tasks.

Pedestrian Crossing Facility Types and Types of Users

Facility Types

Pedestrian crossings are a common feature at signalized intersections where they are typically tied to the vehicular signal phasing scheme. The analysis of these types of pedestrian crossings is outlined in the HCM. Besides these signalized crossing locations at intersections, there are five major types of unsignalized pedestrian crossings: crossings at channelized right-turn lanes, mid-block pedestrian crossings, crossings at the approaches to a modern roundabout, crossings at two-way stop-controlled (TWSC) intersections, and crossings at all-way stop-controlled (AWSC) intersections. The type of crossing control (stop, yield, and uncontrolled) will be the focus of our research and data collection. The types of intersection configuration will be discussed below as well as the types of crossing control.

Channelized Right-Turn Lanes

Channelized right-turn lanes (CTL) are commonly found at signalized intersections to create additional capacity for heavy right-turning traffic. These single-lane bypass lanes are typically free-flowing with a yield-controlled merge into downstream traffic and may be outfitted with an acceleration and/or a deceleration lane. A pedestrian crossing at the main signalized intersection inevitably requires the pedestrian movement to also cross these CTLs, which is most commonly done at an unsignalized zebra-striped crosswalk in the center of the turn-lane. For a detailed discussion on CTL geometry and alternative placement for the pedestrian crosswalk, refer to NCHRP Report 279 (TRB, 1985) and the NCHRP 3-72 project (TRB, 2003). While not discussed in detail in this document, a previously published paper (Schroeder, Rouphail and Wall Emerson, 2006) presents a detailed analysis comparing the crossing abilities of blind and sighted travelers at these types of facilities.

Mid-Block Crossings

In addition to crossings at signalized intersections, pedestrian crossings are commonly found at mid-block locations. Contrary to what is implied in the terminology, these crossings are not necessarily located in the middle of a block, but rather can be found anywhere along a roadway at locations away from an intersection crossing. The roadway can range anywhere from one to four or more lanes and may or may not be outfitted with a signal. The decision to place a signal at a mid-block location is regulated by the pedestrian signal warrant in the MUTCD (FHWA 2003).

Pedestrian flows would typically be modeled compliant with both crosswalk location and signal phasing. However, research performed on Hillsborough Street in Raleigh, NC suggests that a significant percentage of pedestrians will cross outside the intended crosswalks and against a 'Don't Walk' indication at a signal (Schroeder, et. al., 2009). A gap acceptance model is required to describe this behavior in simulation and a compliance algorithm would be used to decide when the gap acceptance logic is applied. The data suggest that the frequency of jaywalking behavior is higher at longer midblock segments and further increases during class breaks due to elevated pedestrian activity. The impact of pedestrians on the operations of traffic signals is reduced, as was made evident by skipped WALK phases, with some percentage of pedestrians crossing at midblock. The arrival distributions of pedestrians could no longer be viewed as a random event during class breaks which showed a higher concentration of pedestrian flows. The issues of *path choice* and *time sensitivity* seem to be major contributing factors, where jaywalking behavior is increased along long midblock segments, high (anticipated) wait times, and during class changes.

Roundabout Crossings

Pedestrian crossings are also found at modern roundabouts, which are becoming an increasingly popular traffic control feature in the US. A long-term staple in Europe and Australia, an online database (Kittelson Associates, 2007) now lists more than 1,000 roundabout intersections across the United States justifying their inclusion in this discussion. The pedestrian crossing at modern roundabouts is typically a two-stage crossing with pedestrians being able to find refuge on the splitter island as shown in Figure 32 below. The pedestrian crossing location is also designed to allow storage for one or more vehicles waiting to enter the roundabout downstream of the crosswalk.

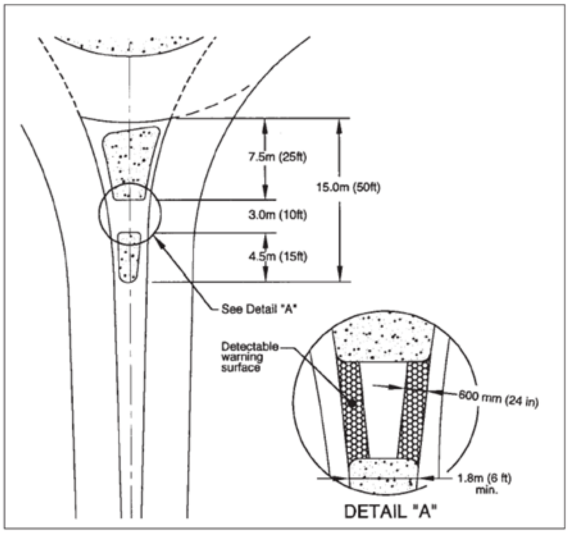


Figure 32: Roundabout Pedestrian Crossing

(SOURCE: FHWA Roundabout Guide, 2000)

For the purpose of discussion it is assumed that the base condition for pedestrian crossings at any of the three types of locations is a zebra-striped unsignalized crosswalk. At these types of crossings, legislation typically gives pedestrians the right-of-way, but motorist compliance varies. To further enhance the crossing and to make it safer for pedestrians, there are several categories of pedestrian crossing treatments that aim to facilitate pedestrian crossings.

Two-Way Stop-Controlled (TWSC)

TWSC intersections are common in the US and have two typical configurations according to the HCM. The first is a four-leg intersection where the major street is uncontrolled and the minor street is controlled by stop signs (TRB, 2010). The second configuration is a three-leg intersection where the single minor-street approach is controlled by a stop sign. The approaches for minor streets can be public streets or private driveways.

At the major crossing of a TWSC intersection, pedestrians must wait for an acceptable gap before crossing. At the minor crossing, pedestrians can cross while a vehicle is stopped at the stop sign or cross if there are no vehicles on the minor approach. Turning movements from the major approach may also affect a pedestrian's ability to cross. Thus, the major crossing for a TWSC intersection is similar to a midblock crossing and the minor crossing is like an AWSC crossing, with the addition of turn conflicts from the major street.

All-Way Stop-Controlled (AWSC)

Every vehicle is required to stop before proceeding at AWSC intersections. Since each driver must stop, the decision to proceed into the intersection is a function of traffic conditions on the other approaches. A driver can proceed immediately after stopping if no other traffic is present at the other approaches. If there is traffic on one or more of the other approaches, a driver proceeds only after determining that no vehicles are currently in the intersection and that it is the driver's turn to proceed (TRB, 2010). Pedestrians at this type of intersection may cross in front of stopped vehicles, but must also be alert of turning vehicles from other approaches.

Crossing Control Types

There are three categories of crossing control: stop, yield, and uncontrolled. For stop-controlled approaches, vehicles and pedestrians must stop and observe their surroundings before continuing through the intersection. AWSC intersections and the minor approach of TWSC intersections are stop-controlled. For yield-controlled approaches, vehicles can continue on their chosen path if it is unobstructed. Channelized right-turn lanes are yield-controlled because the vehicles coming from the channelized lane must yield to any vehicles that are already in major road. Roundabouts are also yield-controlled. Drivers at the roundabout approaches must yield to other drivers who are in the circle. For uncontrolled intersections, drivers can continue through the intersection unheeded, though they should still be aware of any turning vehicles in the intersection. Midblock intersections and the major approach of TWSC intersections are uncontrolled. Uncontrolled intersections are likely to cause the most difficulty for pedestrians.

User Types

For the assessment of pedestrian behavior at road crossings, the *heterogeneous* nature of the pedestrian population needs to be taken into consideration. While gap acceptance for drivers is strongly linked to the acceleration capability of the vehicle, pedestrian decisions are a function of individual attributes. A typical population includes students, elderly, blind pedestrians, children, and people with baby strollers. There are drastic differences in the ability and the willingness to make a crossing decision among these sub-groups. While pedestrian activity in many suburban environments is scarce, roadways in proximity to universities see frequent pedestrian crossings at both signals and mid-block locations. A lot of recent research has focused on gap acceptance by blind pedestrians. Ashmead et al. (2005) found that when attempting to cross at a two-lane roundabout, blind pedestrians waited three times longer than sighted pedestrians and furthermore made about 6% 'risky' decisions. Sighted pedestrians didn't make any. In another example, Sun et al. (2002) found from data at an unsignalized mid-block pedestrian crossing that both the minimum accepted gap time and the average accepted gap were lower for younger than for older pedestrians.

In a study by Avineri, it was found that pedestrian crossing behavior is affected by age and fear of falling within the analysis of head pitches (reflecting attention to traffic) and crossing speed. In an observation, 203 pedestrians were video recorded and later surveyed in order to find the impact of age and FOF. Pedestrian crossing behavior is random, but age and FOF make a difference on speed and head pitch. A limitation of this study was that components were limited and additional characteristics of pedestrians should be considered. More samples should also be observed. The surveys showed that the result of age and FOF is accurate (Avineri, et. al., 2012).

Levels of experience for different age groups greatly affect pedestrian behavior. Young children are typically short, risky, inexperienced, have limited peripheral vision and have trouble locating sound sources (Fitzpatrick, 2006). Preteens have increased physical abilities, but are still risky and inexperienced. High School Age children are typically more physically fit, but they also lack experience and training and can feel invincible. Senior Adults have likely lost some of their physical abilities and thus may overestimate their abilities. Those with disabilities, permanent or temporary, may have restricted mobility. People new to the urban experience may be used to different motorist behavior and may try to cross unsafely and put themselves at risk (FDOT, 1999). Pedestrian inattention due to distraction from multimedia device usage, such as cell phones, can also have an effect on pedestrian behavior.

Pedestrian Behavioral Attributes

This section describes the behavioral components in the interaction of pedestrians and vehicles at unsignalized crosswalks: pedestrian gap acceptance and driver yielding behavior. Path choice, walking speed, and compliance of pedestrians will also be discussed. Some discussion on signalized intersections has also been included when it would be useful for comparison.

Pedestrian Crossing Attributes

The pedestrian population arriving at the crosswalk needs to be treated differently from a vehicle population, as pedestrian movements are generally much less constrained than vehicles traveling within a travel lane (Blue and Adler, 2000). They concluded that pedestrians are not officially channelized, can vary their speed, can occupy any part of the walkway, can bump into each other, and have almost instantaneous acceleration/deceleration profiles. These attributes have clear implications for microscopic analysis and modeling of pedestrians, as the assumption that pedestrians simply operate like "small vehicles" is not valid. The crossing attributes for pedestrians are divided in this review into pedestrian path choice, pedestrian walking speed, and pedestrian compliance with intersection control.

Pedestrian Path Choice

The most important factor for utilizing a crossing for a pedestrian depends on whether it is located between the point of their origin and their destination (Sisiopiku and Akin 2003). Path choices and the strategic location of pedestrian crossings also improve compliance by pedestrians, which are recognized by guidebooks in the pedestrian safety field (Harkey and Zegeer, 2004).

Several researchers attempted to understand pedestrian path choices. Chu (2002) developed a nested logit model based on theoretical expectations of how pedestrians cross roads in urban settings. The model contains variables descriptive of the street environment including *continuous variables* (such as roadside walking distance, crossing distance, and traffic volume) and *discrete characteristics* (such as the presence of marked crosswalks, traffic signals, and pedestrian signals).

Chu (2002) found that people are more likely to cross at an intersection with a traffic signal or a pedestrian signal head (Walk/Don't Walk signs). Also, the likelihood of people crossing at any location with a marked crosswalk is higher than those without. However, depending on the type of facility, the influence of these discrete characteristics can vary. The results of Chu's work (2002) show that the presence of a marked crosswalk is more influential at an intersection than at a midblock location. The most influential factors for crossing at an intersection are pedestrian signals, marked crosswalks, and traffic signals. An increase in any continuous variable for a given option will result in a decrease in the probability of that option being chosen (i.e., the further a pedestrian has to walk to use a particular crossing option, the less likely it is that the pedestrian will choose that option). The magnitude of the decrease varies across these continuous variables and across options (Chu, 2002).

Simulation of pedestrian path choice and other behaviors can be divided into two categories, one is a macroscopic level and the other is a microscopic level which involves individual units with individual traffic characteristics (Teknomo, 2006). With a more detailed approach, a microscopic pedestrian flow model simulates the walking behavior of each pedestrian, as well as the personal space, speed and interactions between pedestrians and the surrounding environment. It is said that real pedestrians influence the behaviors of each other either with mutual or reciprocal action, meaning that they will avoid or overtake others in order to achieve their own goals, such as a steady pace or enough individual space.

The interaction between pedestrians makes a difference on pedestrian behaviors and movement. Some traffic strategies on pedestrian flow and behavior are required because efficiency and safety will decrease on the basis of pedestrian minimum-interaction behaviors without any traffic treatment. A lane-formation approach is demonstrated as a good way to organize pedestrian flow to improve efficiency without increasing walking space for pedestrians. Pedestrians under the condition of traffic control treatment (i.e. lane-formation) tend to follow others rather than make their own walking path (Teknomo, 2006).

Pedestrian Walking Speed

Pedestrian walking speeds may vary for several reasons. For example, the 2000 Highway Capacity Manual (TRB, 2000) chapter on pedestrians recommends a walking speed of 4.0 ft/sec, with a lowered speed of 3.3 ft/sec if the fraction of elderly pedestrians exceeds 20% and a further reduction by 0.3ft/sec at upgrades exceeding 10%. Bennett et al. (2001) investigated pedestrian walking speeds at signalized intersections and mid-block crossings and found slower average speeds at the mid-block locations. The authors also found differences between the 15^{th} and 85^{th} percentile walking speeds of about 2.5 ft/sec and significant variation between pedestrians with and without walking difficulty at all studied locations.

Fitzpatrick et al. (2006) recommended in a more recent NCHRP report to lower the pedestrian walking speed used by the MUTCD from 4.0 ft/s to 3.5 ft/s, but further acknowledged that even lower speeds may be appropriate in some cases, such as where pedestrians typically walk more slowly or use wheelchairs in the crosswalk. The variability of walking speed is important when discussing crossing behavior, because it is directly proportional to the time required to cross a given distance. The research found that as many as one third of pedestrians travel at a slower pace, specifically children, older pedestrians, and persons with disabilities. The mean start-up time (from the start of the Walk signal to the moment the pedestrian steps off the curb and starts to cross) was 2.5 s for older pedestrians, compared with 1.9 s for younger ones (Fitzpatrick, 2006). The walking speeds for pedestrians with physical disabilities are also lower than the average walking speed assumed for the design of pedestrian crosswalk signal timing. These average speeds range from 1.97 ft/s for an above knee amputee to 3.55 ft/s for those using a wheelchair. Weather conditions can also affect walking speeds. The presence of snow, ice, or slush on sidewalks and roads leads to ill-defined curbs, hidden potholes and obstacles, greater amounts of glare and visual difficulties, and a greater chance of a slip or fall by a pedestrian, especially for an older person (Dewar, 2002).

Tarawneh (2001) found through a study in Jordan that pedestrians tended to walk faster after longer wait times and also that pedestrian walking speed varies by age, gender, distance and group size. A walking speed of 1.11 m/s (3.6 ft/s) was suggested with males, small groups, and those aged 21-30 walking faster than females, groups of three or more, and other age groups.

Pedestrian Compliance

The analysis of the interaction between pedestrians and drivers is complicated by the lack of a clear understanding of right-of-way legislation at unsignalized locations. While many states have legislation in place requiring vehicles to yield to pedestrians in the crosswalk, field observations on busier streets quickly make it evident that compliance varies. A study of six two-lane roundabout approaches in four states showed that although all these states require drivers to yield to pedestrians within the crosswalk the yielding is not perfect (Salamati et al. 2013). In a sample of state right-of-way laws applicable to the study (North Carolina, Maryland, Indiana and Tennessee), it is evident that drivers should yield the right of way to pedestrians *in* the crosswalk (NCDOT 2012, MDOT 2012, Tennessee Traffic Safety Resource Services Agency 2012, and Indiana Legislative Services Agency 2012). However, the study shows that the yielding rate varies from 0% to 85% at two-lane roundabout approaches (Salamati et al. 2013).

More commonly, drivers and pedestrians use methods of non-verbal communication to determine crossing priority. The willingness of a driver to yield and the assertiveness with which a pedestrian seizes the crosswalk are two of many factors that may influence this interaction. Other factors may include the cross-section of the road, the type of crossing treatment or the general level of congestion at the crossing location.

Studies of pedestrian behavior show that not everyone complies with traffic laws. The enforcement of pedestrian compliance with traffic signals is rarely emphasized relative to drivers being held accountable for having to stop at a red traffic signal (for example through automatic enforcement). Ishaque and Noland (2008) found that crossing speed, gap acceptance and signal compliance with relation to age and gender are among the factors influencing the crossing behavior of pedestrians. Perception of one's own safety is one of the internal factors identified by Yagil (2000) to influence pedestrian compliance. Yagil also found that a group of pedestrians behaves differently than individuals. Pedestrians were more likely to wait at an intersection if encountering a group of pedestrians already waiting at the crosswalk (Yagil, 2000).

Several researchers studied changes in pedestrian behavior by increasing pedestrian wait time at the intersection. Dunn and Petty (1984) found that pedestrians at midblock crossings who have been waiting for 30 or more seconds showed more risky behavior. A study of pedestrian risk exposure at signalized intersections in India by Tiwari et al. (2007) showed similar results. One of the characteristics of risky behaviors and non-compliance at the crosswalk is accepting shorter gaps in traffic.

However, Sun et al. (2002) found that pedestrians who continue to wait at the crosswalk for a long time are still careful in nature and would not accept shorter gaps. The authors found an increase in the average accepted gaps as the waiting time (delay) increases.

Accordingly, the HCM predicts an increasing likelihood of non-compliance as pedestrian delay increases (TRB, 2010). Therefore, non-compliance can be translated into adjusting the critical gap to a lower value as the pedestrian delay or waiting time increases. This phenomenon should be considered for developing gap acceptance models for pedestrians. Pedestrian delay in the HCM at signalized intersections is a function of signal timing parameters and assumes all pedestrians comply with the signal phase (TRB, 2010).

For unsignalized crossings, the Highway Capacity Manual offers a methodology for estimating delay based on gap acceptance characteristics that could feasibly be applied to non-compliant pedestrians at signals. Guo et al. (2004) created a model for pedestrian delay of non-compliant pedestrians at signalized crossings by taking into account the effect of traffic platooning resulting from upstream signals. The authors found an increase in delay for pedestrians who would otherwise cross illegally, while having little effect on compliant pedestrians.

Pedestrian-Vehicle Interaction

Current traffic engineering analysis tools and capacity models are of limited use for evaluating the interaction of pedestrians and vehicles at unsignalized crossing facilities. The analysis methodologies for unsignalized intersections in the HCM are traditionally limited to boundary cases, which assume strictly enforced right-of-way rules (TRB, 2000). These assumptions mean that pedestrian operations are analyzed by either assuming pedestrian priority (100% driver yielding) or vehicle priority without yielding right-of-way to pedestrians (Schroeder and Rouphail, 2011). The more complex interaction of the two modes in which some drivers yield to pedestrians and some pedestrians accept gaps in traffic is typically ignored in traditional HCM methods. This type of interaction was previously referred to as a *mixed-priority* crossing (Schroeder and Rouphail, 2010). Changes in the 2010 HCM (TRB, 2010) have made an attempt at combining pedestrian gap acceptance and driver yielding behavior for pedestrian delay analysis, but the revised methodology is not based on empirical observations and has not been calibrated by field observations. In practice, alternative analysis tools in the form of microscopic simulations are frequently used to help overcome some of the limitations of the HCM.

Vehicle-pedestrian interaction can be characterized by four interaction processes that can be expressed in the form of four probability terms, following a framework for evaluating unsignalized pedestrian crossings in a simulation context (Schroeder and Rouphail, 2007):

P(G) - The probability of a gap occurring in the traffic stream

P(GU) - The probability of a gap being utilized by the pedestrian

P(Y) - The probability of a driver yielding

P(YU) - The probability that a yield is utilized by the pedestrian

The probability of gap occurrence, P(G), is a function of vehicle arrivals and the headway distribution in the traffic stream. The behavioral characteristics of pedestrians and drivers are generally described by the probability of crossing in a gap, P(GU), and the probability of a driver yielding to a waiting pedestrian, P(Y). The fourth parameter, P(YU), has been observed as an important crossing attribute for pedestrians with vision impairments, and may also be applicable to other pedestrian populations, who tend to reject or miss a portion of the encountered yields. The focus of this review is on the second and third component of the aforementioned framework: pedestrian gap acceptance and driver yielding behavior.

Pedestrian Gap Acceptance

Pedestrian crossing behavior has not been explored to the same degree that vehicle gap acceptance has been investigated. While similar in concept, there are a variety of pedestrian characteristics and caveats in the interaction between the pedestrian and vehicle modes that give reason to derive separate pedestrian gap acceptance models. This section provides an overview of general vehicle gap acceptance models, followed by a review of reasons why pedestrians are believed to behave differently when making a decision to cross the roadway and summarizes existing research.

Overview of Vehicle Gap Acceptance Models

Traditionally, literature on vehicle gap acceptance has used a constant value of *critical gap* (CG) that is calibrated for local conditions (Troutbeck and Brilon, 2002). It can differ depending on the type of movement and the type of vehicle. For example, the CG for left turns is likely to be larger than for right turns, and heavy vehicles tend to have longer CGs, because of slower acceleration profiles and longer vehicle lengths. In the following, this type of gap acceptance model will be referred to as the *deterministic model* for gap acceptance.

By definition, the critical gap is the time between consecutive vehicles on the major road at which a vehicle waiting at the minor approach is equally likely to accept the gap or reject it. Literature on gap acceptance oftentimes assumes that drivers are both *homogeneous* and *consistent*. In a homogeneous driver population, all drivers have the same critical gap. Under the consistency assumption, the same gap acceptance situation will always cause a driver to make the same (consistent) decision. Although these assumptions are not realistic, Troutbeck and Brilon (2002) justify their use because inconsistencies in driver behavior tend to increase capacity while a heterogeneous driver population will decrease capacity, thereby offsetting the previous effect.

The most common US application of deterministic gap acceptance is in the US Highway Capacity Manual (TRB, 2010). The manual recommends using a constant critical gap from listed default parameters or locally estimating CGs from field conditions. It further recommends a reduction of its tabulated CG values for heavily populated regions (greater than 250,000), suggesting that drivers in those regions may be more likely to encounter frequent congestion and have thus lowered their CG threshold.

There are several ways for estimating CG from field data, including a graphical method (Troutbeck and Brilon, 2002), a regression method (Troutbeck and Brilon, 2002), a statistical method based on maximum likelihood estimation (Troutbeck, 2001), and the Ramsey-Routledge method (ITE, 2010). In application of these methods, the capacity of the minor street flow becomes a function of the CG on the minor approach t_c, the follow-up time on the minor approach t_f and the conflicting major street flow q_p as shown in an HCM2010 equation adopted below:

Equation 33: HCM Capacity Equation for Two-Way Stop Controlled Intersection (TRB 2010)

$$Capacity = \frac{3{,}600 \times q_p \times e^{-q_p t_c}}{1 - e^{-q_p t_f}}$$

[vehicles/hour]

The follow-up time describes the time needed for additional vehicles in a stored queue to accept the same gap. The size of t_f is typically less than t_c, because some of the decision and acceleration times for subsequent vehicles occur during the initial gap.

In addition to deterministic gap acceptance, a report compiled for the Federal Highway Administration (FHWA) Next Generation Microsimulation (NGSIM) research effort (Cambridge Systematics, 2004) discusses probabilistic gap acceptance models, for which the driver response for an identical event (same speed, same gap in conflicting traffic) can be drawn from a probabilistic distribution of possible responses. Such probit models assume a mean CG with a random variance term depending on the specific coefficients defined for a driver and/or situation. Conceptually, probit models could represent inconsistent driver behavior and a heterogeneous population by drawing gap acceptance decisions from random distributions.

Alternatively, probabilistic behavior can be modeled in the form of a binary or multinomial logit model. A logit model could describe the likelihood of gap acceptance as a function of a number of different parameters (for example assertive vs. non-assertive pedestrians, gap time, and type of the arriving vehicle). It thus introduces greater complexity in the gap acceptance model, but in turn requires a lot of data for calibration. Logit Gap Acceptance Models have been proposed by Ben-Akiva and Lerman (1985) and Cassidy (1995) and Probit Models were suggested by Mahmassani and Sheffi (1981) and Madanat (1994).

Some researchers have proposed even more complex algorithms for modeling gap acceptance. Kita (1993) used neural networks to describe the process, under the assumption that gap acceptance is not a linear sequence of events, but that multiple factors affect the decision making process. This modeling approach is capable of removing consistency assumptions, but the authors upheld the assumption of homogeneity.

Models for Pedestrian Gap Acceptance

Above discussion suggests that pedestrian movements, pedestrian gap acceptance, and pedestrian-vehicle interaction are different enough from conventional vehicular traffic to warrant alternate models for pedestrian movements, gap acceptance, and capacity. These differences are discussed below.

Underlying assumptions of pedestrian movements must be revisited before adopting gap acceptance concepts. Queued vehicles on a single-lane approach are subject to a first-in-first-out (FIFO) priority, but multiple pedestrians can generally accept the same gap simultaneously (Blue and Adler 2000). Pedestrian gap acceptance is arguably unaffected by the concept of follow-up time. The critical gap may be similarly estimated for vehicles and pedestrians, but the use of the gaps are different for pedestrians. Also, the critical gap is a function of pedestrian speed, so it may vary more significantly than that of vehicles for the same movement. The deterministic gap acceptance model in the HCM2010 offers a method for estimating pedestrian critical gap t_c as a function of crosswalk length L, Pedestrian Walking Speed S_p and pedestrian start-up time t_s (Equation 34).

Equation 34: Pedestrian Critical Gap after HCM (TRB 2010)

$$t_c = \frac{L}{S_p} + t_s$$

Rouphail et al. (2005) described pedestrian gap acceptance as the sum of latency and actual crossing times, an approach similar to the HCM method discussed above. The authors used field estimates of the median latency time in place of the HCM start-up time. The authors' research compared latency times of blind and sighted pedestrians and found that blind pedestrians exhibited significantly larger latency times, resulting in longer critical gap values and presumably more delay. The increased delay to blind pedestrians is consistent with research findings presented above.

Researchers have also attempted to use advanced gap acceptance models to describe pedestrian crossings. Sun et al. (2002) calibrated probit and binary logit models to describe both pedestrian gap acceptance and driver yielding from actual field data. The authors excluded about 25% of observations for later model validation and found that binary logit models performed best in both cases, correctly predicting 85.6% of gap acceptance and 87.1% of yielding decisions. For comparison, a probit model only resulted in 68.5% correctly predicted gap acceptance decisions, and a deterministic critical gap model actually achieved a surprising 81.5% correct predictions. Regression analysis found the important factors for pedestrian gap acceptance to be gap size, number of pedestrians waiting, and age of pedestrians. The authors recommended the binary-logit model for estimation, stating that the good performance of the deterministic model was likely due to an extraordinarily homogeneous pedestrian population.

From observations in China, Yang et al. (2006) derived a pedestrian gap acceptance formulation for the critical gap (CG) of pedestrians. This equation is shown below, where L is the length of the crossing, S is the walking speed and F is a factor of safety based on the pedestrian's confidence.

Equation 35: Critical Gap (CG) for Pedestrians

$$CG = L/S + F$$

Similar assumptions for pedestrian gap acceptance were used in the analysis of unsignalized pedestrian crossings at roundabouts and channelized right turn lanes by Rouphail et al. (2005) and Schroeder et al. (2006), respectively. Schroeder (2008) developed logistic regression-based gap acceptance models for unsignalized crossings to better describe the process of pedestrian gap acceptance by incorporating vehicle dynamics, pedestrian characteristics and concurrent events at the crosswalk.

Crossing speed, and thus critical gap, can vary widely for pedestrians, based on a variety of reasons. When interpreting the consistency assumption for gap acceptance described above, it is intuitive that pedestrians will tend to alter their gap acceptance attributes if they are in a hurry versus if they are on a leisure trip, for example. Documentation of this behavior was provided in the section on pedestrian compliance above. The consistency assumption then is violated, because a similar situation of vehicular gap and speed at a given geometry will result in different decisions by the pedestrian, depending on his/her state of mind.

Pedestrian Follow-Up Time

The HCM equation referenced above uses the critical gap and follow-up time to calculate minor-street capacity as a function of major street flow (TRB, 2010). While a pedestrian critical gap can be observed from field data, the follow-up time concept proves challenging.

Above reference to a lack of channelization for pedestrian traffic (Blue and Adler 2000) means that pedestrians are not confined to sequential queue storage like vehicles, but can occupy spaces next to each other in the waiting area. In fact, the HCM offers equations for analyzing pedestrian storage space at the crosswalk (TRB, 2010). Therefore the concept of follow-up time is not applicable for pedestrians in the same fashion as for vehicles. For pedestrians, it is possible that for example three pedestrians cross at the exact time, depending on the width of the crossing. In this case then, the classical follow-up time wouldn't apply until the 4^{th} pedestrian, who had to wait behind the other three.

Driver Yielding Behavior

Driver yielding behavior has been linked in research to operational characteristics such as vehicle speeds (Geruschat and Hassan 2005), as well as geometric characteristics of the crosswalk location, for example; entry versus exit leg at a roundabout (Rodegerdts, 2007, Ashmead et al., 2005). But to date, these isolated studies of driver yielding behavior at unsignalized crosswalks have largely been *descriptive*, with little insight gained towards *predicting* driver yielding at such crosswalks. The rate of driver yielding to pedestrians at unsignalized crosswalks varies across locations (Rodegerdts, 2007), but in nearly all cases is less than 100%. A range of treatments exist that are intended to increase the rate of driver yielding (Fitzpatrick, 2006).

Findings of pedestrian operations at roundabouts by NCHRP report 572 (Rodegerdts, 2007) show that 43% of the drivers at two-lane approaches of the roundabout do not yield to pedestrians. The lack of yielding is only 17% for single-lane roundabouts. Lack of yielding is also higher at exit (54% not yielding) compared to 33% not yielding at the entry. Based on these findings, the number of lanes and crosswalk location (entry or exit) are the two design elements that affect pedestrian accessibility at roundabouts.

A study by Salamati et al. (2012) at six two-lane roundabout approaches across the country showed that the yielding rate varies from 0% to 85% at the exit and entry leg of roundabouts depending on pedestrian assertiveness to cross the street, pedestrian disability (blind or sighted), entry or exit leg of the roundabout and the study location.

Predictive Yielding Models

In previous research, Sun et al (2002) collected data on driver yielding and pedestrian gap acceptance at an unsignalized midblock pedestrian crossing and compared the fit of different statistical models. The authors estimated yielding probabilities based on the discrete parameters of driver gender, driver age, type of vehicle, number of pedestrians, and the presence of an opposing yield. They found that drivers are more likely to yield to a group of pedestrians and that older drivers were more likely to yield than younger drivers. Their results showed that a logistics modeling approach outperformed a probit model for driver yielding, as well as for pedestrian gap acceptance. The authors looked at only one crosswalk and did not analyze any pedestrian treatment effects. The authors collected 1.5 hours of each AM and PM peak data over 5 days, for a total of 15 hours of data. The resulting samples included 687 accepted gap, 938 rejected gap and 1254 motorist yield data points, which was sufficient to allow them to estimate statistically significant probit and logit models.

The research findings above can be summarized in that the decision of a driver to yield is a function of both operational and behavioral parameters. In the first category, the yield decision is triggered by both the speed of the vehicle and the assertiveness of the pedestrian. In the behavioral category, drivers are influenced by the number of pedestrians at the crosswalk and the clothing worn by the pedestrian(s) (more willing to yield for a brightly clothed pedestrian than for a drably clothed pedestrian). Similarly, it can be hypothesized that yielding is impacted by the presence of a conflict downstream of the crosswalk. There are also cases where a driver may be forced to yield, because of a pedestrian GO decision in a too-short gap in traffic.

Contributing Factors to Driver Yielding

Conceptually, the probability of a driver to yield when a pedestrian is present at the crosswalk, P(Y), can be expressed as a function of independent parameters β_i; much like it would be done in multi-linear regression analysis. Similarly, the probability of a pedestrian decision to cross the road is a function of some variables. Through statistical modeling, these parameters can be related to the response variable as demonstrated by Sun et al. (2000) and others. Due to the discrete nature of the process (1/0 = Y/NY or GO/NoGO) methods of categorical data analysis need to be applied. The decision of a driver to yield or of the pedestrian to GO should be a function of the following types of variables:

- Vehicle Dynamics: A yield is only feasible if a driver can reasonably come to a complete stop (hard yield) or delay his arrival at the crosswalk enough to allow for the pedestrian to cross (rolling yield). Parameters in this category include travel speeds, distance from the conflict area, and maximum (comfortable) deceleration rates for both drivers and pedestrians. These variables affect the arrival time of the vehicle and thereby also the pedestrian decision.
- Vehicle and Driver Characteristics: These attributes describe things such as the 'willingness' of drivers to yield, driver courtesy, and the type of vehicle.
- Pedestrian Characteristics: Pedestrian attributes include assertiveness, the presence of multiple pedestrians, or the willingness to accept risk.
- Confounding Factors: In addition to the personal attributes above, the circumstances surrounding the interaction may impact the decision-making process. Examples include the presence of a downstream queue/congestion after the crosswalk or a yield event in the opposing direction or adjacent travel lane. Yielding behavior is also intuitively related to whether or not a vehicle is traveling in a platoon of vehicles. Similarly, pedestrians may be more willing to accept a shorter gap for an individual vehicle than a platoon.

A study at two two-lane roundabouts in Maryland shows that the speed of the vehicle at the entry and exit of roundabouts significantly influences the driver yielding rate (Geruschat and Hassan, 2005). The authors found that 65% of the variability of driver yielding rate can be explained by speed. Based on their study, they estimated that the driver yielding rate with speeds lower than 15 miles per hour is about 75% as opposed to 50% with speeds higher than 20 miles per hour. In addition to the speed of the vehicle, they found that drivers yield to pedestrians at the entry 79% of the time as opposed to only 37% of the time at exit. The authors observed that pedestrian behavior influences the driver yielding rate. For example, the further the pedestrian is standing in the crosswalk the higher the likelihood of driver yielding. Drivers are almost twice as likely to yield to a pedestrian standing one foot in the crosswalk as one foot from the curb. Their study also shows that drivers have significantly higher yielding rates to pedestrians carrying a white cane (2.1 times higher at entry and 4.4 times higher at exit).

A study by Salamati et al. (2012) showed that the likelihood of driver yielding at the entry leg of roundabouts is higher than exit. Drivers tend to yield to pedestrians with white canes more often than sighted pedestrians. Drivers traveling in the far lane, relative to pedestrian location, have lower probability of yielding to a pedestrian than drivers in the near lane. As the speed increases the probability of driver yielding decreases. At the exit leg of the roundabout, drivers turning right from the adjacent lane have lower propensity of yielding than drivers coming from other directions. The results show that factors such as vehicle platooning, downstream conflict (for only entry leg of the roundabout) and pedestrian waiting position, "at curb" versus "in crosswalk," do not have a significant impact (with 95% confidence level) on the probability of a driver yielding to pedestrian.

Summary of Behavioral Attributes

This section described various aspects of pedestrian behavior and their interaction with vehicular traffic that should be considered in the development of microscopic algorithms for modeling pedestrians in simulation. In particular, pedestrian crossing behavior was discussed in terms of pedestrian path choice, pedestrian walking speed, and pedestrian compliance. The section further discussed aspects of pedestrian-vehicle interaction with an emphasis on pedestrian gap acceptance and driver yielding behavior. In addition, it should be noted that pedestrian behavior could be different at multimodal facilities. The presence of transit alight maneuvers and bike traffic may make pedestrian crossing behavior much more complex.

Table 21 provides a summary of the literature as it relates to pedestrian behavioral attributes and what studies have explored these attributes for the various facility types. Several gaps in the literature were discovered that remain to be addressed. Research sources involving TWSC and AWSC intersections were not found in literature. Literature on pedestrian walking speed and path choice were only found for midblock locations. There is existing literature on compliance at single- and two-lane roundabouts, as well as midblock crossings, but not for channelized turn lanes.

Table 21: Summary of Existing Literature on Facility Types

Type of Studies	**Facility Type**					
	Single-Lane Roundabout	**Two-Lane Roundabout**	**Channelized Turn Lane**	**Midblock Crosswalk**	**Two-way Stop-Controlled Intersection**	**All-way Stop-Controlled Intersection**
Pedestrian Gap Acceptance	Ashmead et al. (2005), Schroeder and Rouphail (2011), Rouphail et al. (2005)	Ashmead et al. (2005), NCHRP 674	Schroeder, Rouphail, and Wall Emerson (2006)	Schroeder and Rouphail (2011), Wang et al (2010), Ottomanelli et al (2011), Hunt et al. (2011), Schroeder (2008), Schroeder and Rouphail (2011)		
Driver Yielding	NCHRP 672, NCHRP 674	NCHRP 674, Geruschat and Hassan (2005), Salamati et al. (2013)	NCHRP 674	NCHRP 674, Schroeder and Rouphail (2011), Fitzpatrick et al. (2006), Schroeder (2008), Sun et al. (2002)		
Pedestrian Walking Speed				Fitzpatrick et al. (2006), Ishaque and Noland (2007), Tarawneh (2001)		
Pedestrian Path Choice				Chu et al. (2002),		

				Dunn and Petty (1984), Ishaque and Noland (2008), Wang et al (2010), Schroeder et al. (2009), Sisiopiku and Akin (2003)		
Compliance (Pedestrian Delay)	Schroeder and Rouphail (2010)	RCOC (2011)		Havard and Willis (2011), Hunt el al. (2011), Chu et al. (2002), Schroeder et al. (2009)		

Pedestrian Stream Models

Early research on pedestrian movements focused on the pedestrians that did not interact with other modes of transportation. Modeling their movements required different approaches from methods that are applied to vehicles. There are three typical approaches to modeling pedestrians without consideration of vehicle interactions, namely discrete cellular automata, continuous force-based models, and macroscopic pedestrian stream models.

The first is a discrete cellular automata approach, which segments the walkway into individual cells and assigns pedestrians to specific cells. Pedestrians move forward/backwards, left/right, or diagonally within the walkway in order to reach their destination. Conditional bounds are set so that pedestrians do not share the same cell. Lu et al. (2012) were able to replicate transient channelized flow that is observed in two-way walkways. Cellular automata models tend to require large amounts of memory when modeling larger areas (Blue and Adler, 200l, and Schadschneider, 2002).

A second set of models are continuous force-based models of pedestrian movement. Pedestrians are motivated by social forces, including location, velocity, mass, direction, and repulsive forces of other pedestrians. These models are much more flexible in the types of conditions that can be modeled as obstacles of any size and shape which can be placed within the simulated walking space. This flexibility allows for these models to replicate complex scenarios, such as the evacuation of a building during an emergency. These models have limited application to pedestrian-vehicle interaction, but they may be relevant in replicating the decision-making process of each pedestrian in deciding when to cross as a function of the actions of other pedestrians in the vicinity. Force-based models tend to require large amounts of processing power when modeling larger areas and many pedestrians (Helbing and Molnár 1995, Chraibi et. al. 2009, Teknomo 2006, Toyama 2006).

Finally, macroscopic pedestrian stream models are used that resemble typical traffic stream models. These models utilize parameters such as pedestrian space, speed, and flow per width. Effective width is used when determining the level of service for pedestrians, with common obstructions included that limit effective width. These models also allow for mixed use facilities that include bicycles (HCM 2010).

Current Analysis Methods

HCM Methods

The Highway Capacity Manual 2010 provides three approaches to estimating the effects of pedestrian-vehicle interaction. The first two, applicable at signalized intersections and free flowing approaches to unsignalized intersections, have clearly defined methodologies and relationships, while crossings at stop-controlled approaches are assumed to have minimal delay to both pedestrians and vehicles.

Signalized Intersections

The method assumes that pedestrians cross during the walk and clear phases. The clearance interval is the time required to cross at an assumed pedestrian speed of 3.5 mph. The available Time-Space at each crosswalk is calculated and includes permitted lefts, right turns, and right turns on red (pedestrians yield to vehicles). Pedestrian delay is estimated for each crosswalk (major and minor).

Unsignalized Intersections (Midblock and TWSC Major Street Crossings)

Pedestrian methodologies were developed separately from the vehicle methodology and have different limitations. Up to 4 lanes can be crossed, 8 with a center median. Critical headway is used and pedestrians can platoon. Driver yielding is included with a model for each of 1, 2, 3, and 4 lane crossings that considers likelihood of yielding, pedestrian gap acceptance and vehicle flow rates. Some field yielding rates are included with staged (told how to cross) and un-staged pedestrians. These models have not been empirically developed or calibrated to field observations.

Unsignalized Intersections (AWSC)

The HCM does not include pedestrians in the AWSC methodology as the methodology assumes that all vehicles will wait at the stop bar for pedestrians to cross. Though no methodologies are included, there is discussion on factors that may affect pedestrian delay including traffic volume, number of approach lanes, proportion of turning traffic, and pedestrian volumes.

Microsimulation Methodologies

The following is a list of possible microsimulation approaches to modeling pedestrian and vehicle interactions. The list includes examples of the approach applied in research where found and the type of crossing modeled.

No Pedestrian-Vehicle Interaction

Pedestrians do not interact with vehicles in the microsimulation tool. Pedestrians may be simulated separately without consideration to vehicles or not at all.

Pedestrian Impedance on Vehicle Traffic

Adjustments to saturated-flow of turning movements are based on pedestrian volumes. Vehicles yield to pedestrians, causing a restriction on the saturation flow rate (Alhajyaseen, et. al., 2012 and Rouphail and Eads, 1997). These models were applied at signalized crossings.

All Yield/Signal Controlled

Pedestrians cross only during protected phase at signalized intersections. Only interactions are during the pedestrian phase when permitted lefts or right turns interrupt pedestrian flow. Pedestrians yield to drivers at any conflict. This is analogous to a microsimulation implementation of the HCM signalized intersection methodology.

Pedestrians as Small Vehicles

Pedestrians follow standard vehicle gap-acceptance models using priority similarly to vehicle to vehicle interaction. The gap acceptance model may be calibrated to pedestrian characteristics. This methodology can be applied at signalized and unsignalized intersections (Yang, et. al., 2006).

Cellular Automata

The CA approach is applied with vehicles traveling perpendicular to pedestrians. In the model developed by Ottomanelli, the vehicle is first considered when entering the a calculated sight distance to the crossing, stops for pedestrians already in crosswalk, decides whether to yield for waiting pedestrians, and then crosses after yielding/not yielding. The model was calibrated at a midblock crossing near a college campus (Ottomanelli, et. al., 2012).

Force-based Model

While not well documented in the literature, it is theoretically possible to consider a force-based microsimulation of pedestrian-vehicle interactions where the social forces are defined separately for pedestrians and vehicles (Fellendorf, et. al., 2012).

Current Commercial Microsimulation Approaches to Pedestrian Modeling

A sample of microsimulation packages were surveyed to determine if and how pedestrians and pedestrian-vehicle interaction are modeled. Table 22 shows current commercial microsimulation approaches to pedestrian modeling, along with the types of facilities modeled, where examples were found, and if parameters for varying pedestrian behavior are supplied.

Table 22: Current Commercial Microsimulation Approaches to Pedestrian Modeling

Simulation Tool/Package	Pedestrian-Only Modeling	Pedestrian-Vehicle Interaction	Facility Types	Pedestrian User Types Available
CORSIM	None	Vehicle Impedance Model	Signalized	No
VISSIM	Force-Based	Peds as Small Vehicles	Both	Yes- Gap Acc.
AIMSUN	None	“Legion” Plug-in	Both	TBD
PARAMICS	TBD	TBD	Both	Yes
SimTraffic	TBD	TBD	Both	TBD
HUTSIM	TBD	TBD	TBD	TBD
Other??	TBD	TBD	TBD	TBD

Data Collection Approaches

Background

A scan of relevant literature revealed several research efforts that aimed at studying driver attitudes or pedestrian crossing behaviors. For obtaining data in such studies, three different data collection techniques have been adopted: observational, instrumented vehicle, and driving/pedestrian simulator approaches.

Observational studies are the most traditional method employed in the collection of empirical driving and pedestrian behavior data. They can be used to obtain data from attributes that are fixed (such as vehicle type, pedestrian characteristics, geometric characteristics, etc.), those that change dynamically (e.g., vehicle speeds, pedestrian speeds, distance headways, traffic signal indications, etc.) as well as to record qualitative observations (such as driver or pedestrian distraction). Observational data are obtained from trained observers with the help of tally sheets, count boards, video surveillance equipment, and radar detection devices.

Instrumented vehicles, on the other hand, permit quantitative assessments of driver performance in the field, under actual road conditions. These measurements are not subject to the type of human bias that affects inter-rater reliability on a standard road test. Moreover, the internal network of modern vehicles makes it possible to obtain information from the driver's own automobile, providing opportunities to study in depth driver strategy, vehicle usage, upkeep, drive lengths, route choices, and decision-making (Rizzo et al., 2002). The instrumentation enables researchers to record driver characteristics and vehicle operational parameters. Driver characteristics include galvanic skin response, heart rate, and muscle activity. Examples of vehicle operating characteristics that can be gathered using an instrumented vehicle include steering motion, braking actions, speed, distance and tri-axial accelerations (Helander and Hagvall, 1976). Figure 33 shows a typical instrumented vehicle with relevant connections.

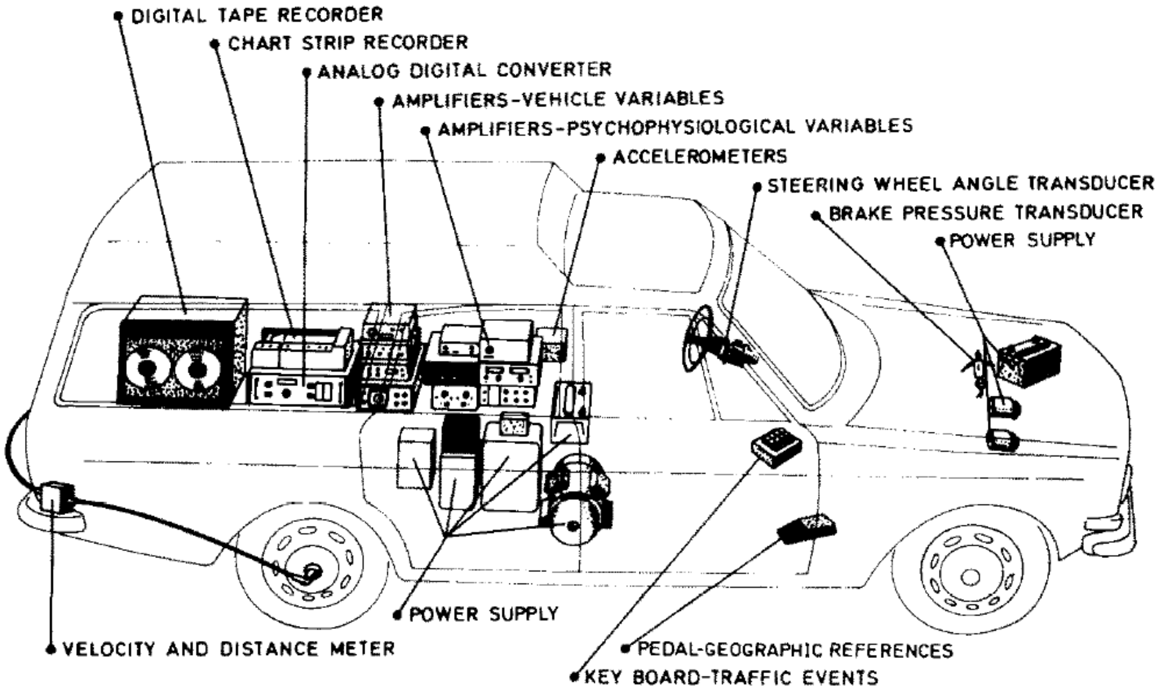


Figure 33: Typical Instrumented vehicle (Helander and Hagvall, 1976)

Laboratory simulators can also be employed to assess behavior in response to synthetic reality. Driving simulators make it possible to observe driver behavior in controlled environments without the risk of driving on the road. It offers a cost-effective alternative to real world naturalistic studies and allows for independent variables to be systematically manipulated so that driver behavior can be measured precisely and safely (Rizzo et al., 2002). Since their introduction in the 1960s they have undergone many advances in terms of computing, visual display, and vehicle dynamics capabilities (Rudin-Brown et al., 2002). Even the lower fidelity simulators are able to collect vast amounts of data, which is one of their reported advantages over naturalistic investigative methods (Moroney and Lilienthal, 2009). Typical dependent measures of driving performance that are collected in driving simulation research studies include vehicle speed, acceleration, braking reaction time, lane position, etc. Similar to the driving simulators, pedestrian simulators also exist that can be used to study pedestrian behavior in controlled environments.

Data Collection Studies

Observational

Observational data collection methods are widely employed in pedestrian behavior analysis. They are leveraged for manifold purposes; a few examples include crash analysis, en-route choice modeling, and assessment of level of service for various facilities. A cursory note of these methods indicate either direct observation approach or video recording based approach as the major means to collect data. Many of the research efforts introduced in Section 3 used observational methods to collect data for studying pedestrian crossing attributes (such as pedestrian crossing speed and pedestrian compliance) as well as pedestrian vehicle interactions (such as gap acceptance, and driver yielding behavior) for a variety of users and crossing types.

In other studies, Zeedyk and Kelly (2003) used unobtrusive observations of 123 adult-child pairs at pedestrian crossings to model the adult-child crossing behavior. Eight types of maneuvers were considered in this study: crossing within the confines of the crosswalk, curb stoppage, oral instruction from adult to child, pressing the button for pedestrian signal, checking for traffic emerging from either direction before initiating crossing, holding hands during crossing, and walking/running (child). Fischer's exact chi-square was used to compare the observations (Zeedyk and Kelly, 2003).

Hatfield and Murphy (2007) investigated the effect of mobile phone usage on crossing speed of pedestrians based of field observation data. The study group comprised of 270 females and 276 males. Both genders were observed to walk slowly when using a mobile phone during crossing. Females were found to be more likely to not look at the traffic before starting a crossing maneuver (Hatfield and Murphy, 2007).

Overall, observational data collection methods require minimal investment in equipment, allow for direct observation of natural pedestrian crossings and driver decision making, and provide first source information to calibrate simulation models. The main shortcoming is the lack of control to cover a specific range of parameters as part of the experiment and isolate others that may bias the data sample.

Instrumented Vehicle

The review of the literature confirms that the use of instrumented vehicles to gather driver behavior measures in the context of driver-pedestrian interaction has gained little attention as of now. Still, studies that utilized instrumented vehicles for gathering driving behavior data can provide some useful insights on experimental design, resource requirements, advantages and limitations.

For example, a study by Boyce and Geller (2000) conducted experiments using instrumented vehicles to assess risky driving behavior. The participant group comprised of 61 licensed drivers with ages ranging from 18 to 82 years. They were paid $10 per hour for participation in the study. The participants were distributed in three groups: younger, middle aged and older. The risky behavior was assessed by means of speeding, on-task behavior, turn-signal use, and following distance (Boyce and Geller, 2002). A study by Rizzo et al. (2002) used instrumented vehicle to collect data for the assessment of fitness and know-how of diverse young and old driving population and develop objective measures to distinguish normal and potentially unfit drivers (Rizzo et al., 2002).

Other research efforts used instrumented vehicles to study driver distraction. For example, an experiment by Texas Transportation Institute (TTI) tried to describe driver behavior under distraction using an instrumented vehicle. The experiment had three tasks, control, reading and writing, on which the driver performance was evaluated. Upon acquaintance with test driving conditions, they were sent two stories via MMS for reading and a short story to be written. Using in-vehicle instrumentation and Psychopy data collection software, data related to speed, lateral lane position, steering, brake, accelerator, light response times, and reading/texting rates were gathered. Some of the major observations of the study include lower mean speed than posted speed while texting and difficulty in maintaining lane discipline while texting (Cooper et al., 2011).

Driving Simulators

Another widely used approach to evaluate driver behavior is using driving simulators. However, there are not many instances where driving simulators were used to examine driver behavior with pedestrian crossing stimulus. In a relevant study, Pradhan et al. (2005) researched the yielding propensity of drivers at mid-block crossings using a driving simulator. The participants were grouped in novice drivers (16-17 years), young drivers (19-29 years), and older drivers (60-75 years). Each group had 24 participants. The position of vehicle, velocity and point of driver's gaze were recorded. The stopping propensity and eye movement were used for developing indices of safe driving behavior. Some scenarios presented to drivers included right turn with walk signal, an intersection with hidden sidewalk, and truck parked in front of sidewalk (Pradhan et al., 2005).

In another study, Fisher and Garay-Vega (2012) conducted simulator based experiments for assessing driver behavior in sight limited, multi threat scenarios. The subjects were divided into two groups, each comprising of 18 in number. The subjects were paid $20 for participation in the experiment. A fixed base Saturn Sedan was used as the simulator vehicle and three screens with 150° horizontal and 30° vertical vision were used in the experiment. The simulator was equipped with audio input. The study assessed the likelihood that sight limited drivers who were presented with a multi threat scenario would skim for pedestrians in the expected zone. The likelihood of yielding at the last minute appearance of pedestrian on provision of advance yield signage was also assessed. The experiment recorded whether or not the driver identified the target zone, crosswalk time upon locating a pedestrian, and percentage of vehicle yielding (Fisher and Garay-Vega, 2012).

Edquist et al. (2012) investigated the effects of on street parking and visual complexity associated with the roadside environment on speed and reaction time. A low complexity and a high complexity scenario with different curb side parking assumptions were assessed in the simulator. The participant group comprised of 29 drivers, 15 of which were male. The ages of the study subjects varied from 20 to 53 years. They were paid $30 for their time. Upon unexpected sight of pedestrian event, variables such as time to accelerator release, time to brake, minimum distance, minimum time to collision, and number of collisions were recorded and evaluated (Edquist et al., 2012).

Hazard perception of elderly drivers and experienced drivers in regards to pedestrian presence was compared using two different approaches by Bromberg et al. (2012). They compared the response to a traffic scene video against the response in a driving simulator. The participants were divided into two groups, experienced (28-40 years) and elderly experienced drivers (65 and above). The first group consisted of 22 participants and the second one with 20 participants. The participants had different visual acuity profiles ranging from 6/6 to 6/12. Participants were paid $15 for their participation (Bromberg et al., 2012).

The validity of a driving simulator, in terms of its ability to reliably measure a given aspect of driving performance, depends on a number of factors associated with physical validity (simulator "fidelity") and behavioral validity (Rudin-Brown et al., 2009). The choice of whether to use a driving simulator should be based on whether the simulator is sufficiently valid for the specific task or behavior under investigation (Kaptein et al., 1996).

Pedestrian Simulators

Charron et al. (2012) used a pedestrian simulator to gauge the risk taking behavior in child pedestrians. In this study, 80 children with median age of 10 years took the simulator test that requested subjects to maneuver the crosswalk. The experiment design consisted of reaching two targets (mailbox, cinema) one after another within a 3 minute timeframe. The targets were connected in such a way that it will take greater time to reach the targets by crosswalk usage. Variables recorded in this study include the subjects' decision to use the crosswalk or not, to walk or run, and to observe the vehicles while crossing (Charron et al., 2012).

Several studies point to marked differences in pedestrian crossing behavior based on age, able bodied condition, or crossing in groups. For instance, Simpson et al. (2003) used a virtual reality system to investigate the differences in crossing behavior between children and young adults. The study comprised of 24 participants equally distributed in the following age groups: 5-9, 10-14, 15-19, and >19 years. Each age group had equal number of male and female participants. The youngest age group was found to make the most unsafe crossings. The system collected collisions, tight fits (potential collisions with vehicle less than 1.5 s away from pedestrian), time headway, and rejected gaps for each crossing maneuver (Simpson et al., 2003).

Several studies reported use of pedestrian simulators to study behavior of subjects (especially young adults) crossing the street with potential distraction due to multimedia devices. A study by Schwebel et al. (2012) found small but meaningful impacts caused by distraction due to multimedia devices. The participant group consisted of 138 college students subjected to cross a virtual street. The participants were randomly assigned to three distinct groups with distraction: talking on phone, texting, listening, with a fourth group without any of these distractions. The variables recorded to model the distraction included elapsed time after pedestrian finished the crossing maneuver and arrival of next vehicle in the crosswalk, left/right observation, looking away, hit instances, and missed crossing opportunities (Schwebel et al., 2012).

Summary of Data Collection Options

Among the existent data collection alternatives, nature of study and available resources govern the choice of preferred alternative. Preliminary findings from literature review shows some advantages related to simulator based data gathering techniques for modeling driver and/or pedestrian distraction. However, the development of experiments appropriate to realistically model pedestrian/vehicle interactions for a wide range of users and facility types is a complex and expensive proposition.

The instrumented vehicle technique renders valuable insights in driver behavior analysis and can be effectively utilized in controlled experiments to study driver yielding behavior associated with pedestrian presence. However, they cannot provide insights about pedestrian gap selection, which is an important element of our work.

Field observational studies, on the other hand, allow for observation of naturally occurring pedestrian crossings as well as driver actions in a coordinated fashion. Such studies allow observation of vehicle type, pedestrian type, gap size, pedestrian-vehicle conflicts etc. as well as gathering of data for determining the percentage of driver yielding, average observed speeds, pedestrian delay, and other variables important for the development of behavioral models for drivers and pedestrians in our study.

- This page intentionally left blank -

APPENDIX B: MIDBLOCK CROSSING STUDY

Data Collection Methodology

The team proposes to collect a variety of empirical data on pedestrian-vehicle interaction. Some of these data will be related to attributes that are changing dynamically (e.g. vehicle speed, vehicle relative distance to the crosswalk), while others will be static descriptors of the pedestrian-vehicle interaction event (e.g. vehicle type, pedestrian characteristics). To evaluate the effect of these variables, they need to be collected accurately using reliable measurement devices. They also need to be coded consistently between crossing events, sites and locations. In this data collection methodology, the temporal beginning of an interaction event is defined as follows:

> *A pedestrian-driver interaction event commences as a pedestrian arrives in the crosswalk influence area (CIA) or waiting location while a driver is on the approach of the crosswalk.*

All interaction variables will be coded relative to this point in time, and the methodology assumes that the following statements are true:

- The pedestrian has indicated that he or she is intending to cross at the facility (rather than continuing along the sidewalk).
- The pedestrian is aware of the approaching vehicle and decides whether or not he or she feels comfortable to cross the road.
- The driver is aware of the pedestrian's intention and must react in some fashion (make the decision to yield or continue through the crosswalk).
- The observer is aware that an event sequence (action-reaction) is about to take place (from video observation) and records the attributes of the interaction event.

The assumptions above are valid if both driver and pedestrian are consciously aware of each other's presence. Clearly, there are cases where that is not true, as driver or pedestrian may be distracted. In an observatory study the cognitive awareness of the involved parties is not discernible, but can only be presumed from erratic or unexpected behavior. For example, a pedestrian may step into the roadway and then retreat quickly, realizing that he or she misjudged the position of the vehicle. Similarly, a driver may perform an emergency braking maneuver after belatedly recognizing the presence of the pedestrian. In the case of a pedestrian retreating, this event should be coded as a separate event. Speed and other vehicle dynamics would be recorded and a note would be made such as "pedestrian pull-back".

From the onset of a pedestrian-driver interaction event, there are three potential outcomes to the interaction of the two modes:

1. Pedestrian GO Decision [GO] – The pedestrian decides that there is sufficient time for a safe crossing and steps into the crosswalk.
2. Pedestrian NO-GO Decision [NOGO]/ Driver Non-Yield Decision [NY] – The pedestrian decides that the time until the expected vehicle arrival time to the crossing point is too short to safely cross the facility, i.e. he/she rejects the lag or gap. At the same time, the driver decides that it is either physically impossible to yield to the pedestrian, or he/she is unwilling to yield.
3. Driver Yield Decision [Y] – The approaching driver decelerates and creates a crossing opportunity for the pedestrian, which may occur with or without coming to a complete stop.

The team proposes to use a three-pronged data collection approach that combines real-time observations by a trained observer on a tally sheet, video recording of the crosswalk, and Lidar speed measurements of approaching vehicles. Figure 34 shows a schematic of the data collection set-up.

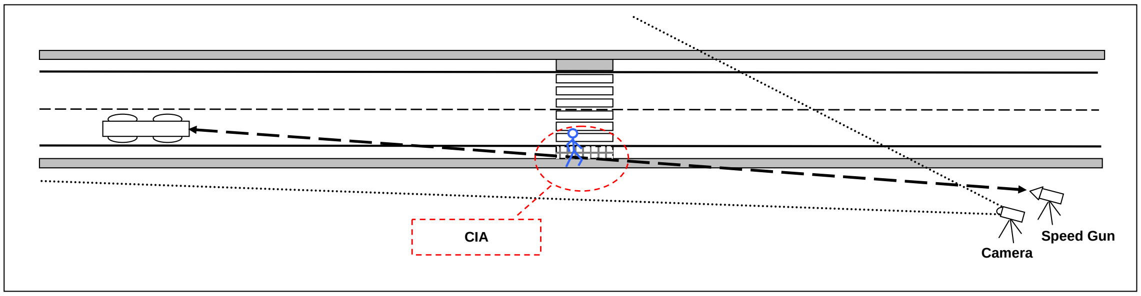


Figure 34: Proposed Field Data Collection Set-Up

In order to capture all relevant data, the video angle has to cover events concurrent to the interaction, such as the presence of an adjacent yield or multiple pedestrians. As shown in the diagram, the video camera angle is wide enough to cover the crosswalk influence area (CIA) or waiting location, and the approach to the crosswalk. The speed recordings from Lidar need to be visible on the video camera and to the field observer, or alternatively need to be recorded audibly on the video. This setup will also allow us to record “illegal” pedestrian crossings in the immediate vicinity of the pedestrian crossing. These will also be analyzed and considered within the broader context of replicating pedestrian crossings along various types of facilities and for various types of crossings.

The heading of the data collection sheet provides the date/time, observer(s), distance to crosswalk (dist. to CW), intersection, approach, and crossing distance. The entire distance must be visible in the video, so that walking speed may later be calculated using this distance and the crossing time (TIME on the data collection sheet). Average speed should also be calculated from unimpeded speeds at each location. These can be done by collecting speeds and distances for vehicles where the speed is not affected by pedestrians or platoons. Record these speeds and distances (for a sample of 30 vehicles) on a separate sheet of paper or along the edge of the data collection sheet. The following table summarizes all data collection elements used in determining the driver yielding behavior, followed by some additional details on variable definitions. Additional variables were needed to determine gap acceptance behavior and will be discussed later. It was discovered that adding a column in excel for the time stamp (and another for the file name, when video spans several files) from the video would help identify each observation.

Data Collection Manual

Table 23: Data collection measures obtained and used in modeling

	Factor	**Description**	**Value**
First Vehicle Variables	SPD	The speed of the first vehicle (mph), at the time the pedestrian arrives at crosswalk influence area (waiting location), recorded from speed gun	Mph
	DIST	The distance from the first vehicle to the researcher recorded from the laser speed gun	Ft
	YIELD	Whether the first vehicle yielded and if it was a hard or soft yield	No=NY, Soft= SY, Hard=HY

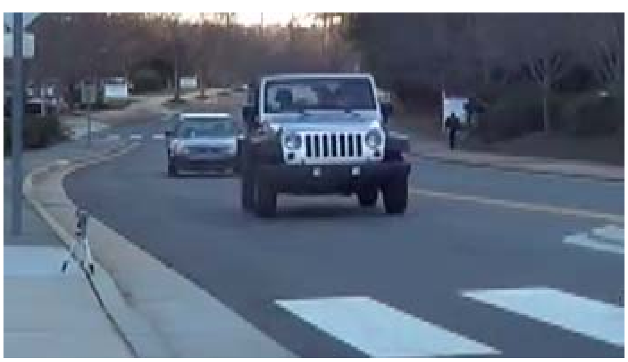
	NEAR	Whether the vehicle for which speed and distance was recorded (first vehicle) was in the lane nearest or farthest from the pedestrian	Near=1, Far=0
	TRIG	If the first vehicle yielded, was it triggered (forced) by the pedestrian. In other words, if the yield happened before pedestrian stepping into the crosswalk (0) or after (1)	Triggered Yield, Yes=1, No=0
	STP	Whether the first vehicle had already stopped at the time that the pedestrian arrived	Stopped=1
	ADJ	Whether there was a yield on the other side of the road (opposite direction) or a yield in an adjacent lane (same direction)	Adj. Yield=1
	PLT	If the first vehicle was in a platoon or had a close follower	Platoon=1
	LSPLT	If the first vehicle was in a platoon or had a close follower and was travelling at a speed less than or equal to 15 mph	Low-speed Platoon=1
	HGV	First vehicle type: passenger car or heavy vehicle (bus or truck)	Heavy Vehicle=1
Pedestrian Variables	MUP	If there were other pedestrians present near the crosswalk; if any pedestrian is at either sides of the street or the splitter island and intends to cross	Multiple Pedestrian=1
	MED	Whether the pedestrian crossed from the median or the curb	Median=1, Curb=0
	CTRL	Whether the crossing pedestrian was controlled (researcher) or random (observational study)	Controlled=1, Random=0
	CROSS	Whether the pedestrian crossed in a gap or a yield	Gap/Yield (G/Y)
	IN_CW	Whether the pedestrian stopped in the crosswalk or at the curb. This variable shows the behavior of the pedestrian. A passive pedestrian is defined to wait at the curb for a crossing opportunity however an assertive pedestrian is defined to be waiting in the crosswalk or walking toward the crosswalk	Crosswalk=1, Curb=0
	AGE	Researcher's estimate of the pedestrian's age group	Young=1, Middle/Older=0
	DISTR	Whether the pedestrian was distracted by an outside source, such as a cell phone	Distracted, Yes=1, No=0
	BUSINESS	Researcher's observation of the pedestrian's attire or clothing	Business=1, Casual=0
	FEMALE	Pedestrian's gender	Female=1, Male=0
Site Variables	CAMPUS	This variable distinguishes sites on-campus (1) from those off-campus (0)	On-Campus=1
	FLORIDA	This variable distinguishes sites in the state of Florida (1) from those in the other two states (0)	Florida=1

	NCAROLINA	This variable distinguishes sites in the state of North Carolina (1) from those in the other two states (0)	North Carolina=1
	Distance to Crosswalk	The distance from the researcher using the laser speed gun to the middle of the crosswalk along the curb.	Ft
	Crossing Distance	The distance from the curb to a measured location, such as a specific white crosswalk marking, a center line, or the opposite side of the crosswalk.	Ft
Other	COUNT	If the first vehicle did not yield, how many vehicles went through before the pedestrian crossed	Number
Video	DIST_DEL	Delay between when the speed should have been taken (at the time pedestrian arrives at the waiting location) and when the gun beeped	Seconds
	ADJDIST	Vehicle position at the time of pedestrian arrival in crosswalk influence area measured in feet using a LIDAR speed measurement device; ADJDIST is calculated from measured distance, speed, distance delay and Distance to Crosswalk; ADJDIST=DIST+SPD*1.467*DIST_DEL-Distance to Crosswalk	Ft
	TTC	Time until vehicle would theoretically arrive at the crosswalk; TTC is calculated from the measured speed and distance at the time pedestrian arrives in the crosswalk influence area; TTC=ADJDIST/(SPD*1.467)	Seconds
	DECEL	Deceleration rate necessary to come to a full stop prior to crosswalk; DECEL is calculated from measured speed and adjusted distance; DECEL=(SPD*SPD)/(2*ADJDIST)	Ft/s^2

Factors in Further Detail

PLT: If the first vehicle was in a platoon or had a close follower (Platoon=1)

- It is predicted that if the first vehicle is part of a platoon, they are less likely to yield
- The platoon should happen upstream, not right at the crossing
- What distance between vehicles is considered a platoon? (The following two examples were considered as platoons)



HGV: First vehicle type, passenger car or heavy vehicle (bus or truck) (Heavy Vehicle=1)

The following are NOT considered heavy vehicles:

MUP: If there were other pedestrians present near the crosswalk; if any pedestrian is at either sides of the street or the splitter island and intends to cross (Multiple Pedestrian=1)

> Question: When the first pedestrian has already crossed one direction and is crossing the opposite side of the road, the second pedestrian arrives. Should it be considered as MUP=1, if the driver of the vehicle may not consider it as a pedestrian group and has already moved forward?
>
> Response: For the condition of multiple pedestrians, we count any pedestrian that intends to cross from any directions or side of the street. So this case is a MUP=1. The pedestrian (other than our subject pedestrian) could be crossing the studied direction or opposite direction.

IN_CW: Whether the pedestrian stopped in the crosswalk or at the curb. This variable shows the behavior of the pedestrian. A passive pedestrian is defined to wait at the curb for a crossing opportunity however an assertive pedestrian is defined to be waiting in the crosswalk or walking toward the crosswalk (Crosswalk=1, Curb=0)

> Question: If there is a big gap when a pedestrian arrives at the crosswalk, and he/she just walked across, should it be considered as an aggressive one (IN_CW=1)? In this condition, what if his/her walking speed is very low?

Response: If the gap is really large and does not require the driver of the vehicle to brake or slow down, the condition is not aggressive (IN_CW=0). But if the pedestrian walks fast to utilize the gap in traffic and/or vehicle slows down to increase the headway the pedestrian behavior is aggressive (IN_CW=1).

W_SP: Pedestrian walking speed while crossing.

Walking speed can be calculated in excel using the following formula:

$$W_{SP} = CrossingDistance/TIME$$

Crossing distance should be entered in the header, while TIME would be entered into the preceding column.

DELAY: Time from the pedestrian arriving at the crosswalk influence area (waiting location) to stepping into the crosswalk to cross (recorded in seconds)

- Where does delay start?

Gap Acceptance

Several additional variables are needed to adequately explain pedestrian gap acceptance behavior. These variables may be added as new columns in the excel spreadsheets, to the right of DIST_DEL. The table below shows the additional variables. A lag event is defined as occurring between a pedestrian who has just reached the crosswalk and the next vehicle to arrive at the crosswalk. A gap event occurs between successive vehicles while a pedestrian waits at the crosswalk.

Table 24: Additional data collection elements used in determining the gap acceptance behavior

Factor	Description	Value
GO	Whether the pedestrian accepted (GO) or rejected (No-GO) a gap/lag event.	GO=1, No-GO=0
LAG	Whether the pedestrian event is a lag or gap	Lag=1, Gap=0

TIME	Time from the pedestrian stepping into the crosswalk to reaching a measured location (such as a specific white crosswalk marking, a center line, or the opposite side of the crosswalk). This value is used in calculating the pedestrian walking speed.	Seconds
W_SP	Pedestrian walking speed while crossing	Ft/s
OBS	Observed lag or gap time, measured with a stopwatch between pedestrian and vehicle arrival (lag) or successive vehicles (gap)	Seconds
DELAY	Time from the pedestrian arriving at the crosswalk influence area (waiting location) to stepping into the crosswalk to cross	Seconds

These variables will not have values for all observations and some observations may require additional rows. The first part is true for soft and hard yields, since the driver decided to yield to the pedestrian. It is not true for triggered yields, because the pedestrian utilized what they saw as an acceptable gap size. Crossing events that were listed as gaps under the pedestrian variable CROSS will have at least one observation. Gap crossings with vehicle count values greater than zero may require additional rows to account for the rejected gaps, in addition to the accepted gap.

This page intentionally left blank

APPENDIX C: DESCRIPTION OF THE INSTRUMENTED VEHICLE STUDY

Instrumented Vehicle Description

The instrumented vehicle used in this study is a Honda Pilot SUV, owned by the University of Florida – Transportation Research Center (TRC). The vehicle is equipped with a Honeywell Mobile Digital Recorder (HTDR400) system. The vehicle has an inbuilt GPS where all information about vehicle position and speed data is displayed and recorded on the HTDR400. In addition to the GPS unit, the vehicle includes four wide coverage digital cameras (DCs) that capture video clips facing the front, the back and the two sides of the vehicle. The video data, as well as audio data during the driving task, are recorded on the HTDR400, and stored at a local hard drive that is located at the trunk of the vehicle. An additional camera facing the driver will be installed on the dashboard, to capture possible facial reactions of the driver during the experiments, and to record the participant's comments while driving. An internal view of the instrumented vehicle is provided in Figure 35. The data that can be collected directly through the instrumented vehicle include: Instrumented vehicle geographical position, speed, throttle, and left-right turn signal activation, Video clips of the vehicles in front, behind and adjacent to the instrumented vehicle, Audio and video recordings of the participant during the driving task.

A laptop is connected to the system which allows for reviewing the display of all four cameras, through the HTRD BusView software. It is also possible to download the video clips from the hard drive to the laptop.

Figure 35: Inside View of the TRC Instrumented Vehicle

Instrumented Vehicle Study Route Descriptions

Route Description

Route #1:

1. Start from the parking lot, go south along Gale Lemerand Dr.
2. Turn right to Museum Rd.
3. Turn right to Radio Rd.
4. Turn left to SW 34th Rd.
5. Turn left to Hull Rd.

6. Turn left to Mowry Rd.
7. Turn left back to Gale Lemerand Dr.
8. Proceed back to the parking lot across the football stadium on the left side.

Total length: 4.7 mile. Total duration: Approximately 16 min.

Route #2:

1. Start from the parking lot, go south along Gale Lemerand Dr.
2. Take the 1st left to Stadium Rd.
3. Continue to Buckman Dr.
4. Turn right to Union Rd.
5. Take the 1st right to Newell Dr.
6. Turn right to SW Archer Rd.
7. Take the 1st right to Center Dr.
8. Take the 3rd left to Museum Dr.
9. Turn right back to Gale Lemerand Dr.
10. Proceed to the parking lot across the football stadium on the left side.

Total length: 2.8 mile. Total duration: Approximately 20 min.

APPENDIX D: YIELD MODELING DETAILED RESULTS

Predicting Probability of Yielding, P(Yield)

Correlation Table (Y vs. NY)

The following table shows a correlation analysis of the response variable (Y_NY = 1 for yield events), as a function of various independent variables. Each cell contains the correlation coefficient (ranges from -1 to +1) in the top row, and the p-value of the Chi-Square correlation test in the second row.

Table 25: Probability of Yielding Correlation Table

Pearson Correlation Coefficients, N = 975
Prob > |r| under H0: Rho=0

	Y_NY	SPD	ADJDIST	NEAR	ADJ	PLT	HGV	MUP	MED	CTRL	TTC	DECEL	STUDENT	DISTR	BUSINESS	FEMALE
Y_NY	1.00000	-0.37698 <.0001	-0.02631 0.4119	-0.06577 0.0400	0.19581 <.0001	0.09812 0.0022	0.01653 0.6061	0.24610 <.0001	0.21079 <.0001	-0.14696 <.0001	0.21460 <.0001	-0.15638 <.0001	0.23039 <.0001	0.01167 0.7160	-0.06696 0.0366	0.10427 0.0011
SPD	-0.37698 <.0001	1.00000	0.36158 <.0001	0.05869 0.0670	-0.15163 <.0001	-0.09315 0.0036	-0.14360 <.0001	-0.25649 <.0001	-0.02244 0.4840	0.31240 <.0001	-0.31806 <.0001	0.33970 <.0001	-0.24174 <.0001	-0.02384 0.4571	-0.11683 0.0003	-0.25464 <.0001
ADJDIST	-0.02631 0.4119	0.36158 <.0001	1.00000	-0.09417 0.0032	-0.02610 0.4156	0.02504 0.4347	0.09996 0.0018	-0.09842 0.0021	0.09946 0.0019	0.01135 0.7233	0.70652 <.0001	-0.33511 <.0001	0.06402 0.0456	0.06568 0.0403	-0.02299 0.4733	-0.18821 <.0001
NEAR	-0.06577 0.0400	0.05869 0.0670	-0.09417 0.0032	1.00000	-0.03318 0.3006	0.07319 0.0223	-0.02888 0.3677	-0.09352 0.0035	-0.54721 <.0001	0.30739 <.0001	-0.13092 <.0001	0.03724 0.2454	-0.01371 0.6690	-0.07316 0.0223	-0.06026 0.0600	-0.02080 0.5165
ADJ	0.19581 <.0001	-0.15163 <.0001	-0.02610 0.4156	-0.03318 0.3006	1.00000	0.06436 0.0445	-0.03734 0.2441	0.09436 0.0032	0.12620 <.0001	-0.00967 0.7629	0.08887 0.0055	-0.04453 0.1647	0.05329 0.0963	0.00423 0.8950	-0.00673 0.8338	0.08597 0.0072
PLT	0.09812 0.0022	-0.09315 0.0036	0.02504 0.4347	0.07319 0.0223	0.06436 0.0445	1.00000	0.09491 0.0030	0.03982 0.2141	0.04096 0.2013	0.03208 0.3170	0.09995 0.0018	-0.10581 0.0009	0.00049 0.9878	0.03353 0.2956	0.00803 0.8022	0.02182 0.4962
HGV	0.01653 0.6061	-0.14360 <.0001	0.09996 0.0018	-0.02888 0.3677	-0.03734 0.2441	0.09491 0.0030	1.00000	0.12388 0.0001	-0.01339 0.6762	-0.10871 0.0007	0.21782 <.0001	-0.08961 0.0051	0.08209 0.0103	0.07094 0.0268	0.01157 0.7182	0.01172 0.7148
MUP	0.24610 <.0001	-0.25649 <.0001	-0.09842 0.0021	-0.09352 0.0035	0.09436 0.0032	0.03982 0.2141	0.12388 0.0001	1.00000	0.01919 0.5495	-0.48996 <.0001	0.10353 0.0012	-0.09046 0.0047	0.03030 0.3447	0.05651 0.0778	0.03319 0.3005	0.03250 0.3106
MED	0.21079 <.0001	-0.02244 0.4840	0.09946 0.0019	-0.54721 <.0001	0.12620 <.0001	0.04096 0.2013	-0.01339 0.6762	0.01919 0.5495	1.00000	0.05617 0.0796	0.11924 0.0002	0.03762 0.2405	0.22320 <.0001	-0.04926 0.1243	-0.10065 0.0017	-0.02150 0.5026
CTRL	-0.14696 <.0001	0.31240 <.0001	0.01135 0.7233	0.30739 <.0001	-0.00967 0.7629	0.03208 0.3170	-0.10871 0.0007	-0.48996 <.0001	0.05617 0.0796	1.00000	-0.21989 <.0001	0.12273 0.0001	0.11919 0.0002	-0.12683 <.0001	-0.28967 <.0001	-0.02401 0.4539
TTC	0.21460 <.0001	-0.31806 <.0001	0.70652 <.0001	-0.13092 <.0001	0.08887 0.0055	0.09995 0.0018	0.21782 <.0001	0.10353 0.0012	0.11924 0.0002	-0.21989 <.0001	1.00000	-0.55213 <.0001	0.19072 <.0001	0.09253 0.0038	0.05629 0.0790	-0.03702 0.2481
DECEL	-0.15638 <.0001	0.33970 <.0001	-0.33511 <.0001	0.03724 0.2454	-0.04453 0.1647	-0.10581 0.0009	-0.08961 0.0051	-0.09046 0.0047	0.03762 0.2405	0.12273 0.0001	-0.55213 <.0001	1.00000	-0.03817 0.2338	-0.05502 0.0860	-0.04083 0.2027	0.03580 0.2640
STUDENT	0.23039 <.0001	-0.24174 <.0001	0.06402 0.0456	-0.01371 0.6690	0.05329 0.0963	0.00049 0.9878	0.08209 0.0103	0.03030 0.3447	0.22320 <.0001	0.11919 0.0002	0.19072 <.0001	-0.03817 0.2338	1.00000	-0.02272 0.4786	-0.23479 <.0001	0.22034 <.0001
DISTR	0.01167 0.7160	-0.02384 0.4571	0.06568 0.0403	-0.07316 0.0223	0.00423 0.8950	0.03353 0.2956	0.07094 0.0268	0.05651 0.0778	-0.04926 0.1243	-0.12683 <.0001	0.09253 0.0038	-0.05502 0.0860	-0.02272 0.4786	1.00000	0.02248 0.4831	-0.00674 0.8334
BUSINESS	-0.06696 0.0366	-0.11683 0.0003	-0.02299 0.4733	-0.06026 0.0600	-0.00673 0.8338	0.00803 0.8022	0.01157 0.7182	0.03319 0.3005	-0.10065 0.0017	-0.28967 <.0001	0.05629 0.0790	-0.04083 0.2027	-0.23479 <.0001	0.02248 0.4831	1.00000	-0.03417 0.2865
FEMALE	0.10427 0.0011	-0.25464 <.0001	-0.18821 <.0001	-0.02080 0.5165	0.08597 0.0072	0.02182 0.4962	0.01172 0.7148	0.03250 0.3106	-0.02150 0.5026	-0.02401 0.4539	-0.03702 0.2481	0.03580 0.2640	0.22034 <.0001	-0.00674 0.8334	-0.03417 0.2865	1.00000

- The following variables show a significant positive correlation with the dependent variable, suggesting an increase in yielding with an increase in the variable (or binary variable change from 0 to 1): ADJ, PLT, MUP, MED, TTC, AGE, and FEMALE
- The following variables show a negative or inverse correlation (<-0.3) with the dependent variable, suggesting a decrease in yielding with an increase in the variable (or binary variable change from 0 to 1): SPD, NEAR, CTRL, DECEL, and BUSINESS
- The following variables are intercorrelated: ADJDIST to TTC (0.70652), NEAR to MED (-0.54721), MUP to CTRL (-0.48996), TTC to DECEL (-0.55213)
- It is expected that ADJDIST, TTC, and DECEL are intercorrelated since TTC and DECEL are calculated using the ADJDIST

Binary Logit (Y vs. NY, full model) [Y-1]

Table 26: Yield Model Y-1 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.8820	0.6443	8.5328	0.0035
SPD	1	-0.1846	0.0296	38.9600	<.0001
ADJDIST	1	0.0107	0.00364	8.5578	0.0034
NEAR	1	0.5844	0.2121	7.5951	0.0059
ADJ	1	1.0466	0.2983	12.3067	0.0005
PLT	1	0.2567	0.1622	2.5037	0.1136
HGV	1	-0.7799	0.3489	4.9960	0.0254
MUP	1	1.0254	0.2346	19.1008	<.0001
MED	1	1.4558	0.2614	31.0254	<.0001
CTRL	1	-0.3252	0.2072	2.4646	0.1164
TTC	1	-0.2036	0.0991	4.2207	0.0399
DECEL	1	-0.00168	0.0335	0.0025	0.9601
STUDENT	1	0.5377	0.2315	5.3961	0.0202
DISTR	1	0.2498	0.7270	0.1181	0.7311
BUSINESS	1	-0.9817	0.3753	6.8441	0.0089
FEMALE	1	-0.0104	0.1601	0.0042	0.9483

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.831	0.785	0.881
ADJDIST	1.011	1.004	1.018
NEAR	1.794	1.184	2.719
ADJ	2.848	1.587	5.111
PLT	1.293	0.941	1.777
HGV	0.458	0.231	0.908
MUP	2.788	1.760	4.416
MED	4.288	2.569	7.157
CTRL	0.722	0.481	1.084
TTC	0.816	0.672	0.991
DECEL	0.998	0.935	1.066
STUDENT	1.712	1.088	2.695
DISTR	1.284	0.309	5.337
BUSINESS	0.375	0.180	0.782
FEMALE	0.990	0.723	1.354

- AIC (1078.595), SC (1156.714), -2 Log L (1046.595), R-Square (0.2591), Max-rescaled R-Square (0.3470)
- This model shows the following variables to be significant: SPD, ADJDIST, NEAR, ADJ, HGV, MUP, MED, TTC, AGE, and BUSINESS
- The following variables are shown to be insignificant: PLT, CTRL, DECEL, DISTR, FEMALE
- Only 10 of the 975 observations involved a distracted pedestrian, so this variable should not be used in further modeling.
- It is expected that forward selection and backward elimination will show the same variables to be significant and insignificant.

Binary Logit (Y vs. NY, forward selection) [Y-2]

Table 27: Yield Model Y-2 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.8782	0.6246	9.0423	0.0026
SPD	1	-0.1888	0.0283	44.5603	<.0001
ADJDIST	1	0.0107	0.00355	9.1045	0.0025
NEAR	1	0.4910	0.1949	6.3444	0.0118
ADJ	1	1.0192	0.2944	11.9861	0.0005
HGV	1	-0.7287	0.3450	4.4612	0.0347
MUP	1	1.1791	0.2121	30.8934	<.0001
MED	1	1.3798	0.2459	31.4786	<.0001
TTC	1	-0.1907	0.0971	3.8551	0.0496
STUDENT	1	0.4682	0.2231	4.4044	0.0358
BUSINESS	1	-0.8490	0.3675	5.3383	0.0209

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.828	0.783	0.875
ADJDIST	1.011	1.004	1.018
NEAR	1.634	1.115	2.394
ADJ	2.771	1.556	4.934
HGV	0.483	0.245	0.949
MUP	3.251	2.145	4.928
MED	3.974	2.454	6.435
TTC	0.826	0.683	1.000
STUDENT	1.597	1.031	2.473
BUSINESS	0.428	0.208	0.879

- AIC (1073.611), SC (1127.318), -2 Log L (1051.611), R-Square (0.2553), Max-rescaled R-Square (0.3418)
- Reasonable that these factors affect yielding decision in the manner shown
- Increased speed reduces chance of yielding. Heavy vehicles are less likely to yield. Drivers are less likely to yield to pedestrians in business attire. Increased time to contact also reduces chance of yielding.
- Driver distance from crosswalk, driver in near lane, adjacent yield, presence of multiple pedestrians, pedestrian crossing from median (pedestrian in refuge or already crossing from opposite side of street), and student age pedestrians increase the chance of yielding
- It was somewhat unexpected that pedestrians in business attire reduce the chance that a driver decides to yield and student age pedestrians increase the chance of yielding. This may reflect a difference in driver behavior by location, since pedestrians in business attire are more likely to be found in downtown areas and students are seen on-campus. A campus variable will be explored in a future modeling step.
- These variables were shown to be insignificant: PLT, CTRL, DECEL, FEMALE
    - It was seen in previous research that driver yielding at roundabouts is affected by platoons, with vehicles traveling in platoons being less likely to yield. It is therefore somewhat surprising that the effect did not show up at the studied midblock sites. In the full model, the PLT variable had a p-value of 0.1136 and an odds ratio of 1.293, which is not significant and further in the opposite direction than expected.
    - Control being insignificant means that staged pedestrians can be used in research, we are not forced to rely on natural pedestrian observations
    - Necessary deceleration does not show up, but speed, adjusted distance, and time to contact were shown to be significant. With these variables already in the model, required deceleration showed less of an effect than if it had been forced into the model
    - It was expected that the gender of the pedestrian would not have an effect on the driver's decision to yield or not

- Backward elimination shows the same results as forward selection. This was expected to occur, but sometimes forward and backward selection produce slightly different results – especially when independent variables are inter-correlated (like TTC and DECEL).

Binary Logit Manual Selection 1 (Y vs. NY, TTC removed) [Y-3]

In the model, the team manually selected variables to add to the model. The variable selection was motivated by the ability to implement this model into a microsimulation environment, which is the primary objective of this project. The team therefore explored different model combinations that could more readily be implemented. In this first attempt, only the SPEED and ADJDIST variables were used, and the interaction variable TTC excluded.

Table 28: Yield Model Y-3 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.9432	0.3894	5.8656	0.0154
SPD	1	-0.1435	0.0150	92.0423	<.0001
ADJDIST	1	0.00418	0.00119	12.3099	0.0005
NEAR	1	0.4824	0.1943	6.1651	0.0130
ADJ	1	1.0006	0.2937	11.6085	0.0007
HGV	1	-0.7919	0.3464	5.2261	0.0223
MUP	1	1.1661	0.2117	30.3435	<.0001
MED	1	1.3593	0.2449	30.8101	<.0001
STUDENT	1	0.4905	0.2218	4.8881	0.0270
BUSINESS	1	-0.8579	0.3707	5.3558	0.0207

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.866	0.841	0.892
ADJDIST	1.004	1.002	1.007
NEAR	1.620	1.107	2.371
ADJ	2.720	1.530	4.836
HGV	0.453	0.230	0.893
MUP	3.209	2.120	4.860
MED	3.893	2.409	6.292
STUDENT	1.633	1.057	2.523
BUSINESS	0.424	0.205	0.877

- AIC (1075.208), SC (1124.032), -2 Log L (1055.208), R-Square (0.2525), Max-rescaled R-Square (0.3382)
- It was seen that removing TTC (a variable that was intercorrelated with ADJDIST and DECEL) from the model only affected the statistics, coefficients, and p-values by small amounts.
- The following variables were shown to be insignificant in the previous models, so they were not used in this model: PLT, CTRL, DECEL, FEMALE
- The following variables were used and determined to be significant: SPD, ADJDIST, NEAR, ADJ, HGV, MUP, MED, AGE, and BUSINESS
- Increasing the speed reduces chance of yielding. Heavy vehicles are less likely to yield. Drivers are less likely to yield to pedestrians in business attire.
- Driver distance from crosswalk, driver in near lane, adjacent yield, presence of multiple pedestrians, pedestrian crossing from median (pedestrian in refuge or already crossing from opposite side of street), and student age pedestrians increase the chance of yielding
- Pedestrians in business attire reduce the chance that a driver decides to yield and student age pedestrians increase the chance of yielding. This may reflect a difference in driver behavior by location, since pedestrians in business attire are more likely to be found in downtown areas and students are seen on-campus.

Binary Logit Manual Selection 2 (Y vs. NY, SPD and ADJDIST removed) [Y-4]

In this second manual selection model, the team removed SPD and ADJDIST, and instead only used the TTC variable. The objective is that the TTC variable more directly describes the time to arrival at the crosswalk, making the case that the interaction of the speed and distance variables is critical in predicting yields.

Table 29: Yield Model Y-4 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.8894	0.3435	30.2600	<.0001
NEAR	1	0.6385	0.2009	10.0975	0.0015
ADJ	1	1.2146	0.2859	18.0461	<.0001
MUP	1	1.0613	0.2209	23.0853	<.0001
MED	1	1.3883	0.2424	32.8091	<.0001
CTRL	1	-0.5115	0.1908	7.1858	0.0073
TTC	1	0.0701	0.0386	3.2970	0.0694
DECEL	1	-0.0903	0.0380	5.6430	0.0175
STUDENT	1	0.9662	0.2097	21.2295	<.0001
FEMALE	1	0.3114	0.1488	4.3788	0.0364

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
NEAR	1.894	1.277	2.808
ADJ	3.369	1.924	5.900
MUP	2.890	1.875	4.456
MED	4.008	2.492	6.445
CTRL	0.600	0.412	0.872
TTC	1.073	0.994	1.157
DECEL	0.914	0.848	0.984
STUDENT	2.628	1.742	3.964
FEMALE	1.365	1.020	1.828

- AIC (1150.536), SC (1199.360), -2 Log L (1130.536), R-Square (0.1925), Max-rescaled R-Square (0.2578)
- It was seen that removing SPD and ADJDIST (ADJDIST was intercorrelated with TTC and both of these variables are used to calculate TTC) from the model affected the statistics, coefficients, and p-values by larger amounts than removing TTC. The overall model fit decreased notably, suggesting that the individual parameters are needed in the model.
- The following variables were shown to be insignificant after the removal of the SPD and ADJDIST variables, so they were not used in this model: PLT, HGV, BUSINESS
- It was seen that after removing these three variables that TTC is no longer shown to be significant. It is recommended that SPD and ADJDIST be kept in the models and TTC be removed.

Binary Logit (Y vs. NY, NC data only) [Y-5]

In the next step, a separate model for NC data only was developed using forward selection.

Table 30: Yield Model Y-5 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.1830	0.8055	2.1570	0.1419
SPD	1	-0.2453	0.0387	40.1743	<.0001
MUP	1	1.1523	0.3691	9.7452	0.0018
MED	1	1.6749	0.7032	5.6725	0.0172
STUDENT	1	2.0210	0.5901	11.7289	0.0006

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.782	0.725	0.844
MUP	3.166	1.535	6.526
MED	5.338	1.345	21.184
STUDENT	7.546	2.374	23.988

- AIC (314.133), SC (333.005), -2 Log L (304.133), R-Square (0.2272), Max-rescaled R-Square (0.3248)
- Sample Size: 93 Yields and 229 Non-yields
- The following variables were determined to be significant: SPD, MUP, MED, and AGE. Increasing speed reduces the likelihood of yielding, while multiple pedestrians, a crossing from the median, and a student increase the likelihood of yielding.

Binary Logit (Y vs. NY, FL data only) [Y-6]

In the next model, only the Florida data were used in model development.

Table 31: Yield Model Y-6 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.9340	0.6156	22.7149	<.0001
SPD	1	-0.1030	0.0279	13.6179	0.0002
PLT	1	0.8589	0.3471	6.1235	0.0133
DECEL	1	-0.1057	0.0432	5.9711	0.0145
FEMALE	1	1.2002	0.3572	11.2903	0.0008

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.902	0.854	0.953
PLT	2.361	1.196	4.661
DECEL	0.900	0.827	0.979
FEMALE	3.321	1.649	6.687

- AIC (272.796), SC (291.315), -2 Log L (262.796), R-Square (0.1834), Max-rescaled R-Square (0.2779)
- Sample Size: 231 Yields and 69 Non-yields
- The following variables were determined to be significant: SPD, PLT, DECEL, and FEMALE. Increasing speed and increasing necessary deceleration rates decrease the likelihood of yielding, while platoon and female increase yielding.
- The PLT effect is surprising, as platooned vehicles have previously been shown to be associated with a decrease in yielding – presumably due to an increased risk of rear-end collisions. Seeing more yielding for platoons is counterintuitive.

Binary Logit (Y vs. NY, AL data only) [Y-7]

In the following model, only the data from the Alabama sites was used in model development.

Table 32: Yield Model Y-7 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	5.2150	1.4085	13.7087	0.0002
SPD	1	-0.3081	0.0645	22.8299	<.0001
ADJDIST	1	0.0344	0.00975	12.4419	0.0004
NEAR	1	1.0411	0.3737	7.7634	0.0053
ADJ	1	2.2172	0.8494	6.8146	0.0090
CTRL	1	-1.8109	0.3710	23.8247	<.0001
TTC	1	-0.7708	0.3070	6.3034	0.0121

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.735	0.648	0.834
ADJDIST	1.035	1.015	1.055
NEAR	2.832	1.362	5.891
ADJ	9.182	1.738	48.519
CTRL	0.164	0.079	0.338
TTC	0.463	0.253	0.844

- AIC (327.088), SC (354.153), -2 Log L (313.088), R-Square (0.2916), Max-rescaled R-Square (0.4117)
- Sample Size: 108 Yields and 245 Non-yields
- The following variables were determined to be significant: SPD, ADJDIST, NEAR, ADJ, CTRL, and TTC. Higher speed decreases the likelihood of yielding, as do CTRL and TTC. The CTRL effect means less yielding to the experimenter, which may be the sign of some bias in the experiment. The TTC effect is opposite from the team's expectation, as greater TTC should lead to more yielding!
- ADJDIST, NEAR, and ADJ all increase the likelihood of yielding, which is as expected.

It can be seen above that each state shows different variables to be significant. Two dummy variables were then created, one with FLORIDA = 1 and one with ALABAMA = 1. The yielding in Florida and Alabama will thus be estimated relative to North Carolina.

Binary Logit (Y vs. NY, On-campus data only, forward selection) [Y-8]

In this next model, only the on-campus data were used in the model development. The results below show the forward selection model.

Table 33: Yield Model Y-8 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.3582	0.4484	9.1764	0.0025
SPD	1	-0.0860	0.0210	16.8284	<.0001
ADJ	1	1.0645	0.5188	4.2100	0.0402
PLT	1	0.8969	0.2693	11.0914	0.0009
HGV	1	-1.2984	0.4224	9.4473	0.0021
MUP	1	1.1235	0.3057	13.5068	0.0002
MED	1	1.1104	0.3187	12.1425	0.0005
FEMALE	1	0.6801	0.2466	7.6086	0.0058

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.918	0.881	0.956
ADJ	2.899	1.049	8.015
PLT	2.452	1.446	4.157
HGV	0.273	0.119	0.625
MUP	3.076	1.689	5.599
MED	3.036	1.626	5.669
FEMALE	1.974	1.218	3.201

- AIC(479.760), SC(512.634), -2 Log L (578.249), R-Square (0.2246), Max-rescaled R-Square (0.3105)
- Sample Size: 296 Yields and 154 Non-yields
- The following variables were determined to be significant: SPD, ADJ, PLT, HGV, MUP, MED, FEMALE
- SPD and HGV reduce the likelihood of yielding on campus, which is expected
- ADJ, PLT, MUP, MED, and FEMALE all increase the likelihood of yielding. The effects of ADJ, MUP, and MED are as expected. The effect of FEMALE is an interesting finding, suggesting more yielding to women on campus.
- The effects of PLT are once again counter to the team's expectation. A potential explanation is the low-speed campus environment, where platooning is a function of general vehicular congestion. In those cases, drivers may be more likely to yield, as they are already slow and delayed. It was determined that a low-speed platoon (LSPLT) variable should be added. This variable had a threshold of 15 mph, where we saw a difference in yielding decisions.
- Backward elimination showed slightly different results, but told a similar story as forward selection.

Binary Logit (Y vs. NY, Off-campus data only, forward selection) [Y-9]

In the next model, only the off-campus data were evaluated with regard to their effect on yielding in a forward selection model.

Table 34: Yield Model Y-9 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.4795	0.3848	14.7798	0.0001
SPD	1	-0.1553	0.0200	60.0913	<.0001
ADJDIST	1	0.00487	0.00192	6.4384	0.0112
MED	1	0.8093	0.3156	6.5733	0.0104

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.856	0.823	0.890
ADJDIST	1.005	1.001	1.009
MED	2.246	1.210	4.170

- AIC(527.985), SC(545.039), -2 Log L (519.985), R-Square (0.1424), Max-rescaled R-Square (0.2090)
- Sample Size: 136 Yields and 389 Non-yields
- The following variables were determined to be significant: SPD, ADJDIST, and MED
- SPD reduces the likelihood of yielding
- ADJDIST and MED increase the likelihood of yielding. All these effects are as expected
- Backward elimination showed slightly different results, but told a similar story as forward selection.

It can be seen above that the models for on-campus data only and off-campus data only show different variables to be significant. As a result, a new variable for CAMPUS=1 will be included in the model development.

Correlation Table (Y vs. NY, Campus and State variable added)

The following table shows a correlation analysis of the response variable (Y_NY = 1 for yield events), as a function of various independent variables. Each cell contains the correlation coefficient (ranges from -1 to +1) in the top row, and the p-value of the Chi-Square correlation test in the second row.

Table 35: Probability Yielding Correlation Table with Campus

	Y_NY	SPD	ADJDIST	NEAR	ADJ	PLT	HGV	MUP	MED	CTRL	TTC	DECEL	AGE	BUSINESS	FEMALE	CAMPUS	FLORIDA	ALABAMA
Y_NY	1.00000	-0.37698 <.0001	-0.02631 0.4119	-0.06577 0.0400	0.19581 <.0001	0.09812 0.0022	0.01653 0.6061	0.24610 <.0001	0.21079 <.0001	-0.14696 <.0001	0.21460 <.0001	-0.15638 <.0001	0.23039 <.0001	-0.06696 0.0366	0.10427 0.0011	0.40015 <.0001	0.43875 <.0001	-0.20796 <.0001
SPD	-0.37698 <.0001	1.00000	0.36158 <.0001	0.05869 0.0670	-0.15163 <.0001	-0.09315 0.0036	-0.14360 <.0001	-0.25649 <.0001	-0.02244 0.4840	0.31240 <.0001	-0.31806 <.0001	0.33970 <.0001	-0.24174 <.0001	-0.11683 0.0003	-0.25464 <.0001	-0.28481 <.0001	-0.20801 <.0001	0.44092 <.0001
ADJDIST	-0.02631 0.4119	0.36158 <.0001	1.00000	-0.09417 0.0032	-0.02610 0.4156	0.02504 0.4347	0.09996 0.0018	-0.09842 0.0021	0.09946 0.0019	0.01135 0.7233	0.70652 <.0001	-0.33511 <.0001	0.06402 0.0456	-0.02299 0.4733	-0.18821 <.0001	0.09719 0.0024	0.07179 0.0250	-0.01937 0.5458
NEAR	-0.06577 0.0400	0.05869 0.0670	-0.09417 0.0032	1.00000	-0.03318 0.3006	0.07319 0.0223	-0.02888 0.3677	-0.09352 0.0035	-0.54721 <.0001	0.30739 <.0001	-0.13092 <.0001	0.03724 0.2454	-0.01371 0.6690	-0.06026 0.0600	-0.02080 0.5165	-0.08528 0.0077	-0.10318 0.0013	-0.01056 0.7420
ADJ	0.19581 <.0001	-0.15163 <.0001	-0.02610 0.4156	-0.03318 0.3006	1.00000	0.06436 0.0445	-0.03734 0.2441	0.09436 0.0032	0.12620 <.0001	-0.00967 0.7629	0.08887 0.0055	-0.04453 0.1647	0.05329 0.0963	-0.00673 0.8338	0.08597 0.0072	0.10380 0.0012	0.18445 <.0001	-0.13648 <.0001
PLT	0.09812 0.0022	-0.09315 0.0036	0.02504 0.4347	0.07319 0.0223	0.06436 0.0445	1.00000	0.09491 0.0030	0.03982 0.2141	0.04096 0.2013	0.03208 0.3170	0.09995 0.0018	-0.10581 0.0009	0.00049 0.9878	0.00803 0.8022	0.02182 0.4962	-0.04166 0.1937	0.07451 0.0200	-0.09117 0.0044
HGV	0.01653 0.6061	-0.14360 <.0001	0.09996 0.0018	-0.02888 0.3677	-0.03734 0.2441	0.09491 0.0030	1.00000	0.12388 0.0001	-0.01339 0.6762	-0.10871 0.0007	0.21782 <.0001	-0.08961 0.0051	0.08209 0.0103	0.01157 0.7182	0.01172 0.7148	0.13167 <.0001	0.01264 0.6934	-0.12210 0.0001
MUP	0.24610 <.0001	-0.25649 <.0001	-0.09842 0.0021	-0.09352 0.0035	0.09436 0.0032	0.03982 0.2141	0.12388 0.0001	1.00000	0.01919 0.5495	-0.48996 <.0001	0.10353 0.0012	-0.09046 0.0047	0.03030 0.3447	0.03319 0.3005	0.03250 0.3106	0.22231 <.0001	0.18108 <.0001	-0.13844 <.0001
MED	0.21079 <.0001	-0.02244 0.4840	0.09946 0.0019	-0.54721 <.0001	0.12620 <.0001	0.04096 0.2013	-0.01339 0.6762	0.01919 0.5495	1.00000	0.05617 0.0796	0.11924 0.0002	0.03762 0.2405	0.22320 <.0001	-0.10065 0.0017	-0.02150 0.5026	0.14490 <.0001	0.41416 <.0001	-0.13052 <.0001
CTRL	-0.14696 <.0001	0.31240 <.0001	0.01135 0.7233	0.30739 <.0001	-0.00967 0.7629	0.03208 0.3170	-0.10871 0.0007	-0.48996 <.0001	0.05617 0.0796	1.00000	-0.21989 <.0001	0.12273 0.0001	0.11919 0.0002	-0.28967 <.0001	-0.02401 0.4539	-0.29794 <.0001	0.00700 0.8271	0.08439 0.0084
TTC	0.21460 <.0001	-0.31806 <.0001	0.70652 <.0001	-0.13092 <.0001	0.08887 0.0055	0.09995 0.0018	0.21782 <.0001	0.10353 0.0012	0.11924 0.0002	-0.21989 <.0001	1.00000	-0.55213 <.0001	0.19072 <.0001	0.05629 0.0790	-0.03702 0.2481	0.28831 <.0001	0.22051 <.0001	-0.30288 <.0001
DECEL	-0.15638 <.0001	0.33970 <.0001	-0.33511 <.0001	0.03724 0.2454	-0.04453 0.1647	-0.10581 0.0009	-0.08961 0.0051	-0.09046 0.0047	0.03762 0.2405	0.12273 0.0001	-0.55213 <.0001	1.00000	-0.03817 0.2338	-0.04083 0.2027	0.03580 0.2640	-0.09215 0.0040	0.12411 0.0001	0.01644 0.6082
AGE	0.23039 <.0001	-0.24174 <.0001	0.06402 0.0456	-0.01371 0.6690	0.05329 0.0963	0.00049 0.9878	0.08209 0.0103	0.03030 0.3447	0.22320 <.0001	0.11919 0.0002	0.19072 <.0001	-0.03817 0.2338	1.00000	-0.23479 <.0001	0.22034 <.0001	0.28402 <.0001	0.34292 <.0001	-0.48343 <.0001
BUSINESS	-0.06696 0.0366	-0.11683 0.0003	-0.02299 0.4733	-0.06026 0.0600	-0.00673 0.8338	0.00803 0.8022	0.01157 0.7182	0.03319 0.3005	-0.10065 0.0017	-0.28967 <.0001	0.05629 0.0790	-0.04083 0.2027	-0.23479 <.0001	1.00000	-0.03417 0.2865	-0.12197 0.0001	-0.15500 <.0001	-0.08807 0.0059
FEMALE	0.10427 0.0011	-0.25464 <.0001	-0.18821 <.0001	-0.02080 0.5165	0.08597 0.0072	0.02182 0.4962	0.01172 0.7148	0.03250 0.3106	-0.02150 0.5026	-0.02401 0.4539	-0.03702 0.2481	0.03580 0.2640	0.22034 <.0001	-0.03417 0.2865	1.00000	-0.09455 0.0031	-0.01725 0.5906	-0.38242 <.0001
CAMPUS	0.40015 <.0001	-0.28481 <.0001	0.09719 0.0024	-0.08528 0.0077	0.10380 0.0012	-0.04166 0.1937	0.13167 <.0001	0.22231 <.0001	0.14490 <.0001	-0.29794 <.0001	0.28831 <.0001	-0.09215 0.0040	0.28402 <.0001	-0.12197 0.0001	-0.09455 0.0031	1.00000	0.39022 <.0001	-0.13666 <.0001
FLORIDA	0.43875 <.0001	-0.20801 <.0001	0.07179 0.0250	-0.10318 0.0013	0.18445 <.0001	0.07451 0.0200	0.01264 0.6934	0.18108 <.0001	0.41416 <.0001	0.00700 0.8271	0.22051 <.0001	0.12411 0.0001	0.34292 <.0001	-0.15500 <.0001	-0.01725 0.5906	0.39022 <.0001	1.00000	-0.50223 <.0001
ALABAMA	-0.20796 <.0001	0.44092 <.0001	-0.01937 0.5458	-0.01056 0.7420	-0.13648 <.0001	-0.09117 0.0044	-0.12210 0.0001	-0.13844 <.0001	-0.13052 <.0001	0.08439 0.0084	-0.30288 <.0001	0.01644 0.6082	-0.48343 <.0001	-0.08807 0.0059	-0.38242 <.0001	-0.13666 <.0001	-0.50223 <.0001	1.00000

- The following variables show a significant positive correlation with the dependent variable, suggesting an increase in yielding with an increase in the variable (or binary variable change from 0 to 1): ADJ, PLT, MUP, MED, TTC, AGE, FEMALE, CAMPUS, and FLORIDA
- The following variables show a negative or inverse correlation (<-0.3) with the dependent variable, suggesting a decrease in yielding with an increase in the variable (or binary variable change from 0 to 1): SPD, NEAR, CTRL, DECEL, BUSINESS, and ALABAMA
- The following variables are intercorrelated: ADJDIST to TTC (0.70652), NEAR to MED (-0.54721), MUP to CTRL (-0.48996), TTC to DECEL (-0.55213), AGE to ALABAMA (-0.48343), FLORIDA to ALABAMA (-0.50223)
- It is expected that ADJDIST, TTC, and DECEL are intercorrelated since TTC and DECEL are calculated using the ADJDIST

Binary Logit (Y vs. NY, full model) [Y-10]

Table 36: Yield Model Y-10 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.6177	0.7481	0.6818	0.4090
SPD	1	-0.1601	0.0335	22.8712	<.0001
ADJDIST	1	0.00802	0.00421	3.6268	0.0569
NEAR	1	0.4102	0.2313	3.1453	0.0761
ADJ	1	0.9535	0.3341	8.1437	0.0043
PLT	1	0.3875	0.1819	4.5358	0.0332
HGV	1	-0.7065	0.4003	3.1152	0.0776
MUP	1	0.8386	0.2597	10.4301	0.0012
MED	1	0.5962	0.3067	3.7797	0.0519
CTRL	1	-0.2050	0.2444	0.7031	0.4017
TTC	1	-0.2301	0.1137	4.0932	0.0431
DECEL	1	-0.1356	0.0462	8.6109	0.0033
AGE	1	0.2783	0.2642	1.1094	0.2922
BUSINESS	1	0.1898	0.3966	0.2291	0.6322
FEMALE	1	0.8148	0.2020	16.2684	<.0001
CAMPUS	1	0.9738	0.2014	23.3709	<.0001
FLORIDA	1	2.4487	0.2822	75.2698	<.0001
ALABAMA	1	1.2733	0.2618	23.6518	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.852	0.798	0.910
ADJDIST	1.008	1.000	1.016
NEAR	1.507	0.958	2.372
ADJ	2.595	1.348	4.995
PLT	1.473	1.031	2.104
HGV	0.493	0.225	1.081
MUP	2.313	1.390	3.848
MED	1.815	0.995	3.311
CTRL	0.815	0.505	1.315
TTC	0.794	0.636	0.993
DECEL	0.873	0.798	0.956
AGE	1.321	0.787	2.217
BUSINESS	1.209	0.556	2.631
FEMALE	2.259	1.520	3.356
CAMPUS	2.648	1.784	3.930
FLORIDA	11.573	6.656	20.123
ALABAMA	3.573	2.139	5.968

- AIC (932.495), SC (1020.379), -2 Log L (896.495), R-Square (0.3648), Max-rescaled R-Square (0.4885)
- This model shows the following variables to be significant: SPD, ADJDIST, NEAR, ADJ, PLT, MUP, TTC, DECEL, FEMALE, CAMPUS, FLORIDA, and ALABAMA
- The following variables are shown to be insignificant: HGV, MED, CTRL, AGE, and BUSINESS
- It is expected that forward selection and backward elimination will show the same variables to be significant and insignificant.

<u>Binary Logit (Y vs. NY, forward selection) [Y-11]</u>

Table 37: Yield Model Y-11 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.0433	0.3731	0.0134	0.9077
SPD	1	-0.1093	0.0167	42.8224	<.0001
ADJ	1	0.9211	0.3283	7.8732	0.0050
PLT	1	0.4120	0.1790	5.2967	0.0214
HGV	1	-0.8247	0.3917	4.4328	0.0353
MUP	1	0.8055	0.2294	12.3267	0.0004
DECEL	1	-0.1332	0.0387	11.8210	0.0006
FEMALE	1	0.7987	0.1927	17.1741	<.0001
CAMPUS	1	1.0681	0.1782	35.9425	<.0001
FLORIDA	1	2.5596	0.2540	101.5453	<.0001
ALABAMA	1	1.1337	0.2254	25.2927	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.897	0.868	0.926
ADJ	2.512	1.320	4.780
PLT	1.510	1.063	2.145
HGV	0.438	0.203	0.945
MUP	2.238	1.427	3.509
DECEL	0.875	0.811	0.944
FEMALE	2.223	1.523	3.243
CAMPUS	2.910	2.052	4.126
FLORIDA	12.930	7.860	21.273
ALABAMA	3.107	1.998	4.834

- AIC (928.549), SC (982.256), -2 Log L (906.549), R-Square (0.3582), Max-rescaled R-Square (0.4797)
- The following variables were determined to be significant: SPD, ADJ, PLT, HGV, MUP, DECEL, FEMALE, CAMPUS, FLORIDA, and ALABAMA
- Increased speed reduces chance of yielding, as does increased required deceleration rate.
- Adjacent yield, presence of multiple pedestrians, and female pedestrians increase the chance of yielding
- The coefficient estimates for Florida and Alabama show that drivers are approximately three times as likely to yield in Florida and 1.5 times more likely to yield in Alabama than drivers in North Carolina.
- These variables were shown to be insignificant: ADJDIST, NEAR, MED, CTRL, TTC, AGE
- Backward elimination showed slightly different results, but told a similar story as forward selection.

Correlation Table (Y vs. NY, low speed platoon variable added)

The following table shows a correlation analysis of the response variable (Y_NY = 1 for yield events), as a function of various independent variables. Each cell contains the correlation coefficient (ranges from -1 to +1) in the top row, and the p-value of the Chi-Square correlation test in the second row.

Table 38: Probability of Yielding Correlation Table with low-speed platoon

	Y_NY	SPD	ADJDIST	NEAR	ADJ	PLT	LSPLT	HGV	MUP	MED	CTRL	TTC	DECEL	AGE	BUSINESS	FEMALE	CAMPUS	FLORIDA	ALABAMA
Y_NY	1.00000	-0.37698 <.0001	-0.02631 0.4119	-0.06577 0.0400	0.19581 <.0001	0.09812 0.0022	0.22912 <.0001	0.01653 0.6061	0.24610 <.0001	0.21079 <.0001	-0.14696 <.0001	0.21460 <.0001	-0.15638 <.0001	0.23039 <.0001	-0.06696 0.0366	0.10427 0.0011	0.40015 <.0001	0.43875 <.0001	-0.20796 <.0001
SPD	-0.37698 <.0001	1.00000	0.36158 <.0001	0.05869 0.0670	-0.15163 <.0001	-0.09315 0.0036	-0.40952 <.0001	-0.14360 <.0001	-0.25649 <.0001	-0.02244 0.4840	0.31240 <.0001	-0.31806 <.0001	0.33970 <.0001	-0.24174 <.0001	-0.11683 0.0003	-0.25464 <.0001	-0.28481 <.0001	-0.20801 <.0001	0.44092 <.0001
ADJDIST	-0.02631 0.4119	0.36158 <.0001	1.00000	-0.09417 0.0032	-0.02610 0.4156	0.02504 0.4347	-0.16035 <.0001	0.09996 0.0018	-0.09842 0.0021	0.09946 0.0019	0.01135 0.7233	0.70652 <.0001	-0.33511 <.0001	0.06402 0.0456	-0.02299 0.4733	-0.18821 <.0001	0.09719 0.0024	0.07179 0.0250	-0.01937 0.5458
NEAR	-0.06577 0.0400	0.05869 0.0670	-0.09417 0.0032	1.00000	-0.03318 0.3006	0.07319 0.0223	0.06364 0.0470	-0.02888 0.3677	-0.09352 0.0035	-0.54721 <.0001	0.30739 <.0001	-0.13092 <.0001	0.03724 0.2454	-0.01371 0.6690	-0.06026 0.0600	-0.02080 0.5165	-0.08528 0.0077	-0.10318 0.0013	-0.01056 0.7420
ADJ	0.19581 <.0001	-0.15163 <.0001	-0.02610 0.4156	-0.03318 0.3006	1.00000	0.06436 0.0445	0.10513 0.0010	-0.03734 0.2441	0.09436 0.0032	0.12620 <.0001	-0.00967 0.7629	0.08887 0.0055	-0.04453 0.1647	0.05329 0.0963	-0.00673 0.8338	0.08597 0.0072	0.10380 0.0012	0.18445 <.0001	-0.13648 <.0001
PLT	0.09812 0.0022	-0.09315 0.0036	0.02504 0.4347	0.07319 0.0223	0.06436 0.0445	1.00000	0.45414 <.0001	0.09491 0.0030	0.03982 0.2141	0.04096 0.2013	0.03208 0.3170	0.09995 0.0018	-0.10581 0.0009	0.00049 0.9878	0.00803 0.8022	0.02182 0.4962	-0.04166 0.1937	0.07451 0.0200	-0.09117 0.0044
LSPLT	0.22912 <.0001	-0.40952 <.0001	-0.16035 <.0001	0.06364 0.0470	0.10513 0.0010	0.45414 <.0001	1.00000	0.13305 <.0001	0.19131 <.0001	0.02624 0.4132	-0.14722 <.0001	0.17986 <.0001	-0.17452 <.0001	0.06698 0.0365	0.03337 0.2979	0.07596 0.0177	0.05657 0.0775	0.12566 <.0001	-0.19715 <.0001
HGV	0.01653 0.6061	-0.14360 <.0001	0.09996 0.0018	-0.02888 0.3677	-0.03734 0.2441	0.09491 0.0030	0.13305 <.0001	1.00000	0.12388 0.0001	-0.01339 0.6762	-0.10871 0.0007	0.21782 <.0001	-0.08961 0.0051	0.08209 0.0103	0.01157 0.7182	0.01172 0.7148	0.13167 <.0001	0.01264 0.6934	-0.12210 0.0001
MUP	0.24610 <.0001	-0.25649 <.0001	-0.09842 0.0021	-0.09352 0.0035	0.09436 0.0032	0.03982 0.2141	0.19131 <.0001	0.12388 0.0001	1.00000	0.01919 0.5495	-0.48996 <.0001	0.10353 0.0012	-0.09046 0.0047	0.03030 0.3447	0.03319 0.3005	0.03250 0.3106	0.22231 <.0001	0.18108 <.0001	-0.13844 <.0001
MED	0.21079 <.0001	-0.02244 0.4840	0.09946 0.0019	-0.54721 <.0001	0.12620 <.0001	0.04096 0.2013	0.02624 0.4132	-0.01339 0.6762	0.01919 0.5495	1.00000	0.05617 0.0796	0.11924 0.0002	0.03762 0.2405	0.22320 <.0001	-0.10065 0.0017	-0.02150 0.5026	0.14490 <.0001	0.41416 <.0001	-0.13052 <.0001
CTRL	-0.14696 <.0001	0.31240 <.0001	0.01135 0.7233	0.30739 <.0001	-0.00967 0.7629	0.03208 0.3170	-0.14722 <.0001	-0.10871 0.0007	-0.48996 <.0001	0.05617 0.0796	1.00000	-0.21989 <.0001	0.12273 0.0001	0.11919 0.0002	-0.28967 <.0001	-0.02401 0.4539	-0.29794 <.0001	0.00700 0.8271	0.08439 0.0084
TTC	0.21460 <.0001	-0.31806 <.0001	0.70652 <.0001	-0.13092 <.0001	0.08887 0.0055	0.09995 0.0018	0.17986 <.0001	0.21782 <.0001	0.10353 0.0012	0.11924 0.0002	-0.21989 <.0001	1.00000	-0.55213 <.0001	0.19072 <.0001	0.05629 0.0790	-0.03702 0.2481	0.28831 <.0001	0.22051 <.0001	-0.30288 <.0001
DECEL	-0.15638 <.0001	0.33970 <.0001	-0.33511 <.0001	0.03724 0.2454	-0.04453 0.1647	-0.10581 0.0009	-0.17452 <.0001	-0.08961 0.0051	-0.09046 0.0047	0.03762 0.2405	0.12273 0.0001	-0.55213 <.0001	1.00000	-0.03817 0.2338	-0.04083 0.2027	0.03580 0.2640	-0.09215 0.0040	0.12411 0.0001	0.01644 0.6082
AGE	0.23039 <.0001	-0.24174 <.0001	0.06402 0.0456	-0.01371 0.6690	0.05329 0.0963	0.00049 0.9878	0.06698 0.0365	0.08209 0.0103	0.03030 0.3447	0.22320 <.0001	0.11919 0.0002	0.19072 <.0001	-0.03817 0.2338	1.00000	-0.23479 <.0001	0.22034 <.0001	0.28402 <.0001	0.34292 <.0001	-0.48343 <.0001
BUSINESS	-0.06696 0.0366	-0.11683 0.0003	-0.02299 0.4733	-0.06026 0.0600	-0.00673 0.8338	0.00803 0.8022	0.03337 0.2979	0.01157 0.7182	0.03319 0.3005	-0.10065 0.0017	-0.28967 <.0001	0.05629 0.0790	-0.04083 0.2027	-0.23479 <.0001	1.00000	-0.03417 0.2865	-0.12197 0.0001	-0.15500 <.0001	-0.08807 0.0059
FEMALE	0.10427 0.0011	-0.25464 <.0001	-0.18821 <.0001	-0.02080 0.5165	0.08597 0.0072	0.02182 0.4962	0.07596 0.0177	0.01172 0.7148	0.03250 0.3106	-0.02150 0.5026	-0.02401 0.4539	-0.03702 0.2481	0.03580 0.2640	0.22034 <.0001	-0.03417 0.2865	1.00000	-0.09455 0.0031	-0.01725 0.5906	-0.38242 <.0001
CAMPUS	0.40015 <.0001	-0.28481 <.0001	0.09719 0.0024	-0.08528 0.0077	0.10380 0.0012	-0.04166 0.1937	0.05657 0.0775	0.13167 <.0001	0.22231 <.0001	0.14490 <.0001	-0.29794 <.0001	0.28831 <.0001	-0.09215 0.0040	0.28402 <.0001	-0.12197 0.0001	-0.09455 0.0031	1.00000	0.39022 <.0001	-0.13666 <.0001
FLORIDA	0.43875 <.0001	-0.20801 <.0001	0.07179 0.0250	-0.10318 0.0013	0.18445 <.0001	0.07451 0.0200	0.12566 <.0001	0.01264 0.6934	0.18108 <.0001	0.41416 <.0001	0.00700 0.8271	0.22051 <.0001	0.12411 0.0001	0.34292 <.0001	-0.15500 <.0001	-0.01725 0.5906	0.39022 <.0001	1.00000	-0.50223 <.0001
ALABAMA	-0.20796 <.0001	0.44092 <.0001	-0.01937 0.5458	-0.01056 0.7420	-0.13648 <.0001	-0.09117 0.0044	-0.19715 <.0001	-0.12210 0.0001	-0.13844 <.0001	-0.13052 <.0001	0.08439 0.0084	-0.30288 <.0001	0.01644 0.6082	-0.48343 <.0001	-0.08807 0.0059	-0.38242 <.0001	-0.13666 <.0001	-0.50223 <.0001	1.00000

- The following variables show a significant positive correlation with the dependent variable, suggesting an increase in yielding with an increase in the variable (or binary variable change from 0 to 1): ADJ, PLT, LSPLT, MUP, MED, TTC, AGE, FEMALE, CAMPUS, and FLORIDA
- The following variables show a significant negative or inverse correlation with the dependent variable, suggesting a decrease in yielding with an increase in the variable (or binary variable change from 0 to 1): SPD, NEAR, CTRL, DECEL, BUSINESS, and ALABAMA
- The following variables are intercorrelated: ADJDIST to TTC (0.70652), NEAR to MED (-0.54721), MUP to CTRL (-0.48996), TTC to DECEL (-0.55213), AGE to ALABAMA (-0.48343), FLORIDA to ALABAMA (-0.50223)
- It is expected that ADJDIST, TTC, and DECEL are intercorrelated since TTC and DECEL are calculated using the ADJDIST

Binary Logit (Y vs. NY, full model) [Y-12]

Table 39: Yield Model Y-12 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.4326	0.7582	0.3256	0.5683
SPD	1	-0.1536	0.0336	20.8857	<.0001
ADJDIST	1	0.00956	0.00428	4.9766	0.0257
NEAR	1	0.3508	0.2327	2.2732	0.1316
ADJ	1	0.9433	0.3369	7.8398	0.0051
PLT	1	0.1629	0.2032	0.6429	0.4227
LSPLT	1	0.9654	0.3811	6.4158	0.0113
HGV	1	-0.7545	0.4071	3.4341	0.0639
MUP	1	0.8336	0.2614	10.1678	0.0014
MED	1	0.5690	0.3067	3.4417	0.0636
CTRL	1	-0.1772	0.2457	0.5203	0.4707
TTC	1	-0.2721	0.1168	5.4243	0.0199
DECEL	1	-0.1357	0.0464	8.5723	0.0034
AGE	1	0.3036	0.2648	1.3152	0.2515
BUSINESS	1	0.2330	0.4025	0.3351	0.5627
FEMALE	1	0.8296	0.2031	16.6753	<.0001
CAMPUS	1	1.0273	0.2023	25.7860	<.0001
FLORIDA	1	2.4684	0.2828	76.1639	<.0001
ALABAMA	1	1.3077	0.2624	24.8354	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.858	0.803	0.916
ADJDIST	1.010	1.001	1.018
NEAR	1.420	0.900	2.241
ADJ	2.568	1.327	4.971
PLT	1.177	0.790	1.753
LSPLT	2.626	1.244	5.542
HGV	0.470	0.212	1.044
MUP	2.302	1.379	3.842
MED	1.766	0.968	3.222
CTRL	0.838	0.518	1.356
TTC	0.762	0.606	0.958
DECEL	0.873	0.797	0.956
AGE	1.355	0.806	2.276
BUSINESS	1.262	0.574	2.778
FEMALE	2.292	1.539	3.413
CAMPUS	2.794	1.879	4.153
FLORIDA	11.803	6.780	20.548
ALABAMA	3.698	2.211	6.184

- AIC (927.883), SC (1020.649), -2 Log L (889.883), R-Square (0.3691), Max-rescaled R-Square (0.4943)
- This model shows the following variables to be significant: SPD, ADJDIST, ADJ, LSPLT, MUP, TTC, DECEL, FEMALE, CAMPUS, FLORIDA, and ALABAMA
- The following variables are shown to be insignificant: NEAR, PLT, HGV, MED, CTRL, AGE, and BUSINESS
- It is expected that forward selection and backward elimination will show the same variables to be significant and insignificant.

Binary Logit (Y vs. NY, forward selection) [Y-13]

Table 40: Yield Model Y-13 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.2879	0.3977	0.5239	0.4692
SPD	1	-0.0913	0.0173	27.6875	<.0001
ADJ	1	0.9113	0.3302	7.6156	0.0058
LSPLT	1	1.0428	0.3393	9.4476	0.0021
HGV	1	-0.8461	0.3987	4.5032	0.0338
MUP	1	0.7766	0.2296	11.4440	0.0007
DECEL	1	-0.1392	0.0394	12.4733	0.0004
FEMALE	1	0.8138	0.1929	17.7995	<.0001
CAMPUS	1	1.1059	0.1787	38.2893	<.0001
FLORIDA	1	2.6016	0.2553	103.8132	<.0001
ALABAMA	1	1.1514	0.2262	25.9159	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.913	0.882	0.944
ADJ	2.488	1.302	4.752
LSPLT	2.837	1.459	5.517
HGV	0.429	0.196	0.937
MUP	2.174	1.386	3.409
DECEL	0.870	0.805	0.940
FEMALE	2.256	1.546	3.293
CAMPUS	3.022	2.129	4.290
FLORIDA	13.485	8.175	22.242
ALABAMA	3.163	2.030	4.927

- AIC (923.934), SC (977.641), -2 Log L (901.934), R-Square (0.3613), Max-rescaled R-Square (0.4838)
- The following variables were determined to be significant: SPD, ADJ, LSPLT, HGV, MUP, DECEL, FEMALE, CAMPUS, FLORIDA, and ALABAMA
- Increased speed reduces chance of yielding, as does increased necessary deceleration rate.
- Adjacent yield, low speed platoons, presence of multiple pedestrians, student age pedestrians, and female pedestrians increase the chance of yielding
- The coefficient estimates for Florida and Alabama show that drivers are approximately three times as likely to yield in Florida and 1.5 times more likely to yield in Alabama than drivers in North Carolina.
- These variables were shown to be insignificant: ADJDIST, NEAR, PLT, MED, CTRL, TTC, AGE, and BUSINESS
- Backward elimination showed additional variables to be significant

<u>Correlation Table (Y vs. NY, State variables changed)</u>

The following table shows a correlation analysis of the response variable (Y_NY = 1 for yield events), as a function of various independent variables. Each cell contains the correlation coefficient (ranges from -1 to +1) in the top row, and the p-value of the Chi-Square correlation test in the second row. North Carolina was originally used as the intercept value, but it was decided that since most of the observations in North Carolina were female pedestrians, a different state with more balanced observations should be used as the intercept (Florida was chosen).

Table 41: Probability of Yielding Correlation Table with new State variables

	Y_NY	SPD	ADJDIST	NEAR	ADJ	PLT	LSPLT	HGV	MUP	MED	CTRL	TTC	DECEL	AGE	BUSINESS	FEMALE	CAMPUS	ALABAMA	NCAROLINA
Y_NY	1.00000	-0.37698 <.0001	-0.02631 0.4119	-0.06577 0.0400	0.19581 <.0001	0.09812 0.0022	0.22912 <.0001	0.01653 0.6061	0.24610 <.0001	0.21079 <.0001	-0.14696 <.0001	0.21460 <.0001	-0.15638 <.0001	0.23039 <.0001	-0.06696 0.0366	0.10427 0.0011	0.40015 <.0001	-0.20796 <.0001	-0.21806 <.0001
SPD	-0.37698 <.0001	1.00000	0.36158 <.0001	0.05869 0.0670	-0.15163 <.0001	-0.09315 0.0036	-0.40952 <.0001	-0.14360 <.0001	-0.25649 <.0001	-0.02244 0.4840	0.31240 <.0001	-0.31806 <.0001	0.33970 <.0001	-0.24174 <.0001	-0.11683 0.0003	-0.25464 <.0001	-0.28481 <.0001	0.44092 <.0001	-0.24644 <.0001
ADJDIST	-0.02631 0.4119	0.36158 <.0001	1.00000	-0.09417 0.0032	-0.02610 0.4156	0.02504 0.4347	-0.16035 <.0001	0.09996 0.0018	-0.09842 0.0021	0.09946 0.0019	0.01135 0.7233	0.70652 <.0001	-0.33511 <.0001	0.06402 0.0456	-0.02299 0.4733	-0.18821 <.0001	0.09719 0.0024	-0.01937 0.5458	-0.05066 0.1139
NEAR	-0.06577 0.0400	0.05869 0.0670	-0.09417 0.0032	1.00000	-0.03318 0.3006	0.07319 0.0223	0.06364 0.0470	-0.02888 0.3677	-0.09352 0.0035	-0.54721 <.0001	0.30739 <.0001	-0.13092 <.0001	0.03724 0.2454	-0.01371 0.6690	-0.06026 0.0600	-0.02080 0.5165	-0.08528 0.0077	-0.01056 0.7420	0.11205 0.0005
ADJ	0.19581 <.0001	-0.15163 <.0001	-0.02610 0.4156	-0.03318 0.3006	1.00000	0.06436 0.0445	0.10513 0.0010	-0.03734 0.2441	0.09436 0.0032	0.12620 <.0001	-0.00967 0.7629	0.08887 0.0055	-0.04453 0.1647	0.05329 0.0963	-0.00673 0.8338	0.08597 0.0072	0.10380 0.0012	-0.13648 <.0001	-0.04154 0.1950
PLT	0.09812 0.0022	-0.09315 0.0036	0.02504 0.4347	0.07319 0.0223	0.06436 0.0445	1.00000	0.45414 <.0001	0.09491 0.0030	0.03982 0.2141	0.04096 0.2013	0.03208 0.3170	0.09995 0.0018	-0.10581 0.0009	0.00049 0.9878	0.00803 0.8022	0.02182 0.4962	-0.04166 0.1937	-0.09117 0.0044	0.02004 0.5320
LSPLT	0.22912 <.0001	-0.40952 <.0001	-0.16035 <.0001	0.06364 0.0470	0.10513 0.0010	0.45414 <.0001	1.00000	0.13305 <.0001	0.19131 <.0001	0.02624 0.4132	-0.14722 <.0001	0.17986 <.0001	-0.17452 <.0001	0.06698 0.0365	0.03337 0.2979	0.07596 0.0177	0.05657 0.0775	-0.19715 <.0001	0.07814 0.0147
HGV	0.01653 0.6061	-0.14360 <.0001	0.09996 0.0018	-0.02888 0.3677	-0.03734 0.2441	0.09491 0.0030	0.13305 <.0001	1.00000	0.12388 0.0001	-0.01339 0.6762	-0.10871 0.0007	0.21782 <.0001	-0.08961 0.0051	0.08209 0.0103	0.01157 0.7182	0.01172 0.7148	0.13167 <.0001	-0.12210 0.0001	0.11237 0.0004
MUP	0.24610 <.0001	-0.25649 <.0001	-0.09842 0.0021	-0.09352 0.0035	0.09436 0.0032	0.03982 0.2141	0.19131 <.0001	0.12388 0.0001	1.00000	0.01919 0.5495	-0.48996 <.0001	0.10353 0.0012	-0.09046 0.0047	0.03030 0.3447	0.03319 0.3005	0.03250 0.3106	0.22231 <.0001	-0.13844 <.0001	-0.03623 0.2584
MED	0.21079 <.0001	-0.02244 0.4840	0.09946 0.0019	-0.54721 <.0001	0.12620 <.0001	0.04096 0.2013	0.02624 0.4132	-0.01339 0.6762	0.01919 0.5495	1.00000	0.05617 0.0796	0.11924 0.0002	0.03762 0.2405	0.22320 <.0001	-0.10065 0.0017	-0.02150 0.5026	0.14490 <.0001	-0.13052 <.0001	-0.27307 <.0001
CTRL	-0.14696 <.0001	0.31240 <.0001	0.01135 0.7233	0.30739 <.0001	-0.00967 0.7629	0.03208 0.3170	-0.14722 <.0001	-0.10871 0.0007	-0.48996 <.0001	0.05617 0.0796	1.00000	-0.21989 <.0001	0.12273 0.0001	0.11919 0.0002	-0.28967 <.0001	-0.02401 0.4539	-0.29794 <.0001	0.08439 0.0084	-0.09311 0.0036
TTC	0.21460 <.0001	-0.31806 <.0001	0.70652 <.0001	-0.13092 <.0001	0.08887 0.0055	0.09995 0.0018	0.17986 <.0001	0.21782 <.0001	0.10353 0.0012	0.11924 0.0002	-0.21989 <.0001	1.00000	-0.55213 <.0001	0.19072 <.0001	0.05629 0.0790	-0.03702 0.2481	0.28831 <.0001	-0.30288 <.0001	0.09311 0.0036
DECEL	-0.15638 <.0001	0.33970 <.0001	-0.33511 <.0001	0.03724 0.2454	-0.04453 0.1647	-0.10581 0.0009	-0.17452 <.0001	-0.08961 0.0051	-0.09046 0.0047	0.03762 0.2405	0.12273 0.0001	-0.55213 <.0001	1.00000	-0.03817 0.2338	-0.04083 0.2027	0.03580 0.2640	-0.09215 0.0040	0.01644 0.6082	-0.13859 <.0001
AGE	0.23039 <.0001	-0.24174 <.0001	0.06402 0.0456	-0.01371 0.6690	0.05329 0.0963	0.00049 0.9878	0.06698 0.0365	0.08209 0.0103	0.03030 0.3447	0.22320 <.0001	0.11919 0.0002	0.19072 <.0001	-0.03817 0.2338	1.00000	-0.23479 <.0001	0.22034 <.0001	0.28402 <.0001	-0.48343 <.0001	0.15748 <.0001
BUSINESS	-0.06696 0.0366	-0.11683 0.0003	-0.02299 0.4733	-0.06026 0.0600	-0.00673 0.8338	0.00803 0.8022	0.03337 0.2979	0.01157 0.7182	0.03319 0.3005	-0.10065 0.0017	-0.28967 <.0001	0.05629 0.0790	-0.04083 0.2027	-0.23479 <.0001	1.00000	-0.03417 0.2865	-0.12197 0.0001	-0.08807 0.0059	0.24210 <.0001
FEMALE	0.10427 0.0011	-0.25464 <.0001	-0.18821 <.0001	-0.02080 0.5165	0.08597 0.0072	0.02182 0.4962	0.07596 0.0177	0.01172 0.7148	0.03250 0.3106	-0.02150 0.5026	-0.02401 0.4539	-0.03702 0.2481	0.03580 0.2640	0.22034 <.0001	-0.03417 0.2865	1.00000	-0.09455 0.0031	-0.38242 <.0001	0.40772 <.0001
CAMPUS	0.40015 <.0001	-0.28481 <.0001	0.09719 0.0024	-0.08528 0.0077	0.10380 0.0012	-0.04166 0.1937	0.05657 0.0775	0.13167 <.0001	0.22231 <.0001	0.14490 <.0001	-0.29794 <.0001	0.28831 <.0001	-0.09215 0.0040	0.28402 <.0001	-0.12197 0.0001	-0.09455 0.0031	1.00000	-0.13666 <.0001	-0.24329 <.0001
ALABAMA	-0.20796 <.0001	0.44092 <.0001	-0.01937 0.5458	-0.01056 0.7420	-0.13648 <.0001	-0.09117 0.0044	-0.19715 <.0001	-0.12210 0.0001	-0.13844 <.0001	-0.13052 <.0001	0.08439 0.0084	-0.30288 <.0001	0.01644 0.6082	-0.48343 <.0001	-0.08807 0.0059	-0.38242 <.0001	-0.13666 <.0001	1.00000	-0.52901 <.0001
NCAROLINA	-0.21806 <.0001	-0.24644 <.0001	-0.05066 0.1139	0.11205 0.0005	-0.04154 0.1950	0.02004 0.5320	0.07814 0.0147	0.11237 0.0004	-0.03623 0.2584	-0.27307 <.0001	-0.09311 0.0036	0.09311 0.0036	-0.13859 <.0001	0.15748 <.0001	0.24210 <.0001	0.40772 <.0001	-0.24329 <.0001	-0.52901 <.0001	1.00000

- The following variables show a significant positive correlation with the dependent variable, suggesting an increase in yielding with an increase in the variable (or binary variable change from 0 to 1): ADJ, PLT, LSPLT, MUP, MED, TTC, AGE, FEMALE, and CAMPUS
- The following variables show a significant negative or inverse correlation with the dependent variable, suggesting a decrease in yielding with an increase in the variable (or binary variable change from 0 to 1): SPD, NEAR, CTRL, DECEL, BUSINESS, ALABAMA, and NCAROLINA
- The following variables are intercorrelated: ADJDIST to TTC (0.70652), NEAR to MED (-0.54721), PLT to LSPLT (0.45414), MUP to CTRL (-0.48996), TTC to DECEL (-0.55213), AGE to ALABAMA (-0.48343), ALABAMA to NCAROLINA (-0.52901)
- It is expected that ADJDIST, TTC, and DECEL are intercorrelated since TTC and DECEL are calculated using the ADJDIST

Binary Logit (Y vs. NY, full model) [Y-14]

Table 42: Yield Model Y-14 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.9010	0.7996	13.1641	0.0003
SPD	1	-0.1536	0.0336	20.8857	<.0001
ADJDIST	1	0.00956	0.00428	4.9766	0.0257
NEAR	1	0.3508	0.2327	2.2732	0.1316
ADJ	1	0.9433	0.3369	7.8398	0.0051
PLT	1	0.1629	0.2032	0.6429	0.4227
LSPLT	1	0.9654	0.3811	6.4158	0.0113
HGV	1	-0.7545	0.4071	3.4341	0.0639
MUP	1	0.8336	0.2614	10.1678	0.0014
MED	1	0.5690	0.3067	3.4417	0.0636
CTRL	1	-0.1772	0.2457	0.5203	0.4707
TTC	1	-0.2721	0.1168	5.4243	0.0199
DECEL	1	-0.1357	0.0464	8.5723	0.0034
AGE	1	0.3036	0.2648	1.3152	0.2515
BUSINESS	1	0.2330	0.4025	0.3351	0.5627
FEMALE	1	0.8296	0.2031	16.6753	<.0001
CAMPUS	1	1.0273	0.2023	25.7860	<.0001
ALABAMA	1	-1.1607	0.2829	16.8351	<.0001
NCAROLINA	1	-2.4684	0.2828	76.1639	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.858	0.803	0.916
ADJDIST	1.010	1.001	1.018
NEAR	1.420	0.900	2.241
ADJ	2.568	1.327	4.971
PLT	1.177	0.790	1.753
LSPLT	2.626	1.244	5.542
HGV	0.470	0.212	1.044
MUP	2.302	1.379	3.842
MED	1.766	0.968	3.222
CTRL	0.838	0.518	1.356
TTC	0.762	0.606	0.958
DECEL	0.873	0.797	0.956
AGE	1.355	0.806	2.276
BUSINESS	1.262	0.574	2.778
FEMALE	2.292	1.539	3.413
CAMPUS	2.794	1.879	4.153
ALABAMA	0.313	0.180	0.545
NCAROLINA	0.085	0.049	0.147

- AIC (927.883), SC (1020.649), -2 Log L (889.883), R-Square (0.3691), Max-rescaled R-Square (0.4943)
- This model shows the following variables to be significant: SPD, ADJDIST, ADJ, LSPLT, MUP, TTC, DECEL, FEMALE, CAMPUS, ALABAMA, and NCAROLINA
- The following variables are shown to be insignificant: NEAR, PLT, HGV, MED, CTRL, AGE, and BUSINESS
- It is expected that forward selection and backward elimination will show the same variables to be significant and insignificant.

Binary Logit (Y vs. NY, forward selection) [Y-15]

Table 43: Yield Model Y-15 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.3137	0.4395	27.7176	<.0001
SPD	1	-0.0913	0.0173	27.6875	<.0001
ADJ	1	0.9113	0.3302	7.6156	0.0058
LSPLT	1	1.0428	0.3393	9.4476	0.0021
HGV	1	-0.8461	0.3987	4.5032	0.0338
MUP	1	0.7766	0.2296	11.4440	0.0007
DECEL	1	-0.1392	0.0394	12.4733	0.0004
FEMALE	1	0.8138	0.1929	17.7995	<.0001
CAMPUS	1	1.1059	0.1787	38.2893	<.0001
ALABAMA	1	-1.4502	0.2238	41.9828	<.0001
NCAROLINA	1	-2.6016	0.2553	103.8132	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.913	0.882	0.944
ADJ	2.488	1.302	4.752
LSPLT	2.837	1.459	5.517
HGV	0.429	0.196	0.937
MUP	2.174	1.386	3.409
DECEL	0.870	0.805	0.940
FEMALE	2.256	1.546	3.293
CAMPUS	3.022	2.129	4.290
ALABAMA	0.235	0.151	0.364
NCAROLINA	0.074	0.045	0.122

- AIC (923.934), SC (977.641), -2 Log L (901.934), R-Square (0.3613), Max-rescaled R-Square (0.4838)
- The following variables were determined to be significant: SPD, ADJ, LSPLT, HGV, MUP, DECEL, FEMALE, CAMPUS, ALABAMA, and NCAROLINA
- Increased speed reduces chance of yielding, as does increased necessary deceleration rate.
- Adjacent yield, low speed platoons, presence of multiple pedestrians, and female pedestrians increase the chance of yielding
- The coefficient estimates for Alabama and North Carolina show that drivers are approximately three times less likely to yield in North Carolina and 1.5 times less likely to yield in Alabama than drivers in Florida.
- These variables were shown to be insignificant: ADJDIST, NEAR, PLT, MED, CTRL, TTC, AGE, and BUSINESS
- Backward elimination showed additional variables to be significant

Binary Logit Manual Selection 3 (Y vs. NY, Y-2 but switch out age for campus) [Y-16]

Table 44: Yield Model Y-16 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.5790	0.5987	6.9570	0.0083
SPD	1	-0.1747	0.0285	37.5563	<.0001
ADJDIST	1	0.00929	0.00367	6.4137	0.0113
NEAR	1	0.5648	0.1993	8.0316	0.0046
ADJ	1	1.0049	0.3031	10.9918	0.0009
HGV	1	-0.8907	0.3563	6.2486	0.0124
MUP	1	0.9725	0.2213	19.3147	<.0001
MED	1	1.4413	0.2517	32.7802	<.0001
TTC	1	-0.1970	0.1007	3.8243	0.0505
BUSINESS	1	-0.6223	0.3600	2.9883	0.0839
CAMPUS	1	1.2481	0.1625	58.9818	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.840	0.794	0.888
ADJDIST	1.009	1.002	1.017
NEAR	1.759	1.190	2.600
ADJ	2.732	1.508	4.948
HGV	0.410	0.204	0.825
MUP	2.645	1.714	4.081
MED	4.226	2.580	6.922
TTC	0.821	0.674	1.000
BUSINESS	0.537	0.265	1.087
CAMPUS	3.484	2.534	4.791

- AIC (1017.789), SC (1071.496), -2 Log L (995.789), R-Square (0.2967), Max-rescaled R-Square (0.3973)
- The following variables were determined to be significant: SPD, ADJDIST, NEAR, ADJ, HGV, MUP, MED, and CAMPUS

Correlation Table (Y vs. NY, Alabama as intercept)

The following table shows a correlation analysis of the response variable (Y_NY = 1 for yield events), as a function of various independent variables. Each cell contains the correlation coefficient (ranges from -1 to +1) in the top row, and the p-value of the Chi-Square correlation test in the second row. Alabama was chosen as the intercept, since it has the middle level of yielding for the three sites (Florida has greater yielding and North Carolina has less yielding).

Table 45: Probability of Yielding Correlation with Alabama Intercept

Pearson Correlation Coefficients, N = 975
Prob > |r| under H0: Rho=0

	Y_NY	SPD	ADJDIST	NEAR	ADJ	PLT	LSPLT	HGV	MUP	MED	CTRL	TTC	DECEL	AGE	BUSINESS	FEMALE	CAMPUS	FLORIDA	NCAROLINA
Y_NY	1.00000	-0.37698 <.0001	-0.02631 0.4119	-0.06577 0.0400	0.19581 <.0001	0.09812 0.0022	0.22912 <.0001	0.01653 0.6061	0.24610 <.0001	0.21079 <.0001	-0.14696 <.0001	0.21460 <.0001	-0.15638 <.0001	0.23039 <.0001	-0.06696 0.0366	0.10427 0.0011	0.40015 <.0001	0.43875 <.0001	-0.21806 <.0001
SPD	-0.37698 <.0001	1.00000	0.36158 <.0001	0.05869 0.0670	-0.15163 <.0001	-0.09315 0.0036	-0.40952 <.0001	-0.14360 <.0001	-0.25649 <.0001	-0.02244 0.4840	0.31240 <.0001	-0.31806 <.0001	0.33970 <.0001	-0.24174 <.0001	-0.11683 0.0003	-0.25464 <.0001	-0.28481 <.0001	-0.20801 <.0001	-0.24644 <.0001
ADJDIST	-0.02631 0.4119	0.36158 <.0001	1.00000	-0.09417 0.0032	-0.02610 0.4156	0.02504 0.4347	-0.16035 <.0001	0.09996 0.0018	-0.09842 0.0021	0.09946 0.0019	0.01135 0.7233	0.70652 <.0001	-0.33511 <.0001	0.06402 0.0456	-0.02299 0.4733	-0.18821 <.0001	0.09719 0.0024	0.07179 0.0250	-0.05066 0.1139
NEAR	-0.06577 0.0400	0.05869 0.0670	-0.09417 0.0032	1.00000	-0.03318 0.3006	0.07319 0.0223	0.06364 0.0470	-0.02888 0.3677	-0.09352 0.0035	-0.54721 <.0001	0.30739 <.0001	-0.13092 <.0001	0.03724 0.2454	-0.01371 0.6690	-0.06026 0.0600	-0.02080 0.5165	-0.08528 0.0077	-0.10318 0.0013	0.11205 0.0005
ADJ	0.19581 <.0001	-0.15163 <.0001	-0.02610 0.4156	-0.03318 0.3006	1.00000	0.06436 0.0445	0.10513 0.0010	-0.03734 0.2441	0.09436 0.0032	0.12620 <.0001	-0.00967 0.7629	0.08887 0.0055	-0.04453 0.1647	0.05329 0.0963	-0.00673 0.8338	0.08597 0.0072	0.10380 0.0012	0.18445 <.0001	-0.04154 0.1950
PLT	0.09812 0.0022	-0.09315 0.0036	0.02504 0.4347	0.07319 0.0223	0.06436 0.0445	1.00000	0.45414 <.0001	0.09491 0.0030	0.03982 0.2141	0.04096 0.2013	0.03208 0.3170	0.09995 0.0018	-0.10581 0.0009	0.00049 0.9878	0.00803 0.8022	0.02182 0.4962	-0.04166 0.1937	0.07451 0.0200	0.02004 0.5320
LSPLT	0.22912 <.0001	-0.40952 <.0001	-0.16035 <.0001	0.06364 0.0470	0.10513 0.0010	0.45414 <.0001	1.00000	0.13305 <.0001	0.19131 <.0001	0.02624 0.4132	-0.14722 <.0001	0.17986 <.0001	-0.17452 <.0001	0.06698 0.0365	0.03337 0.2979	0.07596 0.0177	0.05657 0.0775	0.12566 <.0001	0.07814 0.0147
HGV	0.01653 0.6061	-0.14360 <.0001	0.09996 0.0018	-0.02888 0.3677	-0.03734 0.2441	0.09491 0.0030	0.13305 <.0001	1.00000	0.12388 0.0001	-0.01339 0.6762	-0.10871 0.0007	0.21782 <.0001	-0.08961 0.0051	0.08209 0.0103	0.01157 0.7182	0.01172 0.7148	0.13167 <.0001	0.01264 0.6934	0.11237 0.0004
MUP	0.24610 <.0001	-0.25649 <.0001	-0.09842 0.0021	-0.09352 0.0035	0.09436 0.0032	0.03982 0.2141	0.19131 <.0001	0.12388 0.0001	1.00000	0.01919 0.5495	-0.48996 <.0001	0.10353 0.0012	-0.09046 0.0047	0.03030 0.3447	0.03319 0.3005	0.03250 0.3106	0.22231 <.0001	0.18108 <.0001	-0.03623 0.2584
MED	0.21079 <.0001	-0.02244 0.4840	0.09946 0.0019	-0.54721 <.0001	0.12620 <.0001	0.04096 0.2013	0.02624 0.4132	-0.01339 0.6762	0.01919 0.5495	1.00000	0.05617 0.0796	0.11924 0.0002	0.03762 0.2405	0.22320 <.0001	-0.10065 0.0017	-0.02150 0.5026	0.14490 <.0001	0.41416 <.0001	-0.27307 <.0001
CTRL	-0.14696 <.0001	0.31240 <.0001	0.01135 0.7233	0.30739 <.0001	-0.00967 0.7629	0.03208 0.3170	-0.14722 <.0001	-0.10871 0.0007	-0.48996 <.0001	0.05617 0.0796	1.00000	-0.21989 <.0001	0.12273 0.0001	0.11919 0.0002	-0.28967 <.0001	-0.02401 0.4539	-0.29794 <.0001	0.00700 0.8271	-0.09311 0.0036
TTC	0.21460 <.0001	-0.31806 <.0001	0.70652 <.0001	-0.13092 <.0001	0.08887 0.0055	0.09995 0.0018	0.17986 <.0001	0.21782 <.0001	0.10353 0.0012	0.11924 0.0002	-0.21989 <.0001	1.00000	-0.55213 <.0001	0.19072 <.0001	0.05629 0.0790	-0.03702 0.2481	0.28831 <.0001	0.22051 <.0001	0.09311 0.0036
DECEL	-0.15638 <.0001	0.33970 <.0001	-0.33511 <.0001	0.03724 0.2454	-0.04453 0.1647	-0.10581 0.0009	-0.17452 <.0001	-0.08961 0.0051	-0.09046 0.0047	0.03762 0.2405	0.12273 0.0001	-0.55213 <.0001	1.00000	-0.03817 0.2338	-0.04083 0.2027	0.03580 0.2640	-0.09215 0.0040	0.12411 0.0001	-0.13859 <.0001
AGE	0.23039 <.0001	-0.24174 <.0001	0.06402 0.0456	-0.01371 0.6690	0.05329 0.0963	0.00049 0.9878	0.06698 0.0365	0.08209 0.0103	0.03030 0.3447	0.22320 <.0001	0.11919 0.0002	0.19072 <.0001	-0.03817 0.2338	1.00000	-0.23479 <.0001	0.22034 <.0001	0.28402 <.0001	0.34292 <.0001	0.15748 <.0001
BUSINESS	-0.06696 0.0366	-0.11683 0.0003	-0.02299 0.4733	-0.06026 0.0600	-0.00673 0.8338	0.00803 0.8022	0.03337 0.2979	0.01157 0.7182	0.03319 0.3005	-0.10065 0.0017	-0.28967 <.0001	0.05629 0.0790	-0.04083 0.2027	-0.23479 <.0001	1.00000	-0.03417 0.2865	-0.12197 0.0001	-0.15500 <.0001	0.24210 <.0001
FEMALE	0.10427 0.0011	-0.25464 <.0001	-0.18821 <.0001	-0.02080 0.5165	0.08597 0.0072	0.02182 0.4962	0.07596 0.0177	0.01172 0.7148	0.03250 0.3106	-0.02150 0.5026	-0.02401 0.4539	-0.03702 0.2481	0.03580 0.2640	0.22034 <.0001	-0.03417 0.2865	1.00000	-0.09455 0.0031	-0.01725 0.5906	0.40772 <.0001
CAMPUS	0.40015 <.0001	-0.28481 <.0001	0.09719 0.0024	-0.08528 0.0077	0.10380 0.0012	-0.04166 0.1937	0.05657 0.0775	0.13167 <.0001	0.22231 <.0001	0.14490 <.0001	-0.29794 <.0001	0.28831 <.0001	-0.09215 0.0040	0.28402 <.0001	-0.12197 0.0001	-0.09455 0.0031	1.00000	0.39022 <.0001	-0.24329 <.0001
FLORIDA	0.43875 <.0001	-0.20801 <.0001	0.07179 0.0250	-0.10318 0.0013	0.18445 <.0001	0.07451 0.0200	0.12566 <.0001	0.01264 0.6934	0.18108 <.0001	0.41416 <.0001	0.00700 0.8271	0.22051 <.0001	0.12411 0.0001	0.34292 <.0001	-0.15500 <.0001	-0.01725 0.5906	0.39022 <.0001	1.00000	-0.46814 <.0001
NCAROLINA	-0.21806 <.0001	-0.24644 <.0001	-0.05066 0.1139	0.11205 0.0005	-0.04154 0.1950	0.02004 0.5320	0.07814 0.0147	0.11237 0.0004	-0.03623 0.2584	-0.27307 <.0001	-0.09311 0.0036	0.09311 0.0036	-0.13859 <.0001	0.15748 <.0001	0.24210 <.0001	0.40772 <.0001	-0.24329 <.0001	-0.46814 <.0001	1.00000

- The following variables show a significant positive correlation with the dependent variable, suggesting an increase in yielding with an increase in the variable (or binary variable change from 0 to 1): ADJ, PLT, LSPLT, MUP, MED, TTC, AGE, FEMALE, CAMPUS, and FLORIDA
- The following variables show a significant negative or inverse correlation with the dependent variable, suggesting a decrease in yielding with an increase in the variable (or binary variable change from 0 to 1): SPD, NEAR, CTRL, DECEL, BUSINESS, and NCAROLINA
- The following variables are intercorrelated: ADJDIST to TTC (0.70652), NEAR to MED (-0.54721), PLT to LSPLT (0.45414), MUP to CTRL (-0.48996), TTC to DECEL (-0.55213), FLORIDA to NCAROLINA (-0.46814)
- It is expected that ADJDIST, TTC, and DECEL are intercorrelated since TTC and DECEL are calculated using the ADJDIST

Binary Logit (Y vs. NY, full model) [Y-17]

Table 46: Yield Model Y-17 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.7403	0.7325	5.6446	0.0175
SPD	1	-0.1536	0.0336	20.8857	<.0001
ADJDIST	1	0.00956	0.00428	4.9766	0.0257
NEAR	1	0.3508	0.2327	2.2732	0.1316
ADJ	1	0.9433	0.3369	7.8398	0.0051
PLT	1	0.1629	0.2032	0.6429	0.4227
LSPLT	1	0.9654	0.3811	6.4158	0.0113
HGV	1	-0.7545	0.4071	3.4341	0.0639
MUP	1	0.8336	0.2614	10.1678	0.0014
MED	1	0.5690	0.3067	3.4417	0.0636
CTRL	1	-0.1772	0.2457	0.5203	0.4707
TTC	1	-0.2721	0.1168	5.4243	0.0199
DECEL	1	-0.1357	0.0464	8.5723	0.0034
AGE	1	0.3036	0.2648	1.3152	0.2515
BUSINESS	1	0.2330	0.4025	0.3351	0.5627
FEMALE	1	0.8296	0.2031	16.6753	<.0001
CAMPUS	1	1.0273	0.2023	25.7860	<.0001
FLORIDA	1	1.1607	0.2829	16.8351	<.0001
NCAROLINA	1	-1.3077	0.2624	24.8354	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.858	0.803	0.916
ADJDIST	1.010	1.001	1.018
NEAR	1.420	0.900	2.241
ADJ	2.568	1.327	4.971
PLT	1.177	0.790	1.753
LSPLT	2.626	1.244	5.542
HGV	0.470	0.212	1.044
MUP	2.302	1.379	3.842
MED	1.766	0.968	3.222
CTRL	0.838	0.518	1.356
TTC	0.762	0.606	0.958
DECEL	0.873	0.797	0.956
AGE	1.355	0.806	2.276
BUSINESS	1.262	0.574	2.778
FEMALE	2.292	1.539	3.413
CAMPUS	2.794	1.879	4.153
FLORIDA	3.192	1.834	5.557
NCAROLINA	0.270	0.162	0.452

- AIC (927.883), SC (1020.649), -2 Log L (889.883), R-Square (0.3691), Max-rescaled R-Square (0.4943)
- The following variables were determined to be significant: SPD, ADJDIST, ADJ, LSPLT, MUP, TTC, DECEL, FEMALE, CAMPUS, FLORIDA, and NCAROLINA

Binary Logit (Y vs. NY, forward selection) [Y-18]

Table 47: Yield Model Y-18 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.8635	0.4260	4.1087	0.0427
SPD	1	-0.0913	0.0173	27.6875	<.0001
ADJ	1	0.9113	0.3302	7.6156	0.0058
LSPLT	1	1.0428	0.3393	9.4476	0.0021
HGV	1	-0.8461	0.3987	4.5032	0.0338
MUP	1	0.7766	0.2296	11.4440	0.0007
DECEL	1	-0.1392	0.0394	12.4733	0.0004
FEMALE	1	0.8138	0.1929	17.7995	<.0001
CAMPUS	1	1.1059	0.1787	38.2893	<.0001
FLORIDA	1	1.4502	0.2238	41.9828	<.0001
NCAROLINA	1	-1.1514	0.2262	25.9159	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.913	0.882	0.944
ADJ	2.488	1.302	4.752
LSPLT	2.837	1.459	5.517
HGV	0.429	0.196	0.937
MUP	2.174	1.386	3.409
DECEL	0.870	0.805	0.940
FEMALE	2.256	1.546	3.293
CAMPUS	3.022	2.129	4.290
FLORIDA	4.264	2.750	6.612
NCAROLINA	0.316	0.203	0.493

- AIC (923.934), SC (977.641), -2 Log L (901.934), R-Square (0.3613), Max-rescaled R-Square (0.4838)
- The following variables were determined to be significant: SPD, ADJ, LSPLT, HGV, MUP, DECEL, FEMALE, CAMPUS, FLORIDA, and NCAROLINA
- Increased speed reduces chance of yielding, as does increased necessary deceleration rate.
- Adjacent yield, low speed platoons, presence of multiple pedestrians, student age pedestrians, and female pedestrians increase the chance of yielding
- The coefficient estimates for Florida and North Carolina show that drivers are more likely to yield in Florida and less likely to yield in North Carolina than drivers in Alabama.

Binary Logit Manual Selection 4 (Y vs. NY, remove heavy vehicle) [Y-19]

Table 48: Yield Model Y-19 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.8344	0.4258	3.8390	0.0501
SPD	1	-0.0894	0.0173	26.6024	<.0001
ADJ	1	0.9448	0.3290	8.2497	0.0041
LSPLT	1	0.9833	0.3375	8.4873	0.0036
MUP	1	0.7348	0.2285	10.3443	0.0013
DECEL	1	-0.1369	0.0395	11.9812	0.0005
FEMALE	1	0.8247	0.1922	18.3998	<.0001
CAMPUS	1	1.0476	0.1761	35.3997	<.0001
FLORIDA	1	1.4245	0.2226	40.9600	<.0001
NCAROLINA	1	-1.2034	0.2247	28.6882	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.914	0.884	0.946
ADJ	2.572	1.350	4.902
LSPLT	2.673	1.380	5.181
MUP	2.085	1.332	3.263
DECEL	0.872	0.807	0.942
FEMALE	2.281	1.565	3.325
CAMPUS	2.851	2.019	4.026
FLORIDA	4.156	2.687	6.429
NCAROLINA	0.300	0.193	0.466

- AIC (926.518), SC (975.343), -2 Log L (906.518), R-Square (0.3582), Max-rescaled R-Square (0.4797)
- The following variables were determined to be significant: SPD, ADJ, LSPLT, MUP, DECEL, FEMALE, CAMPUS, FLORIDA, and NCAROLINA
- Increased speed reduces chance of yielding, as does increased necessary deceleration rate.
- Adjacent yield, low speed platoons, presence of multiple pedestrians, and female pedestrians increase the chance of yielding.
- The coefficient estimates for Florida and North Carolina show that drivers are more likely to yield in Florida and less likely to yield in North Carolina than drivers in Alabama.

Binary Logit Manual Selection 5 (Y vs. NY, generic model without state variables) [Y-20]

Table 49: Yield Model Y-20 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.1765	0.3589	0.2417	0.6230
SPD	1	-0.0758	0.0150	25.4927	<.0001
ADJ	1	1.1365	0.2924	15.1036	0.0001
LSPLT	1	0.9066	0.3086	8.6318	0.0033
MUP	1	0.7171	0.2076	11.9286	0.0006
DECEL	1	-0.0328	0.0277	1.3977	0.2371
FEMALE	1	0.3765	0.1626	5.3631	0.0206
CAMPUS	1	1.5320	0.1598	91.9651	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	0.927	0.900	0.955
ADJ	3.116	1.757	5.527
LSPLT	2.476	1.352	4.533
MUP	2.049	1.364	3.077
DECEL	0.968	0.917	1.022
FEMALE	1.457	1.060	2.004
CAMPUS	4.627	3.383	6.329

- AIC (1050.753), SC (1089.812), -2 Log L (1034.753), R-Square (0.2680), Max-rescaled R-Square (0.3589)
- The state dummy variables were removed from the model to provide a model that can apply outside of the three states where data was collected.
- Once these dummy variables were removed, it was seen that DECEL is no longer a significant variable. This variable should be kept for consistency.

Table 50: Probability of Yielding Model Comparison

	Y-1	Y-2	Y-11	Y-13	Y-15	Y-17	Y-18	Y-19	Y-20
Intercept	1.882**	1.8782**	0.0433	-0.2879	2.3137***	1.7403**	0.8635**	0.8344*	0.1765
SPD	-0.1846***	-0.1888***	-0.1093***	-0.0913***	-0.0913***	-0.1536***	-0.0913***	-0.0894***	-0.0758***
ADJDIST	0.0107**	0.0107**	---	---	---	0.00956**	---	N/A	N/A
NEAR	0.5844**	0.491**	---	---	---	0.3508*	---	N/A	N/A
ADJ	1.0466**	1.0192**	0.9211**	0.9113**	0.9113**	0.9433**	0.9113**	0.9448**	1.1365***
PLT	0.2567*	---	0.4120**	---	---	0.1629	---	N/A	N/A
LSPLT	N/A	N/A	N/A	1.0428**	1.0428**	0.9654**	1.0428**	0.9833**	0.9066**
HGV	-0.7799**	-0.7287**	-0.8247**	-0.8461**	-0.8461**	-0.7545*	-0.8461**	N/A	N/A
MUP	1.0254***	1.1791***	0.8055**	0.7766**	0.7766**	0.8336**	0.7766**	0.7348**	0.7171**
MED	1.4558***	1.3798***	---	---	---	0.5690*	---	N/A	N/A
CTRL	-0.3252*	---	---	---	---	-0.1772	---	N/A	N/A
IN_CW	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TTC	-0.2036**	-0.1907**	---	---	---	-0.2721**	---	N/A	N/A
DECEL	-0.00168	---	-0.1332**	-0.1392**	-0.1392**	-0.1357**	-0.1392**	-0.1369**	-0.0328
AGE	0.5377**	0.4682**	---	---	---	0.3036	---	N/A	N/A
DISTR	0.2498	---	N/A	N/A	N/A	N/A	N/A	N/A	N/A
BUSINESS	-0.9817**	-0.849**	---	---	---	0.233	---	N/A	N/A
FEMALE	-0.0104	---	0.7987***	0.8138***	0.8138***	0.8296***	0.8138***	0.8247***	0.3765**
CAMPUS	N/A	N/A	1.0681***	1.1059***	1.1059***	1.0273***	1.1059***	1.0476***	1.5320***
FLORIDA	N/A	N/A	2.5596***	2.6016***	N/A	1.1607***	1.4502***	1.4245***	N/A
ALABAMA	N/A	N/A	1.1337***	1.1514***	-1.4502***	N/A	N/A	N/A	N/A
NCAROLINA	N/A	N/A	N/A	N/A	-2.6016***	-1.3077***	-1.1514***	-1.2034***	N/A
AIC	1078.595	1073.611	928.549	923.934	923.934	927.883	923.934	926.518	1050.753
SC	1156.714	1127.318	982.256	977.641	977.641	1020.649	977.641	975.343	1089.812
-2 Log L	1046.595	1051.611	906.549	901.934	901.934	889.883	901.934	906.518	1034.753
R^2	0.2591	0.2553	0.3582	0.3613	0.3613	0.3691	0.3613	0.3582	0.268
Max-rescaled R^2	0.3470	0.3418	0.4797	0.4838	0.4838	0.4943	0.4838	0.4797	0.3589

Predicting Probability of Hard Yield, P(HY|Yield)

<u>Correlation Table (HY vs. SY)</u>

Table 51: Probability of Hard Yield Correlation Table

Pearson Correlation Coefficients, N = 432
Prob > |r| under H0: Rho=0

	HY	SPD	ADJDIST	NEAR	ADJ	PLT	HGV	MUP	MED	CTRL	TTC	DECEL	AGE	DISTR	BUSINESS	FEMALE	CAMPUS	FLORIDA	ALABAMA
HY	1.00000	-0.10352 0.0315	-0.31248 <.0001	-0.13446 0.0051	0.15089 0.0017	0.05782 0.2304	-0.04518 0.3489	0.26306 <.0001	0.07617 0.1139	-0.21577 <.0001	-0.20050 <.0001	0.21495 <.0001	0.02909 0.5465	0.01091 0.8211	0.01912 0.6919	0.02334 0.6285	0.07151 0.1378	0.18194 0.0001	-0.03358 0.4863
SPD	-0.10352 0.0315	1.00000	0.53843 <.0001	0.00024 0.9961	-0.10561 0.0282	-0.11839 0.0138	-0.12276 0.0107	-0.24497 <.0001	-0.02620 0.5871	0.24540 <.0001	-0.24767 <.0001	0.33004 <.0001	-0.18357 0.0001	0.05268 0.2746	-0.07886 0.1016	-0.12354 0.0102	-0.09810 0.0416	-0.07684 0.1108	0.28201 <.0001
ADJDIST	-0.31248 <.0001	0.53843 <.0001	1.00000	-0.06783 0.1593	-0.01781 0.7120	0.02882 0.5502	0.06186 0.1994	-0.11640 0.0155	0.09703 0.0438	0.15634 0.0011	0.60717 <.0001	-0.18933 <.0001	0.03146 0.5143	0.05674 0.2393	-0.08481 0.0783	-0.16935 0.0004	0.05703 0.2369	0.10449 0.0299	0.00355 0.9413
NEAR	-0.13446 0.0051	0.00024 0.9961	-0.06783 0.1593	1.00000	-0.04739 0.3258	0.04749 0.3248	0.02108 0.6622	-0.02826 0.5580	-0.72991 <.0001	0.16337 0.0007	-0.07165 0.1371	0.02859 0.5534	-0.07149 0.1379	-0.09064 0.0598	-0.00342 0.9435	-0.04806 0.3190	-0.10421 0.0303	-0.12963 0.0070	0.02465 0.6094
ADJ	0.15089 0.0017	-0.10561 0.0282	-0.01781 0.7120	-0.04739 0.3258	1.00000	0.08302 0.0848	-0.04212 0.3825	0.08606 0.0740	0.12432 0.0097	0.00901 0.8519	0.07680 0.1109	-0.05392 0.2634	-0.00486 0.9198	0.01499 0.7560	-0.04445 0.3567	0.05312 0.2706	0.06222 0.1968	0.11998 0.0126	-0.07850 0.1032
PLT	0.05782 0.2304	-0.11839 0.0138	0.02882 0.5502	0.04749 0.3248	0.08302 0.0848	1.00000	0.12430 0.0097	0.07420 0.1236	0.01796 0.7097	-0.03082 0.5230	0.16202 0.0007	-0.08429 0.0801	0.04358 0.3662	0.00092 0.9848	-0.10181 0.0344	0.01026 0.8316	-0.02208 0.6472	0.11922 0.0132	-0.11751 0.0145
HGV	-0.04518 0.3489	-0.12276 0.0107	0.06186 0.1994	0.02108 0.6622	-0.04212 0.3825	0.12430 0.0097	1.00000	0.09735 0.0431	-0.01135 0.8141	-0.02352 0.6259	0.17229 0.0003	-0.09285 0.0538	0.08086 0.0932	0.07074 0.1422	-0.04498 0.3510	0.01470 0.7606	0.07194 0.1355	0.05584 0.2468	-0.11310 0.0187
MUP	0.26306 <.0001	-0.24497 <.0001	-0.11640 0.0155	-0.02826 0.5580	0.08606 0.0740	0.07420 0.1236	0.09735 0.0431	1.00000	-0.00660 0.8912	-0.51517 <.0001	0.10008 0.0376	-0.12364 0.0101	0.06877 0.1536	0.07341 0.1277	-0.01043 0.8288	-0.01868 0.6987	0.21565 <.0001	0.11869 0.0136	-0.12349 0.0102
MED	0.07617 0.1139	-0.02620 0.5871	0.09703 0.0438	-0.72991 <.0001	0.12432 0.0097	0.01796 0.7097	-0.01135 0.8141	-0.00660 0.8912	1.00000	0.02689 0.5772	0.13004 0.0068	0.04256 0.3775	0.18025 0.0002	-0.06788 0.1590	-0.09089 0.0591	0.05276 0.2738	0.14844 0.0020	0.39962 <.0001	-0.24344 <.0001
CTRL	-0.21577 <.0001	0.24540 <.0001	0.15634 0.0011	0.16337 0.0007	0.00901 0.8519	-0.03082 0.5230	-0.02352 0.6259	-0.51517 <.0001	0.02689 0.5772	1.00000	-0.07240 0.1330	0.09401 0.0509	0.01138 0.8135	-0.11440 0.0174	-0.20051 <.0001	0.07523 0.1184	-0.30173 <.0001	0.06585 0.1719	-0.10708 0.0260
TTC	-0.20050 <.0001	-0.24767 <.0001	0.60717 <.0001	-0.07165 0.1371	0.07680 0.1109	0.16202 0.0007	0.17229 0.0003	0.10008 0.0376	0.13004 0.0068	-0.07240 0.1330	1.00000	-0.50070 <.0001	0.16192 0.0007	0.04018 0.4048	-0.01566 0.7455	-0.11825 0.0139	0.14979 0.0018	0.21138 <.0001	-0.24457 <.0001
DECEL	0.21495 <.0001	0.33004 <.0001	-0.18933 <.0001	0.02859 0.5534	-0.05392 0.2634	-0.08429 0.0801	-0.09285 0.0538	-0.12364 0.0101	0.04256 0.3775	0.09401 0.0509	-0.50070 <.0001	1.00000	0.03881 0.4210	-0.04195 0.3844	0.02064 0.6688	0.08103 0.0926	0.02816 0.5594	0.26564 <.0001	-0.08968 0.0626
AGE	0.02909 0.5465	-0.18357 0.0001	0.03146 0.5143	-0.07149 0.1379	-0.00486 0.9198	0.04358 0.3662	0.08086 0.0932	0.06877 0.1536	0.18025 0.0002	0.01138 0.8135	0.16192 0.0007	0.03881 0.4210	1.00000	0.03690 0.4443	-0.22506 <.0001	0.16800 0.0005	0.32360 <.0001	0.36556 <.0001	-0.50312 <.0001
DISTR	0.01091 0.8211	0.05268 0.2746	0.05674 0.2393	-0.09064 0.0598	0.01499 0.7560	0.00092 0.9848	0.07074 0.1422	0.07341 0.1277	-0.06788 0.1590	-0.11440 0.0174	0.04018 0.4048	-0.04195 0.3844	0.03690 0.4443	1.00000	-0.02052 0.6706	-0.01964 0.6839	0.07335 0.1280	-0.11601 0.0159	0.03749 0.4371
BUSINESS	0.01912 0.6919	-0.07886 0.1016	-0.08481 0.0783	-0.00342 0.9435	-0.04445 0.3567	-0.10181 0.0344	-0.04498 0.3510	-0.01043 0.8288	-0.09089 0.0591	-0.20051 <.0001	-0.01566 0.7455	0.02064 0.6688	-0.22506 <.0001	-0.02052 0.6706	1.00000	-0.03443 0.4754	-0.17091 0.0004	-0.20332 <.0001	0.03650 0.4492
FEMALE	0.02334 0.6285	-0.12354 0.0102	-0.16935 0.0004	-0.04806 0.3190	0.05312 0.2706	0.01026 0.8316	0.01470 0.7606	-0.01868 0.6987	0.05276 0.2738	0.07523 0.1184	-0.11825 0.0139	0.08103 0.0926	0.16800 0.0005	-0.01964 0.6839	-0.03443 0.4754	1.00000	-0.10228 0.0336	-0.03120 0.5178	-0.23526 <.0001
CAMPUS	0.07151 0.1378	-0.09810 0.0416	0.05703 0.2369	-0.10421 0.0303	0.06222 0.1968	-0.02208 0.6472	0.07194 0.1355	0.21565 <.0001	0.14844 0.0020	-0.30173 <.0001	0.14979 0.0018	0.02816 0.5594	0.32360 <.0001	0.07335 0.1280	-0.17091 0.0004	-0.10228 0.0336	1.00000	0.39692 <.0001	-0.09208 0.0558
FLORIDA	0.18194 0.0001	-0.07684 0.1108	0.10449 0.0299	-0.12963 0.0070	0.11998 0.0126	0.11922 0.0132	0.05584 0.2468	0.11869 0.0136	0.39962 <.0001	0.06585 0.1719	0.21138 <.0001	0.26564 <.0001	0.36556 <.0001	-0.11601 0.0159	-0.20332 <.0001	-0.03120 0.5178	0.39692 <.0001	1.00000	-0.61894 <.0001
ALABAMA	-0.03358 0.4863	0.28201 <.0001	0.00355 0.9413	0.02465 0.6094	-0.07850 0.1032	-0.11751 0.0145	-0.11310 0.0187	-0.12349 0.0102	-0.24344 <.0001	-0.10708 0.0260	-0.24457 <.0001	-0.08968 0.0626	-0.50312 <.0001	0.03749 0.4371	0.03650 0.4492	-0.23526 <.0001	-0.09208 0.0558	-0.61894 <.0001	1.00000

- The following variables show a significant positive correlation with the dependent variable, suggesting an increase in hard yielding with an increase in the variable (or binary variable change from 0 to 1): ADJ, MUP, DECEL, and FLORIDA
- The following variables show a negative or inverse correlation with the dependent variable, suggesting a decrease in hard yielding with an increase in the variable (or binary variable change from 0 to 1): SPD, ADJDIST, NEAR, CTRL, and TTC
- The following variables are intercorrelated: SPD to ADJDIST (0.53843), ADJDIST to TTC (0.60717), NEAR to MED (-0.72991), MUP to CTRL (-0.51517), TTC to DECEL (-0.50070), AGE to ALABAMA (-0.50312), FLORIDA to ALABAMA (-0.61894)
- It is expected that SPD, ADJDIST, TTC, and DECEL are intercorrelated since TTC and DECEL are calculated using the SPD and ADJDIST

Pseudo Nested Logit (HY vs. SY, forward, prior to the addition of the CAMPUS, FLORIDA, and ALABAMA variables) [HY-1]

Table 52: Yield Model HY-1 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.0760	0.4074	0.0348	0.8520
ADJDIST	1	-0.0126	0.00236	28.6614	<.0001
NEAR	1	-0.7733	0.2381	10.5463	0.0012
ADJ	1	0.9601	0.3192	9.0464	0.0026
MUP	1	1.3988	0.2542	30.2834	<.0001
DECEL	1	0.3686	0.0816	20.3974	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
ADJDIST	0.987	0.983	0.992
NEAR	0.462	0.289	0.736
ADJ	2.612	1.397	4.883
MUP	4.050	2.461	6.666
DECEL	1.446	1.232	1.696

- Decision to yield has already been made
- AIC (456.016), SC (480.427), -2 Log L (444.016), R-Square (0.2361), Max-rescaled R-Square (0.3249)
- Reasonable that these factors affect decision between hard or soft yielding in the manner shown
- Longer distances from the crosswalk and being in the lane closest to the pedestrian decreases the chance of hard yielding
    - Drivers who are further away from the pedestrian have time to react with a soft yield
- Adjacent yields, presence of multiple pedestrians, and higher necessary deceleration rates increase the chance that a driver will hard yield
    - If the deceleration rate required to yield is higher, then it is reasonable that the chance of hard yielding is increased

Pseudo Nested Logit (HY vs. SY, forward, North Carolina represented in intercept) [HY-2]

Table 53: Yield Model HY-2 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.1114	0.5333	0.0436	0.8346
SPD	1	0.1025	0.0259	15.7078	<.0001
ADJDIST	1	-0.0220	0.00307	51.1966	<.0001
NEAR	1	-1.4654	0.3924	13.9462	0.0002
ADJ	1	0.9576	0.3350	8.1693	0.0043
MUP	1	1.2373	0.2661	21.6199	<.0001
MED	1	-1.1434	0.4660	6.0192	0.0142
FLORIDA	1	1.4007	0.2987	21.9858	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SPD	1.108	1.053	1.165
ADJDIST	0.978	0.972	0.984
NEAR	0.231	0.107	0.498
ADJ	2.605	1.351	5.024
MUP	3.446	2.046	5.806
MED	0.319	0.128	0.795
FLORIDA	4.058	2.260	7.288

- AIC (448.777), SC (481.324), -2 Log L (432.777), R-Square (0.2557), Max-rescaled R-Square (0.3519)
- Longer distances from the crosswalk and being in the lane closest to the pedestrian decreases the chance of hard yielding, as well as presence of pedestrians crossing from the median (rather than the curb)
    - Drivers who are further away from the pedestrian have time to react with a soft yield
- Driver are more likely to hard yield at Florida sites than North Carolina sites
- Increased speed, adjacent yields, and presence of multiple pedestrians increase the chance that a driver will hard yield
- HY-3 included the LSPLT variable, which turned out to be insignificant

Pseudo Nested Logit (HY vs. SY, forward, Florida represented in intercept) [HY-4]

Table 54: Yield Model HY-4 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.3699	0.4356	0.7212	0.3957
ADJDIST	1	-0.0134	0.00238	31.7843	<.0001
NEAR	1	-0.7114	0.2413	8.6937	0.0032
ADJ	1	0.8860	0.3247	7.4463	0.0064
MUP	1	1.3643	0.2582	27.9260	<.0001
DECEL	1	0.2865	0.0825	12.0688	0.0005
NCAROLINA	1	-0.9443	0.3398	7.7242	0.0054

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
ADJDIST	0.987	0.982	0.991
NEAR	0.491	0.306	0.788
ADJ	2.425	1.284	4.583
MUP	3.913	2.359	6.491
DECEL	1.332	1.133	1.565
NCAROLINA	0.389	0.200	0.757

- AIC (449.722), SC (478.201), -2 Log L (435.722), R-Square (0.2507), Max-rescaled R-Square (0.3449)
- Longer distances from the crosswalk and being in the lane closest to the pedestrian decreases the chance of hard yielding
    - Drivers who are further away from the pedestrian have time to react with a soft yield
- Driver are more likely to hard yield at Florida sites than North Carolina sites
- Adjacent yields, presence of multiple pedestrians, and higher deceleration rates increase the chance that a driver will hard yield

Correlation Table (HY vs. SY, Alabama represented in intercept)

Table 55: Probability of Hard Yield Correlation with Alabama Intercept

Pearson Correlation Coefficients, N = 432
Prob > |r| under H0: Rho=0

	HY	SPD	ADJDIST	NEAR	ADJ	PLT	LSPLT	HGV	MUP	MED	CTRL	TTC	DECEL	BUSINESS	FEMALE	CAMPUS	FLORIDA	NCAROLINA
HY	1.00000	-0.10352 0.0315	-0.31248 <.0001	-0.13446 0.0051	0.15089 0.0017	0.05782 0.2304	0.04621 0.3379	-0.04518 0.3489	0.26306 <.0001	0.07617 0.1139	-0.21577 <.0001	-0.20050 <.0001	0.21495 <.0001	0.01912 0.6919	0.02334 0.6285	0.07151 0.1378	0.18194 0.0001	-0.18542 0.0001
SPD	-0.10352 0.0315	1.00000	0.53843 <.0001	0.00024 0.9961	-0.10561 0.0282	-0.11839 0.0138	-0.47912 <.0001	-0.12276 0.0107	-0.24497 <.0001	-0.02620 0.5871	0.24540 <.0001	-0.24767 <.0001	0.33004 <.0001	-0.07886 0.1016	-0.12354 0.0102	-0.09810 0.0416	-0.07684 0.1108	-0.20385 <.0001
ADJDIST	-0.31248 <.0001	0.53843 <.0001	1.00000	-0.06783 0.1593	-0.01781 0.7120	0.02882 0.5502	-0.23246 <.0001	0.06186 0.1994	-0.11640 0.0155	0.09703 0.0438	0.15634 0.0011	0.60717 <.0001	-0.18933 <.0001	-0.08481 0.0783	-0.16935 0.0004	0.05703 0.2369	0.10449 0.0299	-0.13055 0.0066
NEAR	-0.13446 0.0051	0.00024 0.9961	-0.06783 0.1593	1.00000	-0.04739 0.3258	0.04749 0.3248	0.10444 0.0300	0.02108 0.6622	-0.02826 0.5580	-0.72991 <.0001	0.16337 0.0007	-0.07165 0.1371	0.02859 0.5534	-0.00342 0.9435	-0.04806 0.3190	-0.10421 0.0303	-0.12963 0.0070	0.13135 0.0063
ADJ	0.15089 0.0017	-0.10561 0.0282	-0.01781 0.7120	-0.04739 0.3258	1.00000	0.08302 0.0848	0.06351 0.1877	-0.04212 0.3825	0.08606 0.0740	0.12432 0.0097	0.00901 0.8519	0.07680 0.1109	-0.05392 0.2634	-0.04445 0.3567	0.05312 0.2706	0.06222 0.1968	0.11998 0.0126	-0.06290 0.1919
PLT	0.05782 0.2304	-0.11839 0.0138	0.02882 0.5502	0.04749 0.3248	0.08302 0.0848	1.00000	0.56626 <.0001	0.12430 0.0097	0.07420 0.1236	0.01796 0.7097	-0.03082 0.5230	0.16202 0.0007	-0.08429 0.0801	-0.10181 0.0344	0.01026 0.8316	-0.02208 0.6472	0.11922 0.0132	-0.02087 0.6653
LSPLT	0.04621 0.3379	-0.47912 <.0001	-0.23246 <.0001	0.10444 0.0300	0.06351 0.1877	0.56626 <.0001	1.00000	0.19073 <.0001	0.16303 0.0007	-0.01603 0.7398	-0.12956 0.0070	0.20435 <.0001	-0.21787 <.0001	-0.02017 0.6759	0.05128 0.2876	-0.03143 0.5147	0.05998 0.2134	0.13165 0.0061
HGV	-0.04518 0.3489	-0.12276 0.0107	0.06186 0.1994	0.02108 0.6622	-0.04212 0.3825	0.12430 0.0097	0.19073 <.0001	1.00000	0.09735 0.0431	-0.01135 0.8141	-0.02352 0.6259	0.17229 0.0003	-0.09285 0.0538	-0.04498 0.3510	0.01470 0.7606	0.07194 0.1355	0.05584 0.2468	0.05139 0.2865
MUP	0.26306 <.0001	-0.24497 <.0001	-0.11640 0.0155	-0.02826 0.5580	0.08606 0.0740	0.07420 0.1236	0.16303 0.0007	0.09735 0.0431	1.00000	-0.00660 0.8912	-0.51517 <.0001	0.10008 0.0376	-0.12364 0.0101	-0.01043 0.8288	-0.01868 0.6987	0.21565 <.0001	0.11869 0.0136	-0.01394 0.7727
MED	0.07617 0.1139	-0.02620 0.5871	0.09703 0.0438	-0.72991 <.0001	0.12432 0.0097	0.01796 0.7097	-0.01603 0.7398	-0.01135 0.8141	-0.00660 0.8912	1.00000	0.02689 0.5772	0.13004 0.0068	0.04256 0.3775	-0.09089 0.0591	0.05276 0.2738	0.14844 0.0020	0.39962 <.0001	-0.22849 <.0001
CTRL	-0.21577 <.0001	0.24540 <.0001	0.15634 0.0011	0.16337 0.0007	0.00901 0.8519	-0.03082 0.5230	-0.12956 0.0070	-0.02352 0.6259	-0.51517 <.0001	0.02689 0.5772	1.00000	-0.07240 0.1330	0.09401 0.0509	-0.20051 <.0001	0.07523 0.1184	-0.30173 <.0001	0.06585 0.1719	0.03290 0.4952
TTC	-0.20050 <.0001	-0.24767 <.0001	0.60717 <.0001	-0.07165 0.1371	0.07680 0.1109	0.16202 0.0007	0.20435 <.0001	0.17229 0.0003	0.10008 0.0376	0.13004 0.0068	-0.07240 0.1330	1.00000	-0.50070 <.0001	-0.01566 0.7455	-0.11825 0.0139	0.14979 0.0018	0.21138 <.0001	0.00113 0.9813
DECEL	0.21495 <.0001	0.33004 <.0001	-0.18933 <.0001	0.02859 0.5534	-0.05392 0.2634	-0.08429 0.0801	-0.21787 <.0001	-0.09285 0.0538	-0.12364 0.0101	0.04256 0.3775	0.09401 0.0509	-0.50070 <.0001	1.00000	0.02064 0.6688	0.08103 0.0926	0.02816 0.5594	0.26564 <.0001	-0.22789 <.0001
BUSINESS	0.01912 0.6919	-0.07886 0.1016	-0.08481 0.0783	-0.00342 0.9435	-0.04445 0.3567	-0.10181 0.0344	-0.02017 0.6759	-0.04498 0.3510	-0.01043 0.8288	-0.09089 0.0591	-0.20051 <.0001	-0.01566 0.7455	0.02064 0.6688	1.00000	-0.03443 0.4754	-0.17091 0.0004	-0.20332 <.0001	0.20829 <.0001
FEMALE	0.02334 0.6285	-0.12354 0.0102	-0.16935 0.0004	-0.04806 0.3190	0.05312 0.2706	0.01026 0.8316	0.05128 0.2876	0.01470 0.7606	-0.01868 0.6987	0.05276 0.2738	0.07523 0.1184	-0.11825 0.0139	0.08103 0.0926	-0.03443 0.4754	1.00000	-0.10228 0.0336	-0.03120 0.5178	0.28571 <.0001
CAMPUS	0.07151 0.1378	-0.09810 0.0416	0.05703 0.2369	-0.10421 0.0303	0.06222 0.1968	-0.02208 0.6472	-0.03143 0.5147	0.07194 0.1355	0.21565 <.0001	0.14844 0.0020	-0.30173 <.0001	0.14979 0.0018	0.02816 0.5594	-0.17091 0.0004	-0.10228 0.0336	1.00000	0.39692 <.0001	-0.38467 <.0001
FLORIDA	0.18194 0.0001	-0.07684 0.1108	0.10449 0.0299	-0.12963 0.0070	0.11998 0.0126	0.11922 0.0132	0.05998 0.2134	0.05584 0.2468	0.11869 0.0136	0.39962 <.0001	0.06585 0.1719	0.21138 <.0001	0.26564 <.0001	-0.20332 <.0001	-0.03120 0.5178	0.39692 <.0001	1.00000	-0.56150 <.0001
NCAROLINA	-0.18542 0.0001	-0.20385 <.0001	-0.13055 0.0066	0.13135 0.0063	-0.06290 0.1919	-0.02087 0.6653	0.13165 0.0061	0.05139 0.2865	-0.01394 0.7727	-0.22849 <.0001	0.03290 0.4952	0.00113 0.9813	-0.22789 <.0001	0.20829 <.0001	0.28571 <.0001	-0.38467 <.0001	-0.56150 <.0001	1.00000

- The following variables show a significant positive correlation with the dependent variable, suggesting an increase in hard yielding with an increase in the variable (or binary variable change from 0 to 1): ADJ, MUP, DECEL, and FLORIDA
- The following variables show a negative or inverse correlation with the dependent variable, suggesting a decrease in hard yielding with an increase in the variable (or binary variable change from 0 to 1): SPD, ADJDIST, NEAR, CTRL, TTC, and NCAROLINA
- The following variables are intercorrelated: SPD to ADJDIST (0.53843), SPD to LSPLT (-0.47912), ADJDIST to TTC (0.60717), NEAR to MED (-0.72991), PLT to LSPLT (0.56626), MUP to CTRL (-0.51517), TTC to DECEL (-0.50070), FLORIDA to NCAROLINA (-0.56150)
- It is expected that SPD, ADJDIST, TTC, and DECEL are intercorrelated since TTC and DECEL are calculated using the SPD and ADJDIST

Pseudo Nested Logit (HY vs. SY, forward) [HY-5]

Table 56: Yield Model HY-5 Results

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.3699	0.4356	0.7212	0.3957
ADJDIST	1	-0.0134	0.00238	31.7843	<.0001
NEAR	1	-0.7114	0.2413	8.6937	0.0032
ADJ	1	0.8860	0.3247	7.4463	0.0064
MUP	1	1.3643	0.2582	27.9260	<.0001
DECEL	1	0.2865	0.0825	12.0688	0.0005
NCAROLINA	1	-0.9443	0.3398	7.7242	0.0054

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
ADJDIST	0.987	0.982	0.991
NEAR	0.491	0.306	0.788
ADJ	2.425	1.284	4.583
MUP	3.913	2.359	6.491
DECEL	1.332	1.133	1.565
NCAROLINA	0.389	0.200	0.757

- AIC (449.722), SC (478.201), -2 Log L (435.722), R-Square (0.2507), Max-rescaled R-Square (0.3449)
- Longer distances from the crosswalk and being in the lane closest to the pedestrian decreases the chance of hard yielding
    - Drivers who are further away from the pedestrian have time to react with a soft yield
- Driver are more likely to hard yield at Florida sites than North Carolina sites
- Adjacent yields, presence of multiple pedestrians, and higher necessary deceleration rates increase the chance that a driver will hard yield

Table 57: Probability of Hard Yield Model Comparison

	HY-1	HY-3	HY-5
Intercept	-0.076	1.1114	0.3699
SPD	---	0.1025***	---
ADJDIST	-0.0126***	-0.0220***	-0.0134***
NEAR	-0.7733**	-1.4654**	-0.7114**
ADJ	0.9601**	0.9576**	0.8860**
PLT	---	---	---
LSPLT	N/A	---	---
HGV	---	---	---
MUP	1.3988***	1.2373***	1.3643***
MED	---	-1.1434**	---
CTRL	---	---	---
IN_CW	N/A	N/A	N/A
TTC	---	---	---
DECEL	0.3686***	---	0.2865**
AGE	---	---	N/A
DISTR	---	N/A	N/A
BUSINESS	---	---	---
FEMALE	---	---	---
CAMPUS	N/A	---	---
FLORIDA	N/A	1.4007***	---
ALABAMA	N/A	---	N/A
NCAROLINA	N/A	N/A	-0.9443
AIC	456.016	448.777	449.722
SC	480.427	481.324	478.201
-2 Log L	444.016	432.777	435.722
R^2	0.2361	0.2557	0.2507
Max-rescaled R^2	0.3249	0.3519	0.3449

APPENDIX E: GAP MODELING DETAILED RESULTS

Preliminary Data Preparation

Data preparation for gap acceptance modeling involved the following steps.

1) Normalized Gap Length (N_GL) was introduced as a variable in the dataset. N_GL was arrived at by dividing OBS (observed gap length) by CR_WIDTH (crosswalk width).
2) For analysis, data were segregated into two sets: a) Non-Controlled Crossings b) Staged Crossings.
3) Similar to above datasets, data were separated into single and multilane sites. The number of observations obtained for Single lane Non-Controlled crossings were small in size compared with Multi-lane sites. Table 58 shows sample sizes for segregate datasets excluding missing observations.

Table 58: Sample sizes of segregate datasets for gap acceptance analysis

Dataset	Sample Size
Single Lane (Non-Controlled)	153
Single Lane (Staged)	213
All Sites (Non-Controlled)	394
All Sites(Staged)	836

Detailed Gap Acceptance Results

Single Lane (Non-Controlled) - Model I

Table 59 provides summary statistics for single lane non-controlled dataset. The table shows the number of missing values for each variable in the dataset. Variables like SPD, ADJDIST, TTC, DECEL, and DIST_DEL have around quarter of dataset with missing values.

Table 59: Summary Statistics for Single Lane Non-Controlled dataset

Variable	N	N Miss
_	153	0
SPD	113	40
ADJDIST	113	40
NEAR	153	0
TRIG	153	0
STP	153	0
ADJ	153	0
PLT	153	0
HGV	153	0
MUP	153	0
MED	153	0
CTRL	153	0
IN_CW	153	0
TIME	152	1
W_SP	152	1
COUNT	153	0
DIST_DEL	114	39
GO	153	0
LAG	153	0
TTC	113	40
DECEL	113	40
OBS	153	0
AGE	153	0
DISTR	153	0
ATTIRE	153	0
GENDER	153	0
Cr_Width	153	0
N_GL	153	0

Table 60 summarizes contingency tables that indicate the distribution of accepted and rejected gaps for categorical predictor variables in the dataset.

Table 60: Accepted and rejected gaps for categorical predictor variables- Single Lane (Non-Controlled)

Table of NEAR by GO

NEAR	GO: Rejected	GO: Accepted	GO: Total
Far Lane	33	50	83
Near Lane	33	37	70
Total	66	87	153

Table of PLT by GO

PLT	GO: Rejected	GO: Accepted	GO: Total
Not in Platoon	41	60	101
In Platoon	25	27	52
Total	66	87	153

Table of HGV by GO

HGV	GO: Rejected	GO: Accepted	GO: Total
Not Heavy Vehicle	63	72	135
Heavy Vehicle	3	15	18
Total	66	87	153

Table of MUP by GO

MUP	GO: Rejected	GO: Accepted	GO: Total
Single Pedestrian	52	64	116
Multiple Pedestrians	14	23	37
Total	66	87	153

Table of MED by GO

MED	GO: Rejected	GO: Accepted	GO: Total
Crossing from Kerb	66	84	150
Crossing from Median	0	3	3
Total	66	87	153

Table of LAG by GO

LAG	GO: Rejected	GO: Accepted	GO: Total
Crossing in Gap	32	18	50
Crossing in Lag	34	69	103
Total	66	87	153

Table of AGE by GO

AGE	GO: Rejected	GO: Accepted	GO: Total
Elder Pedestrian	29	23	52
Young Pedestrian	37	64	101
Total	66	87	153

Table of ATTIRE by GO

ATTIRE	GO: Rejected	GO: Accepted	GO: Total
Casual	39	67	106
Business	27	20	47
Total	66	87	153

Table of GENDER by GO

GENDER	GO: Rejected	GO: Accepted	GO: Total
Male	49	58	107
Female	17	29	46
Total	66	87	153

Single Lane (Staged) - Model II

Table 61 shows the number of missing values for each variable in the dataset. Variables like SPD, ADJDIST, TTC, DECEL, and DIST_DEL have around half of dataset with missing values.

Table 61: Summary Statistics for Single Lane Staged dataset

Variable	N	N Miss
_	213	0
SPD	111	102
ADJDIST	111	102
NEAR	213	0
TRIG	213	0
STP	213	0
ADJ	213	0
PLT	213	0
HGV	213	0
MUP	213	0
MED	213	0
CTRL	213	0
IN_CW	213	0
TIME	200	13
W_SP	200	13
COUNT	213	0
DIST_DEL	112	101
GO	213	0
LAG	213	0
TTC	111	102
DECEL	111	102
OBS	213	0
AGE	213	0
DISTR	213	0
ATTIRE	213	0
GENDER	213	0
Cr_Width	213	0
N_GL	213	0

Table 62 summarizes contingency tables that indicate the distribution of accepted and rejected gaps for categorical predictor variables in the dataset.

Table 62: Accepted and rejected gaps for categorical predictor variables, Single Lane (Staged)

Table of NEAR by GO

NEAR	GO Rejected	GO Accepted	GO Total
Far Lane	8	3	11
Near Lane	171	31	202
Total	179	34	213

Table of PLT by GO

PLT	GO Rejected	GO Accepted	GO Total
Not in Platoon	81	24	105
In Platoon	98	10	108
Total	179	34	213

Table of HGV by GO

HGV	GO Rejected	GO Accepted	GO Total
Not Heavy Vehicle	176	34	210
Heavy Vehicle	3	0	3
Total	179	34	213

Table of MUP by GO

MUP	GO Rejected	GO Accepted	GO Total
Single Pedestrian	172	33	205
Multiple Pedestrians	7	1	8
Total	179	34	213

Table of MED by GO

MED	GO Rejected	GO Accepted	GO Total
Crossing from Kerb	176	32	208
Crossing from Median	3	2	5
Total	179	34	213

Table of LAG by GO

LAG	GO Rejected	GO Accepted	GO Total
Crossing in Gap	75	28	103
Crossing in Lag	104	6	110
Total	179	34	213

Table of AGE by GO

AGE	GO Rejected	GO Accepted	GO Total
Young Pedestrian	179	34	213
Total	179	34	213

Table of ATTIRE by GO

ATTIRE	GO Rejected	GO Accepted	GO Total
Casual	179	34	213
Total	179	34	213

Table of GENDER by GO

GENDER	GO Rejected	GO Accepted	GO Total
Female	179	34	213
Total	179	34	213

All Sites Non-Controlled dataset – Model III

Table 63 shows the number of missing values for each variable in the dataset. Variables like SPD, ADJDIST, TTC, DECEL, and DIST_DEL have around half of dataset with missing values.

Table 63: Summary Statistics for All Sites Non-Controlled dataset

Variable	N	N Miss
_	400	0
SPD	223	177
ADJDIST	224	176
NEAR	400	0
TRIG	400	0
STP	400	0
ADJ	400	0
PLT	400	0
HGV	400	0
MUP	400	0
MED	400	0
CTRL	400	0
IN_CW	341	59
TIME	275	125
W_SP	275	125
COUNT	358	42
DIST_DEL	199	201
GO	394	6
LAG	394	6
TTC	220	180
DECEL	220	180
OBS	394	6
AGE	400	0
DISTR	400	0
ATTIRE	400	0
GENDER	400	0
Cr_Width	400	0
N_GL	400	0

Table 64 summarizes contingency tables that indicate the distribution of accepted and rejected gaps for categorical predictor variables in the dataset.

Table 64: Accepted and rejected gaps for categorical predictor variables (All Sites Non-Controlled dataset)

Table of NEAR by GO			
	GO		
NEAR	Rejected	Accepted	Total
Far Lane	152	109	261
Near Lane	63	70	133
Total	215	179	394
Frequency Missing = 6			

Table of PLT by GO			
	GO		
PLT	Rejected	Accepted	Total
Not in Platoon	162	139	301
In Platoon	53	40	93
Total	215	179	394
Frequency Missing = 6			

Table of HGV by GO			
	GO		
HGV	Rejected	Accepted	Total
Not Heavy Vehicle	211	163	374
Heavy Vehicle	4	16	20
Total	215	179	394
Frequency Missing = 6			

Table of MUP by GO			
	GO		
MUP	Rejected	Accepted	Total
Single Pedestrian	149	127	276
Multiple Pedestrians	66	52	118
Total	215	179	394
Frequency Missing = 6			

Table of MED by GO			
	GO		
MED	Rejected	Accepted	Total
Crossing from Kerb	209	157	366
Crossing from Median	6	22	28
Total	215	179	394
Frequency Missing = 6			

Table of LAG by GO			
	GO		
LAG	Rejected	Accepted	Total
Crossing in Gap	110	62	172
Crossing in Lag	105	117	222
Total	215	179	394
Frequency Missing = 6			

Table of AGE by GO			
	GO		
AGE	Rejected	Accepted	Total
Elder Pedestrian	142	65	207
Young Pedestrian	73	114	187
Total	215	179	394
Frequency Missing = 6			

Table of ATTIRE by GO			
	GO		
ATTIRE	Rejected	Accepted	Total
Casual	178	151	329
Business	37	28	65
Total	215	179	394
Frequency Missing = 6			

Table of GENDER by GO			
	GO		
GENDER	Rejected	Accepted	Total
Male	132	103	235
Female	83	76	159
Total	215	179	394
Frequency Missing = 6			

All Sites Staged dataset- Model IV

Table 65 shows the number of missing values for each variable in the dataset. Variables like SPD, ADJDIST, TTC, DECEL, and DIST_DEL have around half of dataset with missing values.

Table 65: Summary Statistics for All Sites Staged dataset

Variable	N	N Miss
_	843	1
SPD	359	485
ADJDIST	359	485
NEAR	843	1
TRIG	844	0
STP	844	0
ADJ	844	0
PLT	844	0
HGV	844	0
MUP	844	0
MED	844	0
CTRL	844	0
IN_CW	753	91
TIME	409	435
W_SP	409	435
COUNT	766	78
DIST_DEL	323	521
GO	836	8
LAG	836	8
TTC	359	485
DECEL	359	485
OBS	836	8
AGE	844	0
DISTR	844	0
ATTIRE	844	0
GENDER	844	0
Cr_Width	844	0
N_GL	844	0

Table 66 summarizes contingency tables that indicate the distribution of accepted and rejected gaps for categorical predictor variables in the dataset.

Table 66: Accepted and rejected gaps for categorical predictor variables- All sites staged

Table of NEAR by GO

NEAR	GO: Rejected	Accepted	Total
Far Lane	170	40	210
Near Lane	521	104	625
Total	691	144	835

Frequency Missing = 9

Table of PLT by GO

PLT	GO: Rejected	Accepted	Total
Not in Platoon	327	103	430
In Platoon	365	41	406
Total	692	144	836

Frequency Missing = 8

Table of HGV by GO

HGV	GO: Rejected	Accepted	Total
Not Heavy Vehicle	684	142	826
Heavy Vehicle	8	2	10
Total	692	144	836

Frequency Missing = 8

Table of MUP by GO

MUP	GO: Rejected	Accepted	Total
Single Pedestrian	685	140	825
Multiple Pedestrians	7	4	11
Total	692	144	836

Frequency Missing = 8

Table of MED by GO

MED	GO: Rejected	Accepted	Total
Crossing from Kerb	562	108	670
Crossing from Median	130	36	166
Total	692	144	836

Frequency Missing = 8

Table of LAG by GO

LAG	GO: Rejected	Accepted	Total
Crossing in Gap	383	91	474
Crossing in Lag	309	53	362
Total	692	144	836

Frequency Missing = 8

Table of AGE by GO

AGE	GO: Rejected	Accepted	Total
Elder Pedestrian	179	23	202
Young Pedestrian	513	121	634
Total	692	144	836

Frequency Missing = 8

Table of ATTIRE by GO

ATTIRE	GO: Rejected	Accepted	Total
Casual	692	144	836
Total	692	144	836

Frequency Missing = 8

Table of GENDER by GO

GENDER	GO: Rejected	Accepted	Total
Male	510	102	612
Female	182	42	224
Total	692	144	836

Frequency Missing = 8

Distribution of Normalized Gap Length for All Sites and Non-Controlled Crossings

The frequency distribution of N_GL for Non-Controlled Crossings (All Sites Combined) is shown in Table 67 and Figure 36 below.

Table 67: Frequency distribution of N_GL for Non-Controlled Crossings (All Sites Combined)

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	1.4583333
99%	1.0516544
95%	0.8118640
90%	0.6962899
75% Q3	0.4516541
50% Median	0.1993631
25% Q1	0.0791457
10%	0.0402797
5%	0.0279180
1%	0.0000000
0% Min	0.0000000

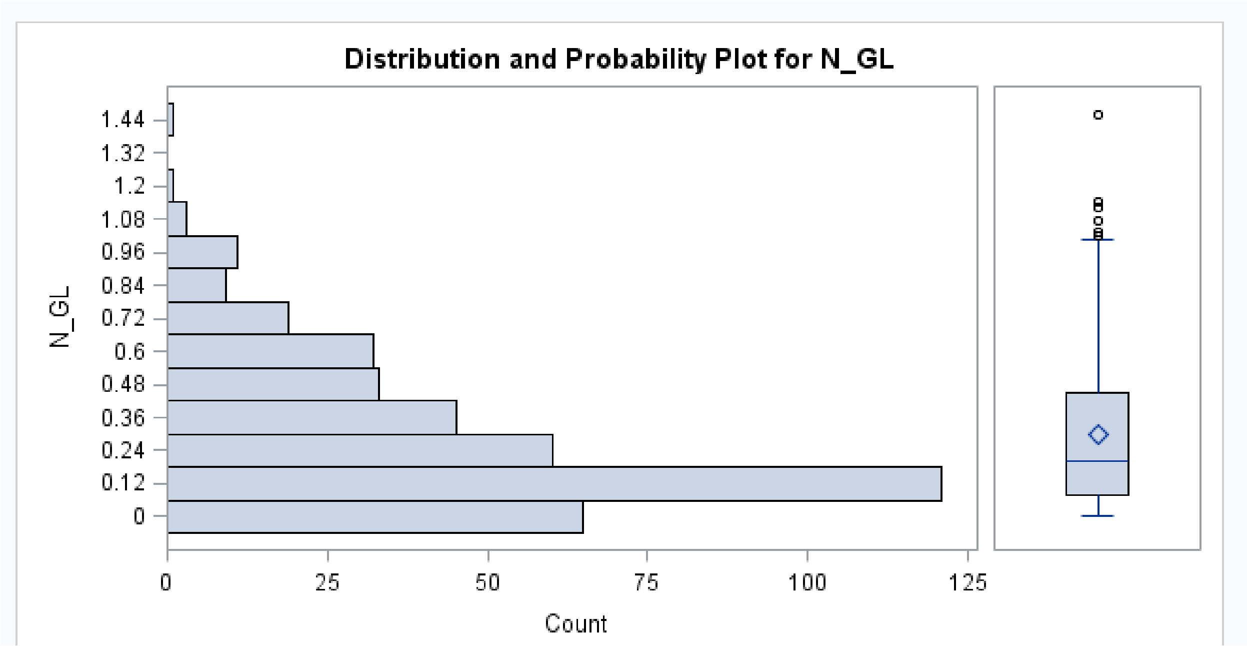


Figure 36: Frequency distribution of N_GL for Non-Controlled Crossings (All Sites Combined)

Figure 36 describes the frequency distribution of N_GL. Much of the data lies between 0.12 and 0.48. The Percentile ranges were used to group the continuous data into discrete bins. Table 68 highlights the distribution of gap acceptance by N_GL. For N_GL values below 50 percentile, rejected gaps are greater in number. Similarly, for N_GL values above 50 percentile mark, gap accepted are greater in number.

Table 68: Distribution of Gap Acceptance by N_GL- Non-Controlled Crossings (All Sites Combined)

Frequency	Table of N_GL by GO		
	GO		
N_GL	Rejected	Accepted	Total
0	0	0	0
10	33	0	33
25	61	0	61
50	82	18	100
75	30	70	100
90	5	55	60
95	2	18	20
100	2	17	19
1.5	0	1	1
Total	215	179	394
Frequency Missing = 6			

Model results for all sites (Non-Controlled) combined w/ normalized gap length: Forward Selection

Table 69 summarizes the Maximum Likelihood Estimates for all sites (Non-Controlled) combined w/ normalized gap length: Forward Selection.

Table 69: Maximum Likelihood Estimates- Non-Controlled Crossings (All Sites Combined)

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.0424	0.2351	75.4681	<.0001
N_GL		1	5.3112	0.4829	120.9762	<.0001
PLT	0	1	0.5543	0.2009	7.6125	0.0058

In earlier logit based gap models, the parameters estimates were not consistent with field observations. Thus, probit models were assessed for the gap dataset. Results above show the parameter estimates for binary probit model (GO/NO-GO).

- The model is constructed for base case (GO = 1)
- The co-efficient for Normalized gap length (N_GL) is 5.3112. This indicates that increase in N_GL increases the predicted probability of gap being accepted.
- Similarly, when the lead vehicle is not in a platoon (indicated by PLT = 0), the predicted probability of gap being accepted increases.

For backward elimination procedure, the model results were similar to forward selection.

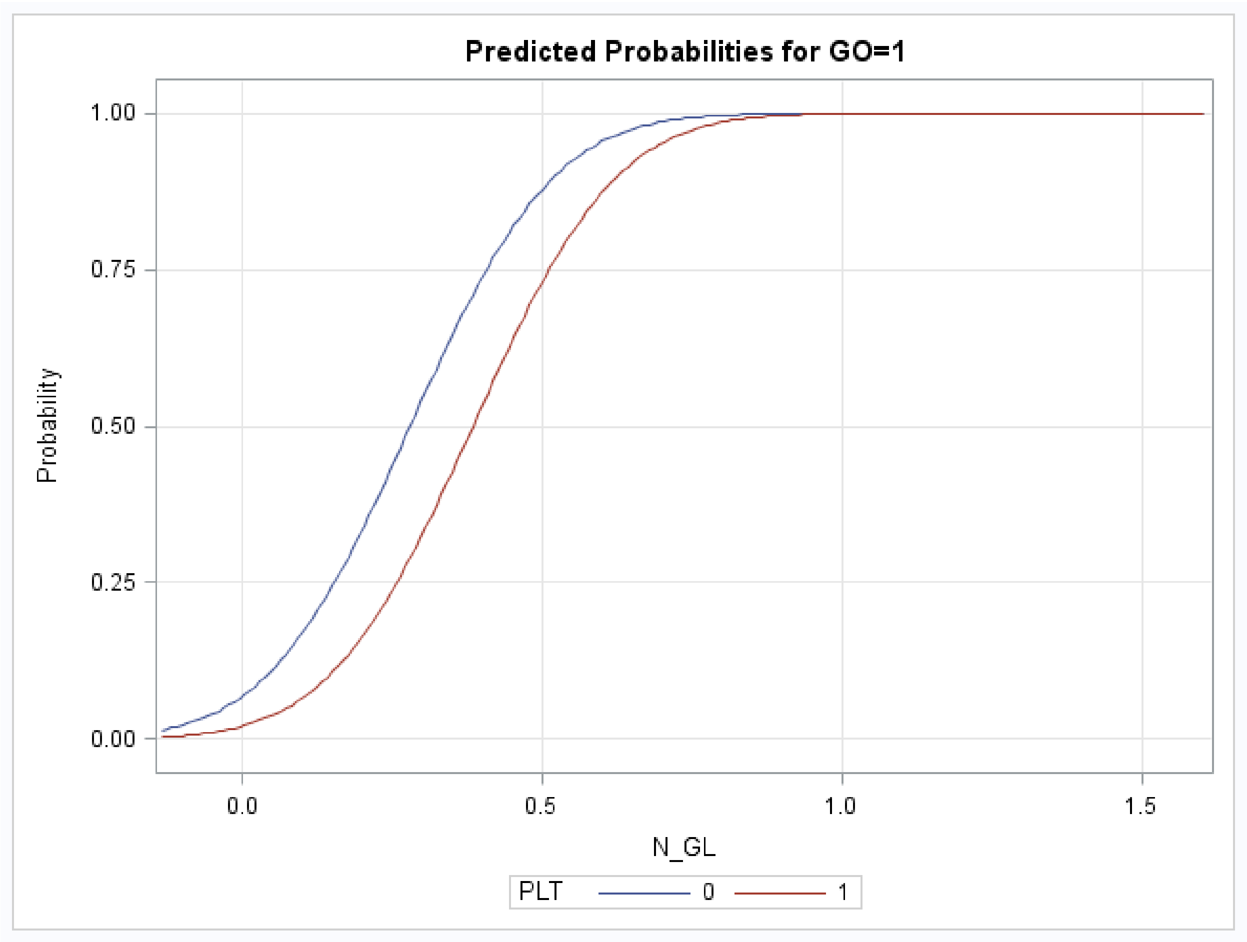


Figure 37: Predicted Gap acceptance as a function of N_GL under platoon and non-platoon conditions

As shown in Figure 37 above, for a given normalized gap length (N_GL), predicted gap acceptance is greater for vehicle not being in platoon.

Distribution of Normalized Gap Length for All Sites and Non-Controlled Crossings

The frequency distribution of N_GL for Non-Controlled Crossings (All Sites Combined) is shown in Table 70 and Figure 38 below. The number of accepted and rejected gaps is shown in Table 71.

Table 70: Frequency distribution of N_GL for Staged Crossings (All Sites Combined)

Quantiles (Definition 5)	
Quantile	**Estimate**
100% Max	1.7800000
99%	1.4054676
95%	0.6211765
90%	0.4288235
75% Q3	0.2369427
50% Median	0.1378312
25% Q1	0.0803439
10%	0.0481250
5%	0.0376000
1%	0.0105556
0% Min	0.0000000

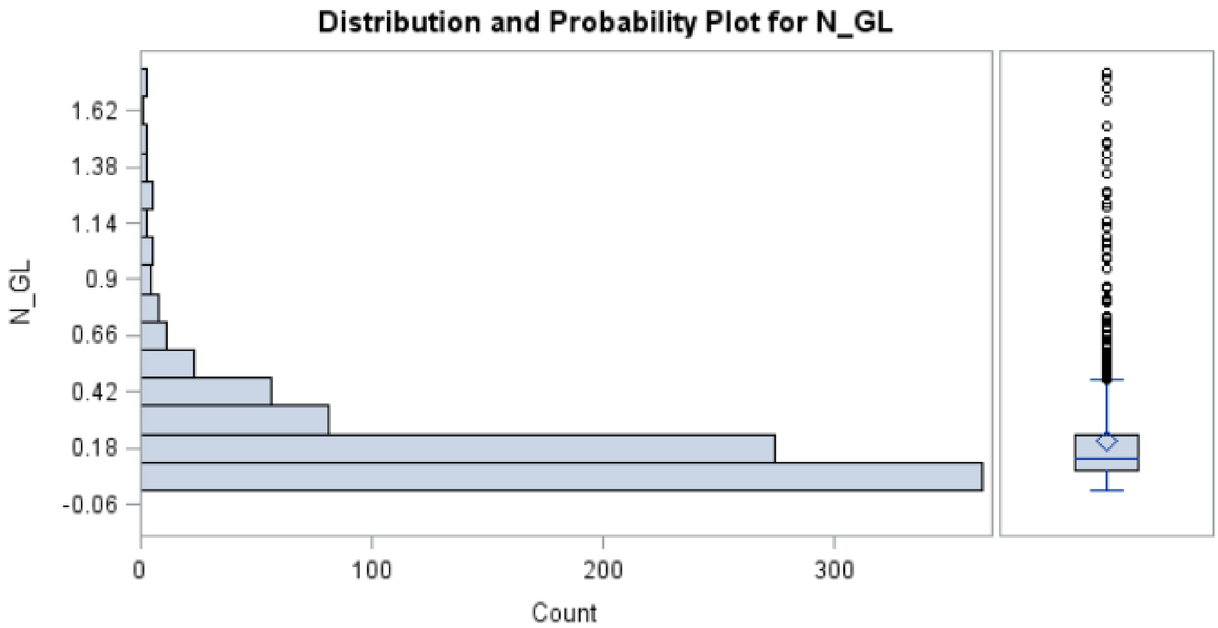


Figure 38: Frequency distribution of N_GL for Staged Crossings (All Sites Combined)

Table 71: Distribution of Gap Acceptance by N_GL- Staged Crossings (All Sites Combined)

Frequency	Table of N_GL by GO			
		GO		
	N_GL	Rejected	Accepted	Total
	0	0	0	0
	10	77	0	77
	25	124	2	126
	50	193	18	211
	75	178	34	212
	90	101	24	125
	95	16	27	43
	99	3	39	42
	Total	692	144	836
	Frequency Missing = 8			

Model results for all sites (Staged) combined w/ normalized gap length: Forward Selection

Table 72 summarizes the Maximum Likelihood Estimates for all sites (staged) combined w/ normalized gap length: Forward Selection.

Table 72: Maximum Likelihood Estimates- Non-Controlled Crossings (All Sites Combined)

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-4.0525	0.3351	146.2384	<.0001
N_GL		1	6.2670	0.6344	97.6016	<.0001
PLT	0	1	0.5371	0.1370	15.3567	<.0001
LAG	0	1	0.2865	0.1349	4.5108	0.0337
GENDER	0	1	1.5585	0.2428	41.1867	<.0001

Results above show the parameter estimates for binary probit model (GO/NO-GO).

- The model is constructed for base case (GO = 1)
- The co-efficient for Normalized gap length (N_GL) is 6.2670. This indicates that increase in N_GL increases the predicted probability of gap being accepted.
- Similarly, when the lead vehicle is not in a platoon (indicated by PLT = 0), the predicted probability of gap being accepted increases.

- Likewise, crossing in a gap and pedestrian crossing the street being male increase the predicted probability of gap being accepted.

For backward elimination procedure, the model results were similar to forward selection. Figure 39 shows predicted gap acceptance under various platoon, lag and gender combinations.

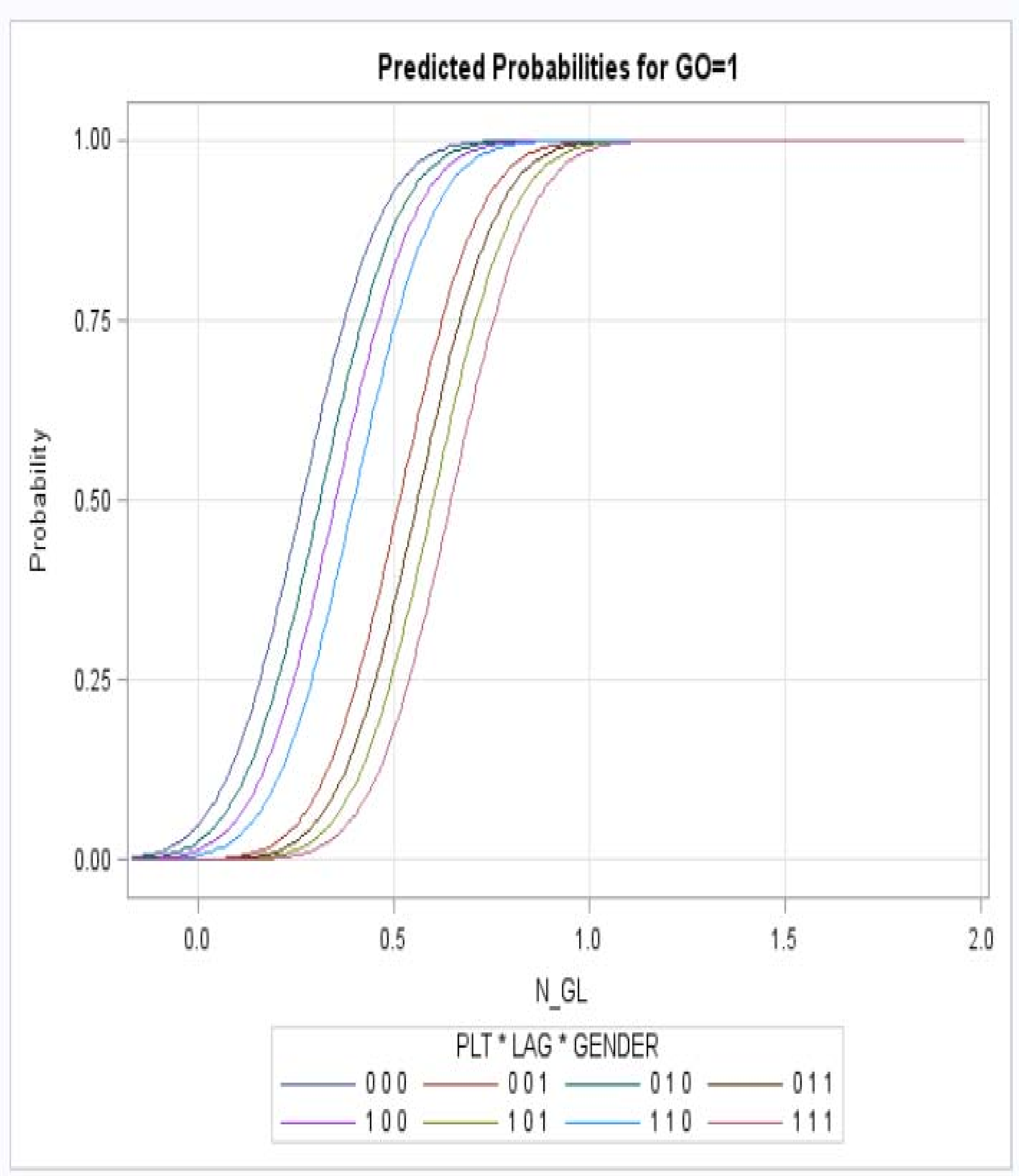


Figure 39: Predicted Gap acceptance under various platoon, lag and gender combinations

Gap Acceptance distribution for Single lane, Non-Controlled crossings by Normalized gap length (N_GL)

The distribution of accepted and rejected gaps by N_GL for non-controlled crossings (single lane) is summarized in Table 73.

Table 73: Distribution of Gap Acceptance by N_GL- Non-Controlled Crossings (Single Lane)

The FREQ Procedure

Frequency	Table of N_GL by GO		
	GO		
N_GL	Rejected	Accepted	Total
10	16	0	16
25	21	1	22
50	23	16	39
75	2	36	38
90	2	21	23
95	1	7	8
100	1	6	7
Total	66	87	153

Model results for Single lane sites (Non-Controlled) combined with normalized gap length for Forward Selection are summarized in Table 74. Figure 40 shows predicted gap acceptance probabilities for non-controlled crossings (single lane sites).

Table 74: Maximum Likelihood Estimates- Non-Controlled Crossings (Single Lane Sites)

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.8904	0.3493	29.2950	<.0001
N_GL		1	5.0483	0.7108	50.4417	<.0001
LAG	0	1	-0.7688	0.2906	7.0007	0.0081

Results above show the parameter estimates for binary probit model (GO/NO-GO).

- The model is constructed for base case (GO = 1)
- The co-efficient for Normalized gap length (N_GL) is 5.0483. This indicates that increase in N_GL increases the predicted probability of gap being accepted.
- Similarly, when the pedestrian crosses street in a gap (indicated by LAG = 0), the predicted probability of gap being accepted decreases.

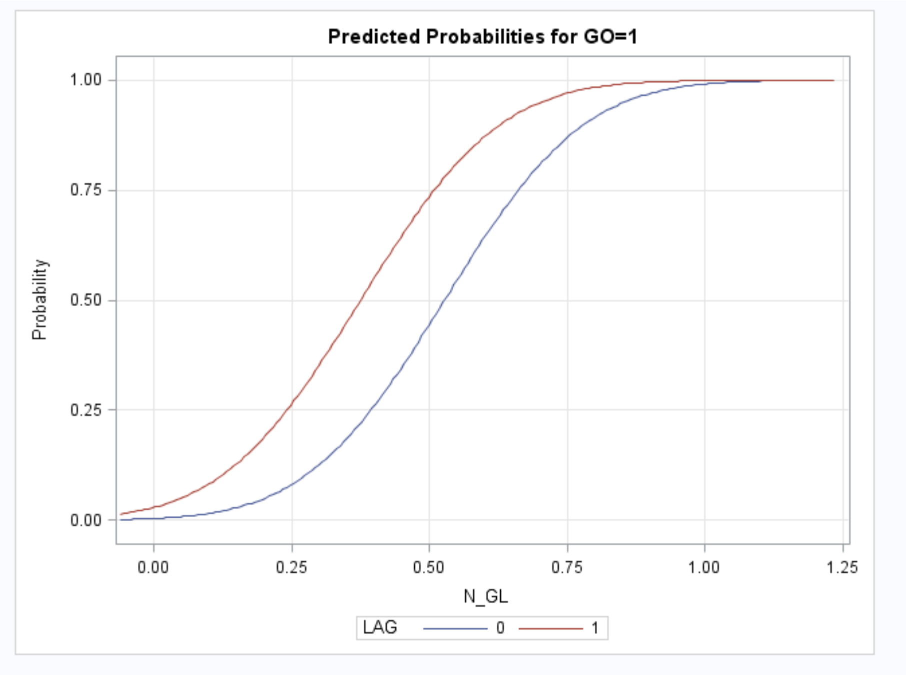


Figure 40: Predicted Gap acceptance probabilities, Non-Controlled Crossings (Single Lane Sites)

Model results for Single lane sites (Staged) crossings w/ normalized gap length: Forward Selection

The distribution of accepted and rejected gaps by N_GL for staged crossings (single lane) is summarized in Table 75.

Table 75: Distribution of Gap Acceptance by N_GL- Staged Crossings (Single Lane)

Frequency	Table of N_GL by GO			
		GO		
	N_GL	Rejected	Accepted	Total
	10	20	0	20
	25	74	0	74
	50	77	5	82
	75	5	16	21
	90	1	3	4
	95	1	5	6
	100	0	2	2
	1.2	0	1	1
	1.3	1	0	1
	1.4	0	1	1
	1.5	0	1	1
	Total	179	34	213

Model results for Single lane sites (Non-Controlled) combined with normalized gap length for Forward Selection are summarized in Table 76. Figure 41 shows predicted gap acceptance probabilities for non-controlled crossings (single lane sites).

Table 76: Maximum Likelihood Estimates- Non-Controlled Crossings (Single Lane Sites)

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-3.3283	0.4027	68.3104	<.0001
N_GL		1	5.1946	0.8320	38.9857	<.0001
NEAR	0	1	1.2189	0.4976	6.0010	0.0143

Results above show the parameter estimates for binary probit model (GO/NO-GO).

- The model is constructed for base case (GO = 1)
- The co-efficient for Normalized gap length (N_GL) is 5.1946. This indicates that increase in N_GL increases the predicted probability of gap being accepted.
- When the lead vehicle is in far lane (indicated by NEAR = 0), the predicted probability of gap being accepted increases.

For backward elimination procedure, the model results were similar to forward selection.

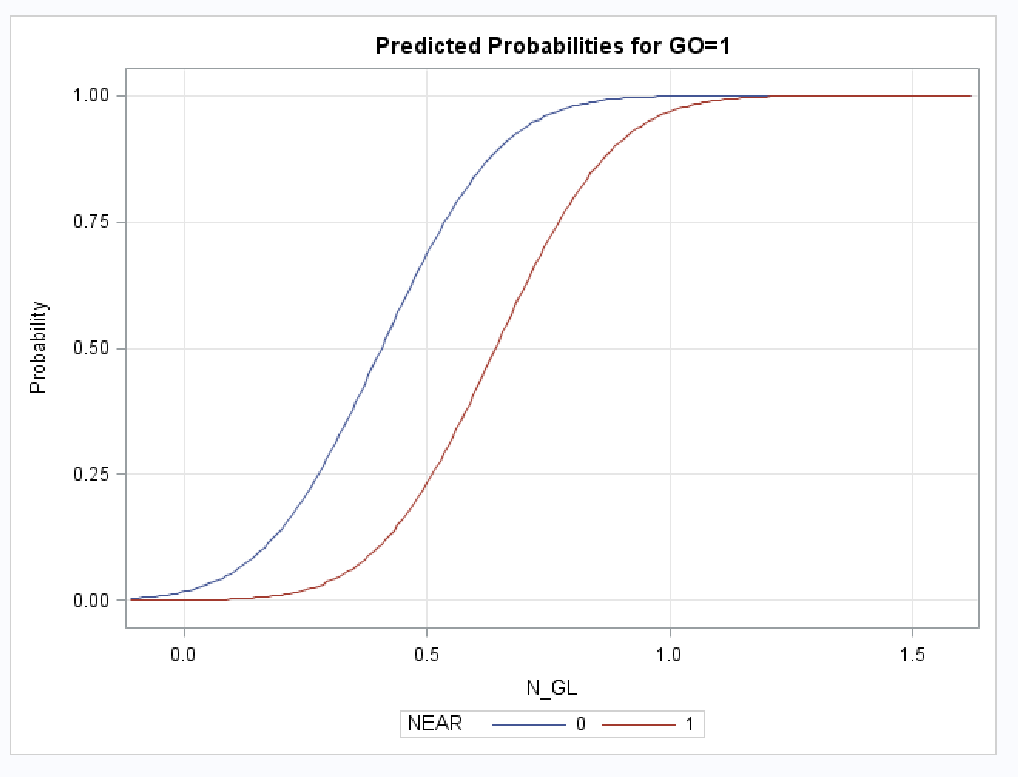


Figure 41: Predicted Gap acceptance probabilities, Staged Crossings (Single Lane Sites)

Validation

Receiver operating characteristic curves are widely used to assess the fit of categorical models in clinical analysis. The curve helps in establishing the threshold values where categorical outcomes predicted with higher degree of accuracy. The concept of sensitivity and specificity are important in this regard.

- Sensitivity: The number of gaps that the test correctly identifies as accepted.
- Specificity: The number of gaps that the test correctly identifies as rejected.

If a higher threshold value for accepting a gap is selected, more accurate gap acceptance will be predicted. However, many gaps that were accepted below the threshold value may not be identified.

Conversely, if a lower threshold value for gap acceptance is fixed, will identify more accepted gaps, but many gaps that are rejected will also be wrongly identified as accepted gap.

Area under curve: The area under a ROC curve quantifies the overall ability of the test to discriminate between the gaps accepted and gaps rejected. A truly useless test (one no better at identifying truly accepted gaps than flipping a coin) has an area of 0.5. A perfect test (one that zero false predicted gap accepted and zero false predicted gap rejected) will have an area of 1.00. Greater the area under of the curve, better is the prediction accuracy of the model.

Figure 42 through Figure 45 show the area under the curve for respective gap models build upon segregate data sets. From ROC curves, it can be observed that Model IV has the largest area under curve (0.9959), followed by Model I (0.8832). All four models fairly predict the gap acceptance outcomes, compared to flipping of coin.

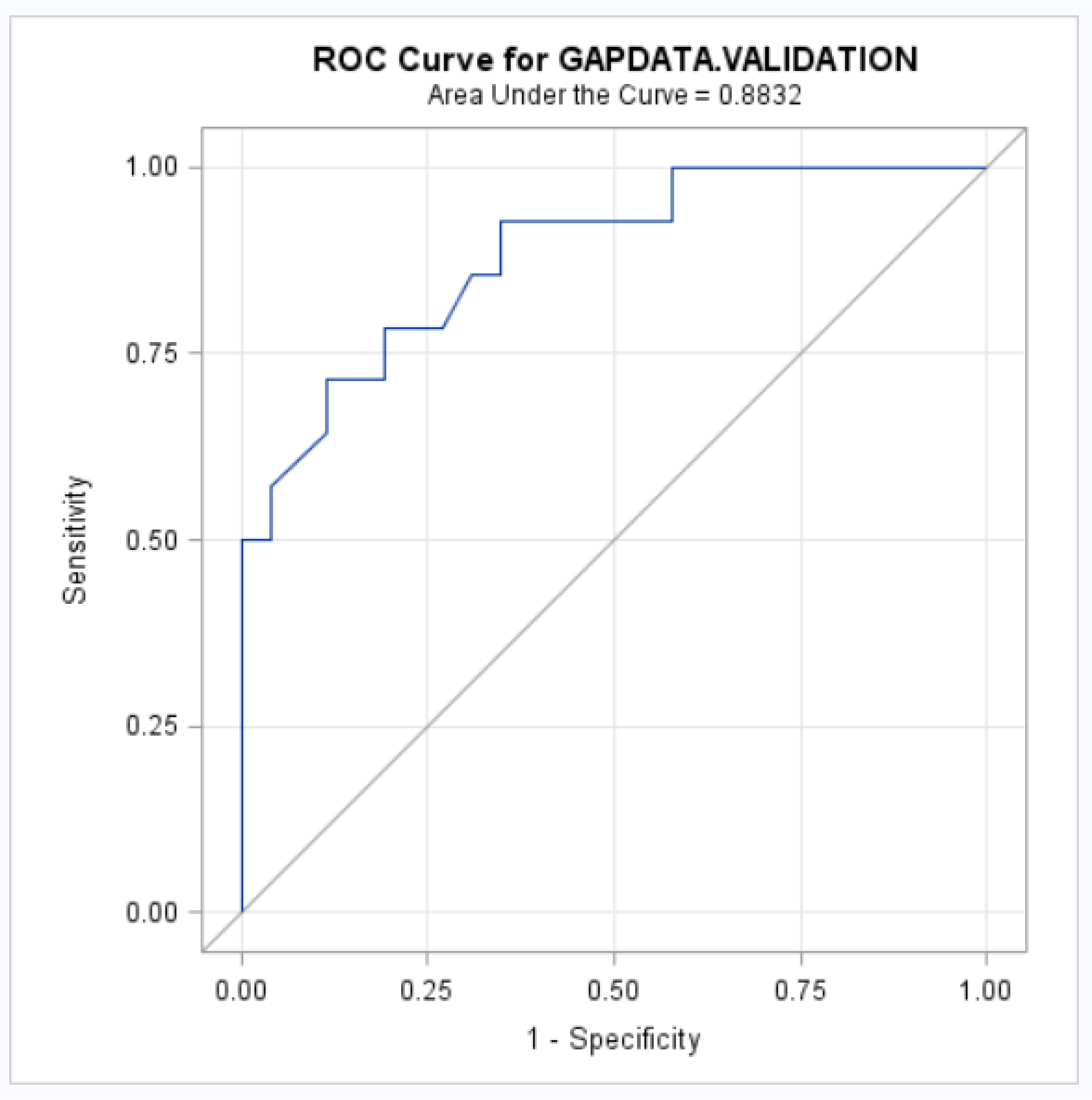


Figure 42: ROC Curve for MODEL I

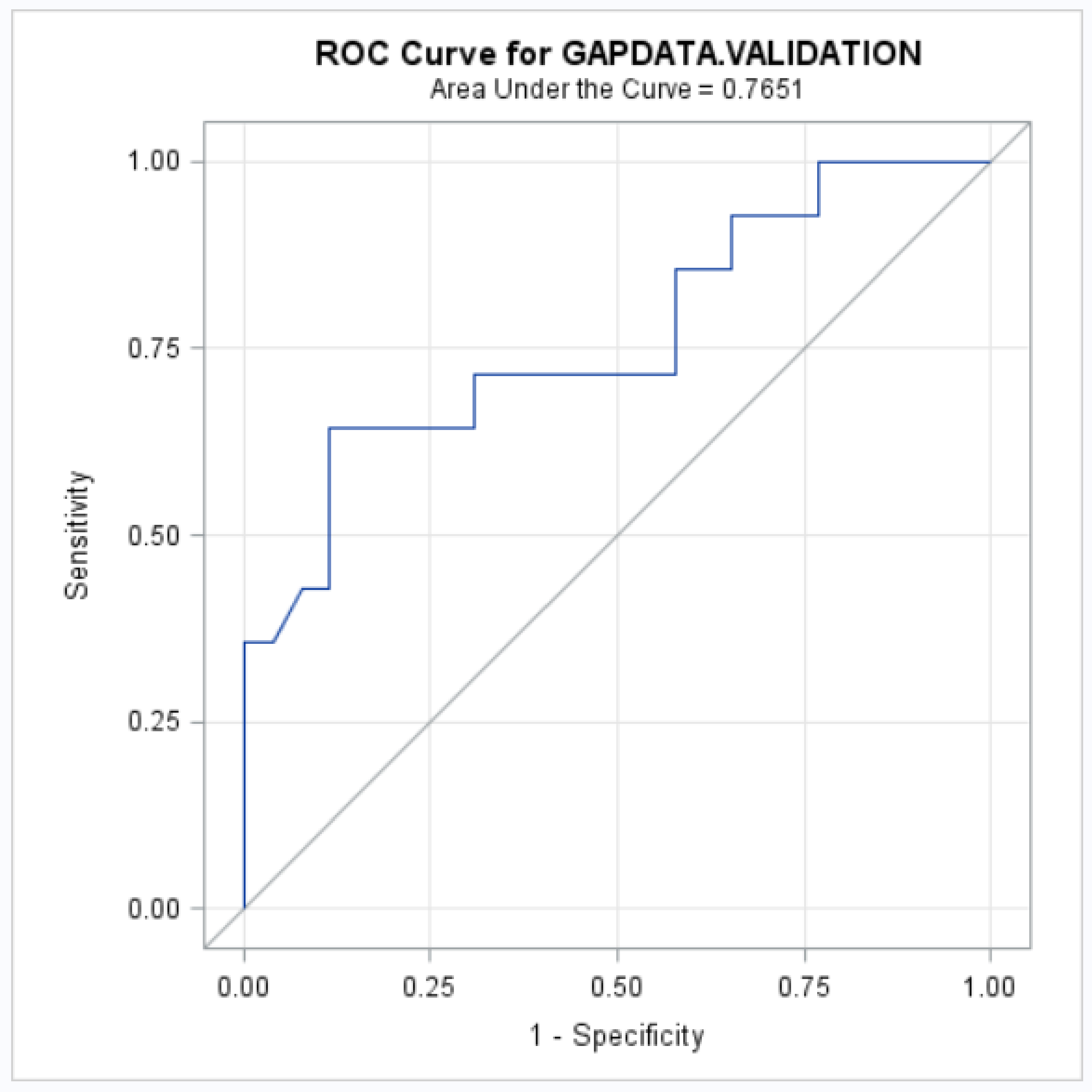


Figure 43: ROC Curve for Model II

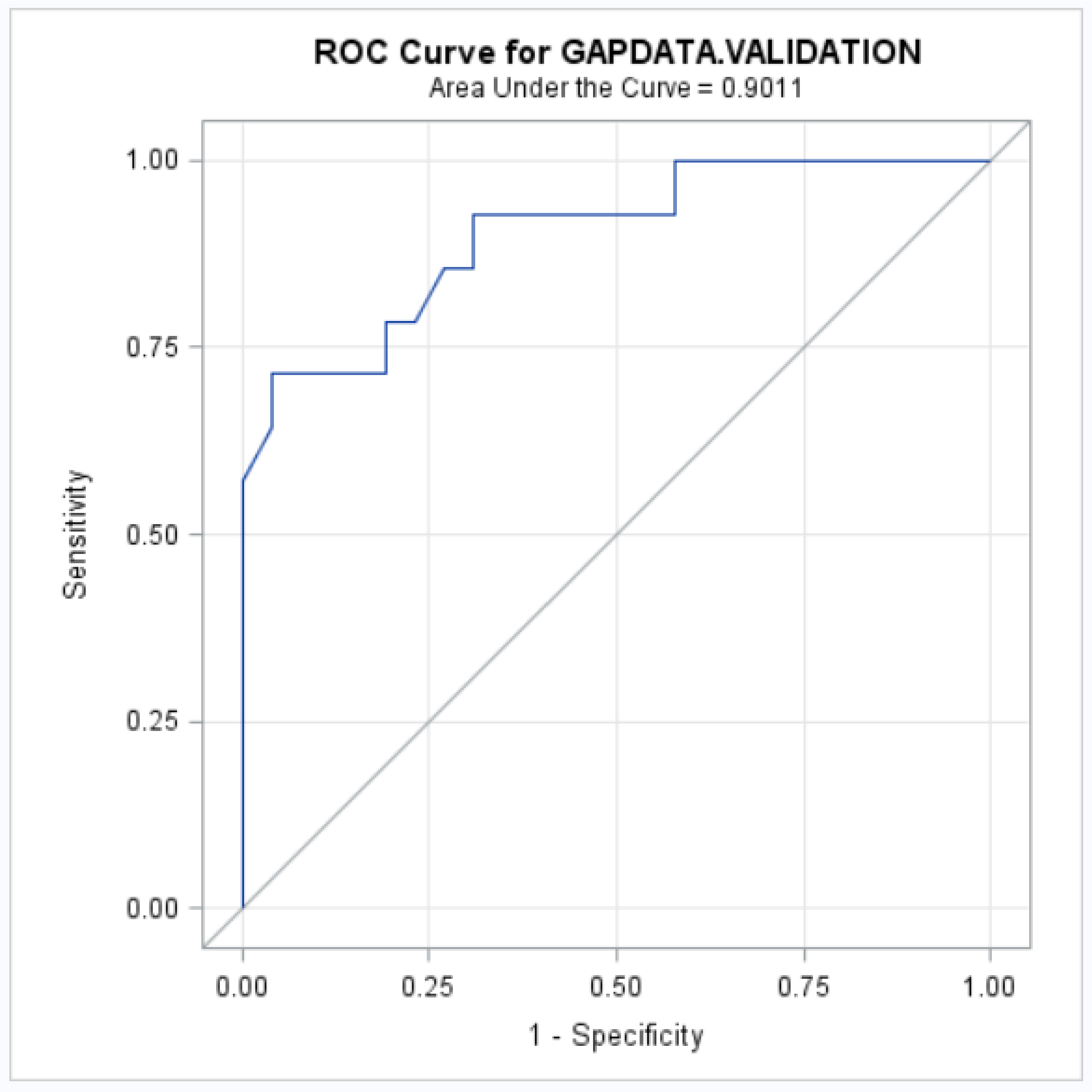


Figure 44: ROC Curve for MODEL III

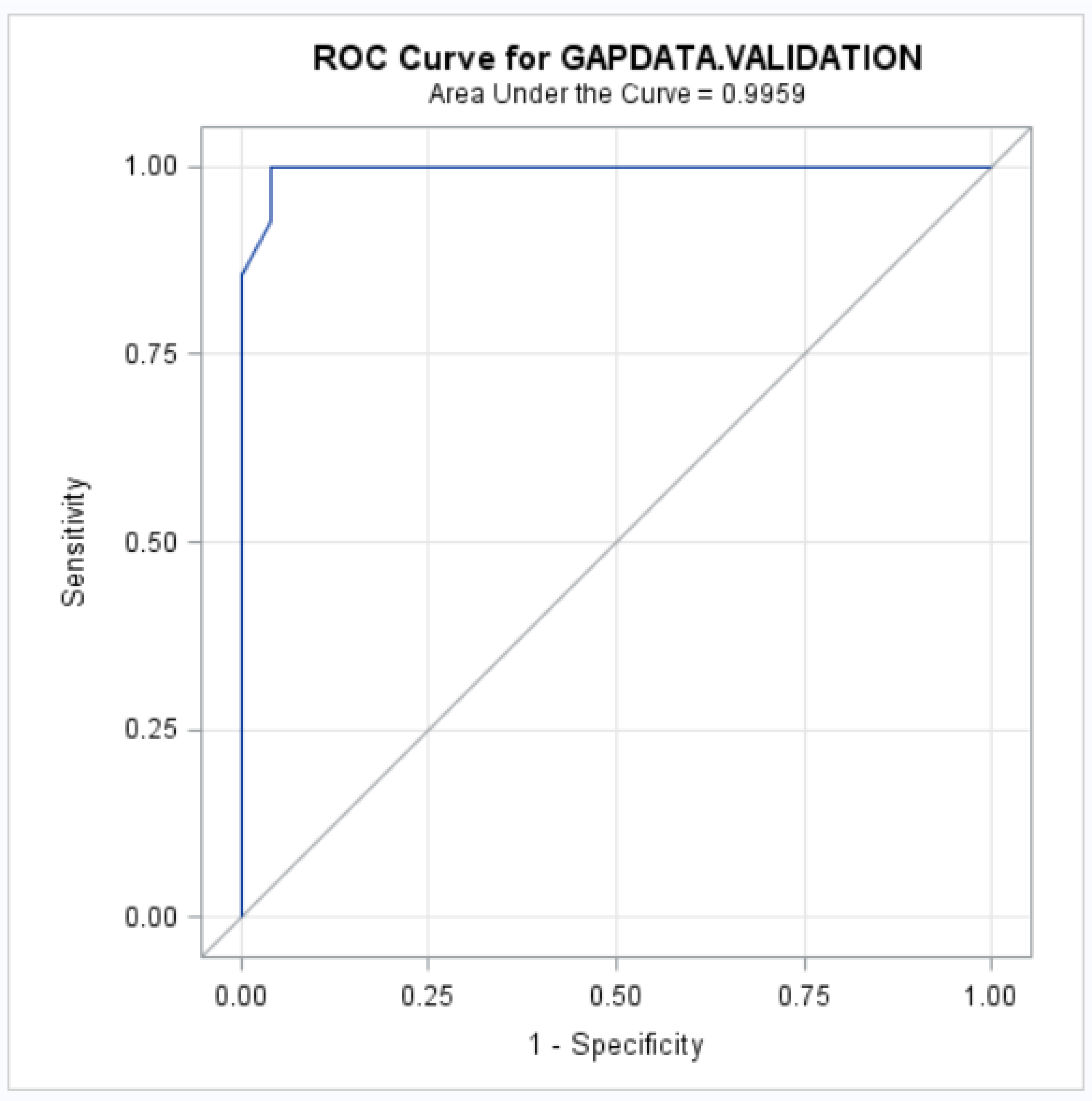


Figure 45: ROC Curve for Model IV

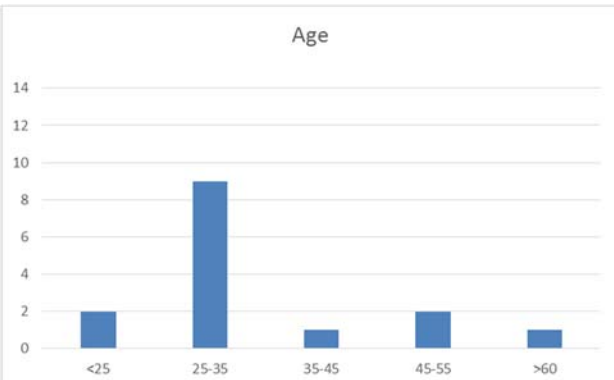
- This page intentionally left blanks -

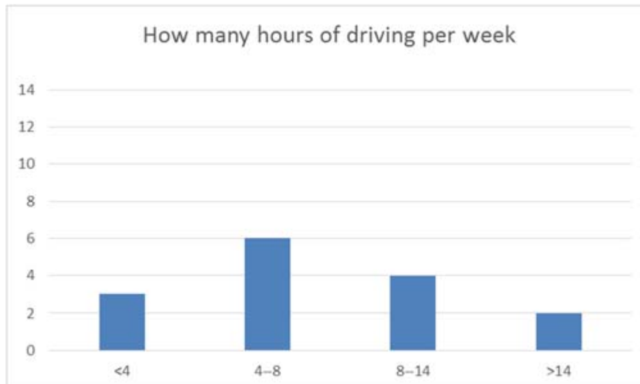
APPENDIX F: DRIVER STUDY DETAILED RESULTS

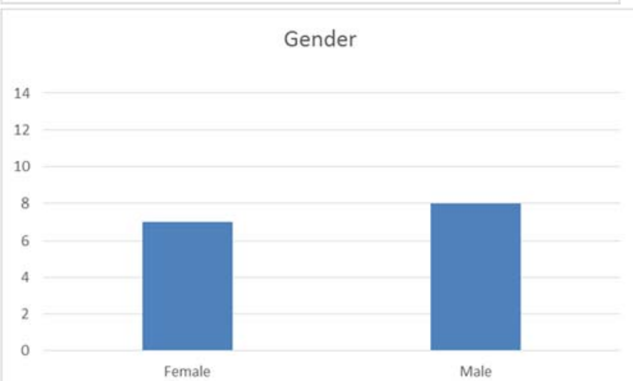
There are 15 drivers involved in the driver behavior study with different characteristics and driving experiences. They were asked to complete several questionnaires in terms of obtaining their demographic characteristics and general driving attitude, such as prescreening form, pre-driving questionnaire, between-route questionnaires and final questionnaires. Their information is summarized as following:

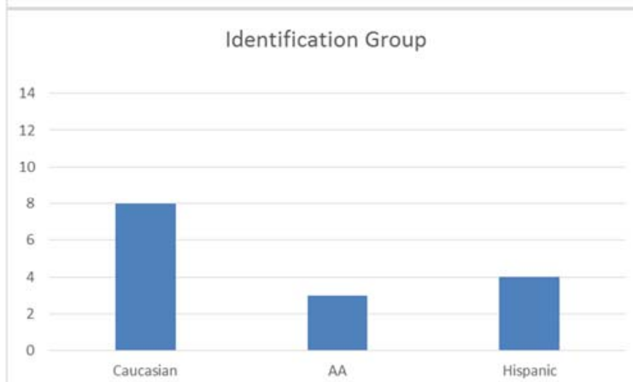
Demographic Characteristics

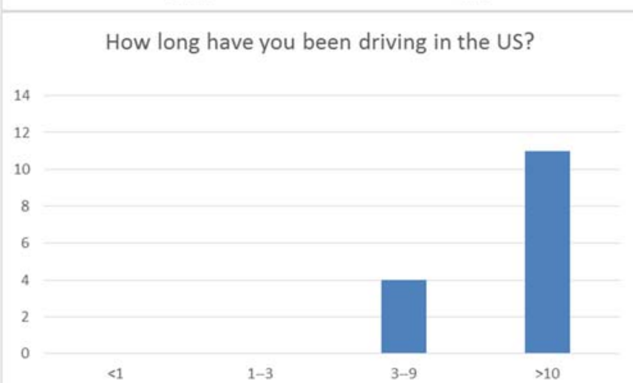
Prescreening form and pre-driving questionnaire provided the demographic characteristics and driving experience of each participant, as a way to classify driver types. Most of them have the driving experiences longer than 3 years. Almost half are females, and drivers in different age group were selected. Their aggressiveness could be observe through their perceived desired speed and lane-changing behaviors.

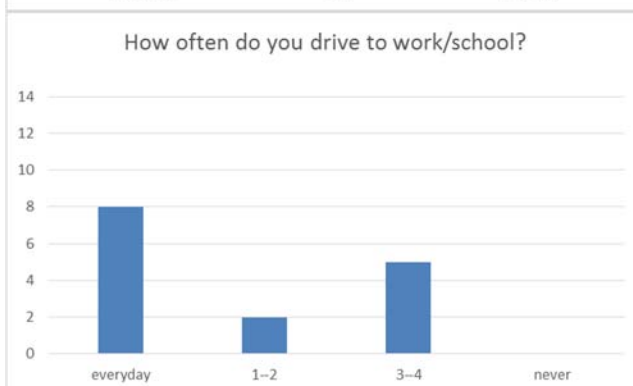












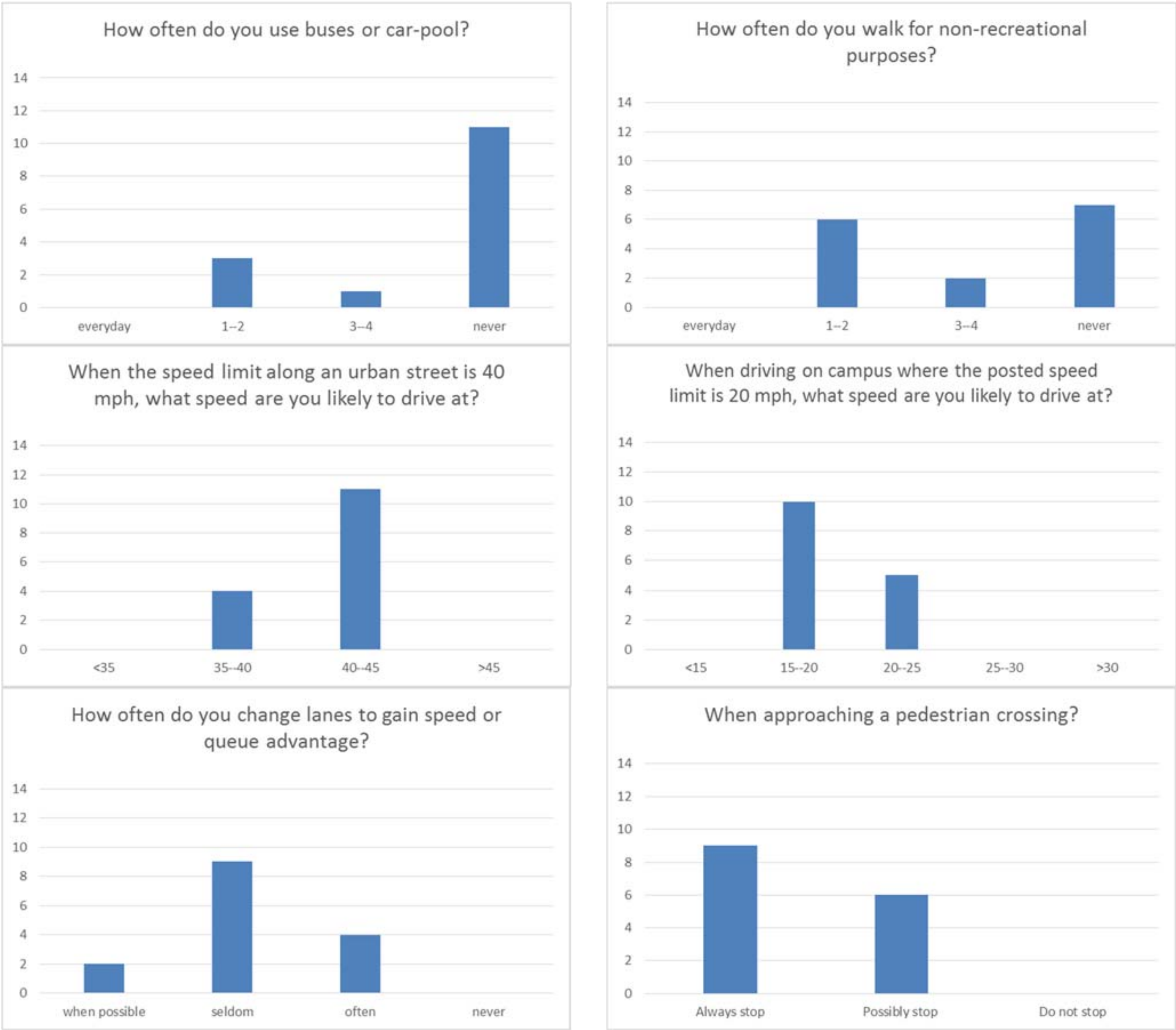


Figure 46: Summary of Driver Study Demographic Characteristics

Driving Experience during Routes

Between-route questionnaires were designed to ask about the participants' experience and reactions during routes towards pedestrians, bicyclist and transit. Most people noticed the pedestrian crossings, bicycle lanes and transit stops, but some of them were not aware of the local laws regarding the right-of-way of pedestrians and bicycles.

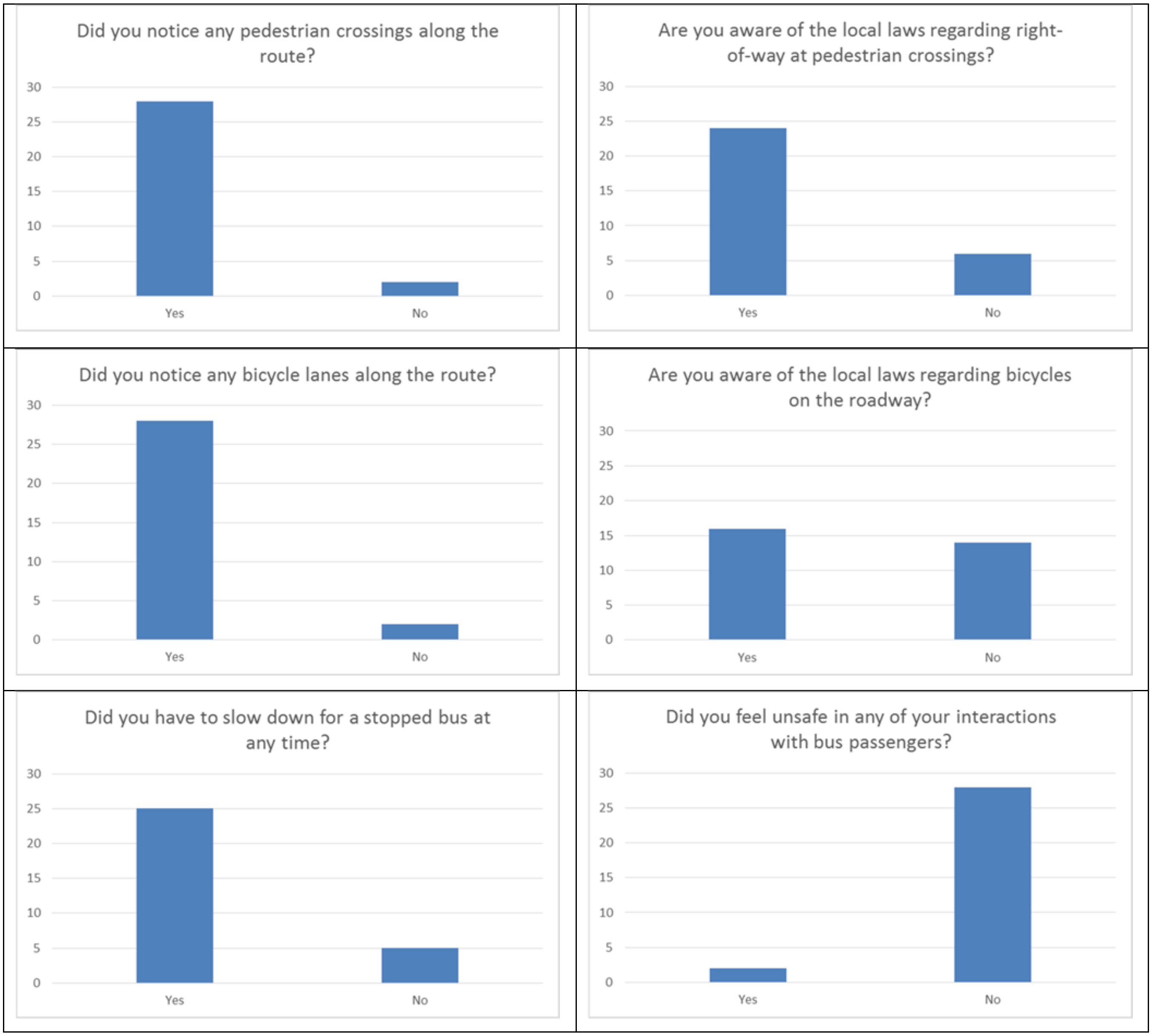


Figure 47: Summary of Driver Experience during Study

Perceptions on Interacting with Pedestrians

Final questionnaire was designed to ask about the participants' reactions and perceptions on pedestrian interactions, especially their yield decisions. It is shown that there are almost half drivers stating that they felt unsafe and did not expect some of the pedestrian actions during the experiment. And, most of the participants stated that they would like to slow down to 5-10 mph to let pedestrian cross if they must. Moreover, participants were provided with several conditions to react. This questionnaire (summarized in the table below) gives the basic idea of how drivers think about yielding to pedestrians at midblock locations in different scenarios.

Table 77: Summary of Driver Study Pedestrian Perception

Question		**1**		**2**		**3**		**4**		**5**		**6**	
Pedestrian	In Group	O		O		O		X		O		O	
	Wait on the Sidewalk	X		X		X		X		O		X	
Subject Vehicle	Eye Contact with Pedestrians	O		X		O		O		O		X	
	In Leading Position	X		X		O		O		X		X	
	Vehicle in Front Yields	-		-		O		O		-		-	
	Currently Yielding	O		O		O		O		X		O	
Answer		S	A	S	A	S	A	S	A	S	A	S	A
		13	2	15	0	13	2	13	2	15	0	13	2

Note: O – Yes; X – No; S – Stop; A – Accelerate.

Discussions

Most of the participants would like to yield if they see someone is currently crossing or behaves to cross. All the participants would stop if pedestrians are looking at them. Some would not yield if there's no eye contact. All the participants would remain stopping if there are other pedestrians coming up. Pedestrian crossing in group or individually does not affect driver's decisions. Some drivers would not yield if the front car does not stop for pedestrians. Crosswalks with flashing pedestrian-crossing light stand out through the routes. Jaywalking pedestrians always act out of their expectation, and jaywalker is the main cause of feeling unsafe.

Survey Forms

Prescreening Form

Table 78: Driver Study Pre-Screening Form Summary

ID #	Gender	Age	Identification Group	Driving Experience	Valid License	Occupation	How often driving to work/school	How many hours per week	Veh Type
1	1	>60	Caucasian	>10	1	other	everyday	8--14	Sedan
2	1	<25	Caucasian	3--9	1	other	usually	<4	sedan
3	1	25-35	Caucasian	>10	1	student	usually	4--8	sedan
4	0	25-35	Hispanic	3--9	1	student	usually	8--14	sedan
5	0	25-35	Caucasian	>10	1	other	sometimes	<4	sedan
6	1	45-55	African Am.	>10	1	other	everyday	>14	sedan
7	0	35-45	African Am	3--9	1	other	everyday	4--8	sedan
8	0	45-55	Caucasian	>10	1	driver	everyday	>14	truck
9	1	25-35	Hispanic	>10	1	other	everyday	4--8	sedan
10	0	25-35	Hispanic	3--9	1	student	usually	4--8	sedan
11	1	25-35	Caucasian	3--9	1	other	everyday	8--14	truck
12	0	25-35	Hispanic	>10	1	other	everyday	8--14	sedan
13	1	<25	Caucasian	3--9	1	other	everyday	4--8	sedan
14	1	25-35	African Am	>10	1	student	usually	4-8	sedan
15	0	25-35	Caucasian	>10	1	other	sometimes	<4	pickup

Pre-driving Form

Table 79 Driver Study Pre-Driving Form Summary

ID #		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
How long have you been driving in the US?	<1															
	1--3															
	3--9		1		1			1						1		
	>10	1		1		1	1		1	1	1	1	1		1	1
How often do you drive to work/school?	everyday	1					1	1	1	1		1	1	1		
	1--2					1										1
	3--4		1	1	1						1				1	
	never															
How often do you use buses or car-pool?	everyday															
	1--2										1				1	1
	3--4					1										
	never	1	1	1	1		1	1	1	1		1	1	1		
How often do you walk for non-recreational purposes?	everyday															
	1--2	1				1					1			1	1	1
	3--4		1	1												
	never				1		1	1	1	1		1	1			
How many hours do you spend driving per week?	<5		1			1								1		1
	5--10	1		1				1		1	1				1	
	>10				1		1		1			1	1			
When the speed limit along an urban street is 40 mph, what speed are you likely to drive at?	<35															
	35--40		1			1	1								1	
	40--45	1		1	1			1	1	1	1	1	1	1		1
	>45															
When driving on campus where the posted speed limit is 20 mph, what speed are you likely to drive at?	<15															
	15--20	1	1			1	1	1			1	1	1	1	1	
	20--25			1	1				1	1						1
	25--30															
	>30															
How often do you change lanes to gain speed or queue advantage?	when possible				1					1						
	seldom	1	1			1	1		1		1		1		1	1
	often			1				1				1		1		
	never															
When approaching a pedestrian crossing?	Always stop			1		1	1	1		1		1	1		1	1
	Possibly stop	1	1		1				1		1			1		
	Do not stop															

Questionnaire 1 (I)

Table 80: Driver Study Questionnaire 1 Results

ID #		1	2	3	4	5	6	7	8
Did you notice any pedestrian crossings along the route?	Yes		1	1	1	1	1	1	1
	No	1							
Which pedestrian crossing stood out for you and why?		Skater	Near Reitz (biker)	SPVD (biggest)	jaywalk	one with flashing ped light	No	one with ped sign	one with flashing ped light
Under what condition do you yield to pedestrians at the crosswalk?			If someone behaves to cross/crossing	if waiting, standing or crossing	Any time	If someone behaves to cross/crossing	No	If someone behaves to cross/crossing	If someone behaves to cross/crossing
Are you aware of the local laws regarding right-of-way at pedestrian crossings?	Yes	1		1	1		1	1	1
	No		1			1			
Did you notice any bicycle lanes along the route?	Yes	1	1	1	1	1	1	1	1
	No								
Did you feel unsafe in any of your interactions with bicyclists?	Yes		1	1		1			1
	No	1			1		1	1	
Are you aware of the local laws regarding bicycles on the roadway?	Yes	1					1	1	
	No		1	1	1	1			1
Did you have to slow down for a stopped bus at any time?	Yes		1	1	1	1	1	1	1
	No	1							
Did you feel unsafe in any of your interactions with bus passengers?	Yes								
	No	1	1	1	1	1	1	1	1

Questionnaire 1 (II)

ID #		9	10	11	12	13	14	15
Did you notice any pedestrian crossings along the route?	Yes	1	1	1	1	1	1	1
	No							
Which pedestrian crossing stood out for you and why?		No	No	No	Headphone ped	One was waiting	Talking on phone ped	guard
Under what condition do you yield to pedestrians at the crosswalk?		Upon seeing	If someone behaves to cross/crossing	Always	Always	If someone behaves to cross/crossing	If someone behaves to cross/crossing	If someone behaves to cross/crossing
Are you aware of the local laws regarding right-of-way at pedestrian crossings?	Yes	1	1	1	1		1	1
	No					1		
Did you notice any bicycle lanes along the route?	Yes	1	1	1	1	1	1	1
	No							
Did you feel unsafe in any of your interactions with bicyclists?	Yes		1	1				1
	No	1			1	1	1	
Are you aware of the local laws regarding bicycles on the roadway?	Yes	1		1		1	1	1
	No		1		1			
Did you have to slow down for a stopped bus at any time?	Yes	1	1	1	1	1	1	1
	No							
Did you feel unsafe in any of your interactions with bus passengers?	Yes						1 (a little)	
	No	1	1	1	1	1		1

Questionnaire 2 (I)

Table 81: Driver Study Questionnaire 2 Results

ID #		1	2	3	4	5	6	7	8
Did you notice any pedestrian crossings along the route?	Yes	1	1	1	1	1	1	1	1
	No								
Which pedestrian crossing stood out for you and why?		Skater	No	In front of the stadium (more of a medium size)	Exercising ped	one with stop sign	No	one with ped sign	jaywalk
Under what condition do you yield to pedestrians at the crosswalk?			If someone behaves to cross/crossing	If someone behaves to cross/crossing	Upon seeing	If someone behaves to cross/crossing	No	If someone behaves to cross/crossing	If someone behaves to cross/crossing
Are you aware of the local laws regarding right-of-way at pedestrian crossings?	Yes	1		1	1		1	1	1
	No		1			1			
Did you notice any bicycle lanes along the route?	Yes	1			1	1	1	1	1
	No		1	1					
Did you feel unsafe in any of your interactions with bicyclists?	Yes					1			1 for them
	No	1	1	1	1		1	1	
Are you aware of the local laws regarding bicycles on the roadway?	Yes	1					1	1	
	No		1	1	1	1			1
Did you have to slow down for a stopped bus at any time?	Yes		1			1	1	1	1
	No	1		1	1				
Did you feel unsafe in any of your interactions with bus passengers?	Yes								1 for them
	No	1	1	1	1	1	1	1	

Questionnaire 2 (II)

ID #		9	10	11	12	13	14	15
Did you notice any pedestrian crossings along the route?	Yes	1	1	1	1		1	1
	No					1		
Which pedestrian crossing stood out for you and why?		No	Wait for each other to go	No	Someone jumped out	No	One waits to cross, but no motorists yield	No
Under what condition do you yield to pedestrians at the crosswalk?		Upon seeing	If someone behaves to cross/crossing	always	Someone jumped out	If someone behaves to cross/crossing	If someone behaves to cross/crossing	If someone behaves to cross/crossing
Are you aware of the local laws regarding right-of-way at pedestrian crossings?	Yes	1	1	1	1		1	1
	No					1		
Did you notice any bicycle lanes along the route?	Yes	1	1	1	1	1	1	1
	No							
Did you feel unsafe in any of your interactions with bicyclists?	Yes		1	1	1			
	No	1				1	1	1
Are you aware of the local laws regarding bicycles on the roadway?	Yes	1		1		1	1	1
	No		1		1			
Did you have to slow down for a stopped bus at any time?	Yes	1	1	1		1	1	1
	No				1			
Did you feel unsafe in any of your interactions with bus passengers?	Yes							
	No	1	1	1	1	1	1	1

Final Questionnaire (I)

Table 82: Driver Study Final Questionnaire Results

ID #		1	2	3	4	5
Were there any conditions that felt unsafe?	Yes			busy intersections		ped/bike don't obey rules
	No	1	1		1	
Did pedestrians act in any ways that you did not expect?	Yes	Skater	Biker	sometimes no notice to ped and too late to stop	jaywalk	jaywalk
	No					
A single pedestrian is stopped on the sidewalk by a crosswalk you are approaching. They are looking down at their phone.	Accelerate					1
	Brake	1	1	1	1	
A single pedestrian is stopped on the sidewalk by a crosswalk you are approaching. They are looking at you and look like they want to cross.	Accelerate					
	Brake	1	1	1	1	1
A single pedestrian is stopped on the sidewalk by a crosswalk you are approaching. They are looking at traffic, but the car in front of you does not stop and there is a car very close behind you that may not see the pedestrian.	Accelerate				1	
	Brake	1	1	1		1
A group of 5 pedestrians are stopped on the sidewalk by a crosswalk you are approaching. They are looking at traffic, but the car in front of you does not stop and there is a car very close behind you that may not see the pedestrians.	Accelerate					
	Brake	1	1	1	1	1
How you drive when you must slow down to let a pedestrian cross, but do not have to come to a full stop. Consider when you begin to slow down, how quickly you slow down, where you stop slowing and begin to coast (if you choose to coast) and the speed at which you coast (if you choose to coast).			Pay attention to car in front to avoid accidents (no coasting)	slow down until ped pass my side (coast at 10 mph)	Slow down when ped approaching; stop when ped crossing	Depends on how close the ped are
You have just stopped to let a pedestrian cross at a crosswalk. You see another pedestrian approaching the crossing.	Remain Stopped	1	1	1	1	1
	Drive Through					
You are approaching a crosswalk and begin to slow to allow a pedestrian to cross. The pedestrian does not step out into the intersection, though they are looking at you.	Accelerate				1	
	Brake	1	1	1		1

Final Questionnaire (II)

ID #		6	7	8	9	10
Were there any conditions that felt unsafe?	Yes		jaywalk	step out of cars	turn left to stadium Rd	bikes
	No	1				
Did pedestrians act in any ways that you did not expect?	Yes	1	jaywalk	someone forced to a yield	jaywalk	jaywalk
	No					
A single pedestrian is stopped on the sidewalk by a crosswalk you are approaching. They are looking down at their phone.	Accelerate					
	Brake	1	1	1	1	1
A single pedestrian is stopped on the sidewalk by a crosswalk you are approaching. They are looking at you and look like they want to cross.	Accelerate					
	Brake	1	1	1	1	1
A single pedestrian is stopped on the sidewalk by a crosswalk you are approaching. They are looking at traffic, but the car in front of you does not stop and there is a car very close behind you that may not see the pedestrian.	Accelerate					
	Brake	1	1	1	1	1
A group of 5 pedestrians are stopped on the sidewalk by a crosswalk you are approaching. They are looking at traffic, but the car in front of you does not stop and there is a car very close behind you that may not see the pedestrians.	Accelerate					
	Brake	1	1	1	1	1
How you drive when you must slow down to let a pedestrian cross, but do not have to come to a full stop. Consider when you begin to slow down, how quickly you slow down, where you stop slowing and begin to coast (if you choose to coast) and the speed at which you coast (if you choose to coast).		Slow down and look both sides	Driving slow when there is ped. 5 mph under speed limit (no coast)	Always cover brake&coast	Slow down quickly	Begin slowing down slowly (not press on brakes)
You have just stopped to let a pedestrian cross at a crosswalk. You see another pedestrian approaching the crossing.	Remain Stopped	1	1	1	1	1
	Drive Through					
You are approaching a crosswalk and begin to slow to allow a pedestrian to cross. The pedestrian does not step out into the intersection, though they are looking at you.	Accelerate			1		
	Brake	1	1		1	1

Final Questionnaire (III)

ID #		11	12	13	14	15
Were there any conditions that felt unsafe?	Yes		1 too dark		1 jaywalk	Cyclist
	No	1		1		
Did pedestrians act in any ways that you did not expect?	Yes		1 jaywalk		ped forces to yield	jaywalk
	No	1		1		
A single pedestrian is stopped on the sidewalk by a crosswalk you are approaching. They are looking down at their phone.	Accelerate			1		
	Brake	1	1		1	1
A single pedestrian is stopped on the sidewalk by a crosswalk you are approaching. They are looking at you and look like they want to cross.	Accelerate					
	Brake	1	1	1	1	1
A single pedestrian is stopped on the sidewalk by a crosswalk you are approaching. They are looking at traffic, but the car in front of you does not stop and there is a car very close behind you that may not see the pedestrian.	Accelerate			1		
	Brake	1	1		1	1
A group of 5 pedestrians are stopped on the sidewalk by a crosswalk you are approaching. They are looking at traffic, but the car in front of you does not stop and there is a car very close behind you that may not see the pedestrians.	Accelerate			1	1	
	Brake	1	1			1
How you drive when you must slow down to let a pedestrian cross, but do not have to come to a full stop. Consider when you begin to slow down, how quickly you slow down, where you stop slowing and begin to coast (if you choose to coast) and the speed at which you coast (if you choose to coast).		Slow down when see	Just slow down if no ped is approaching	Slow down to 10 mph	5-10 mph	5-10 mph
You have just stopped to let a pedestrian cross at a crosswalk. You see another pedestrian approaching the crossing.	Remain Stopped	1	1	1	1	1
	Drive Through					
You are approaching a crosswalk and begin to slow to allow a pedestrian to cross. The pedestrian does not step out into the intersection, though they are looking at you.	Accelerate					
	Brake	1	1	1	1	1

This page intentionally left blank

APPENDIX G: UF INSTITUTIONAL REVIEW BOARD (IRB) FORM

UFIRB 02 – Social & Behavioral Research

Protocol Submission Form

This form must be typed. Send this form and the supporting documents to IRB02, PO Box 112250, Gainesville, FL 32611. Should you have questions about completing this form, call 352-392-0433.

Title of Protocol:	**In-vehicle Experiment on Driver Behavior Along Urban Streets**		
Principal Investigator:	**Dr. Lily Elefteriadou**		**UFID #:**
Degree / Title:	**Professor**	**Mailing Address:** (If on campus include PO Box address): **365 Weil Hall, PO Box 116580**	**Email:** **elefter@ce.ufl.edu**
Department:	**Civil and Coastal Engineering**		**Telephone #:** **352-392-9537, ext.1452**
Co-Investigator(s):	**Yinan Zheng, Thomas Chase Ph.D. students in Civil and Coastal Engineering**		**Email:** **zhengyinan@ufl.edu; rtchase@ufl.edu**
Supervisor *(If PI is student)*:			
Degree / Title:			
Department:			
Date of Proposed Research:	**March 2013 to November 2013**		

Source of Funding *(A copy of the grant proposal must be submitted with this protocol if funding is involved):*	STRIDE Center, US DOT

Scientific Purpose of the Study:
To capture important factors that impact driver behavior along urban streets. We are particularly interested in the behavior of drivers around pedestrian crossings. One of the goals of the study is to provide recommendations on modeling the interaction between pedestrians and drivers at pedestrian crossings.

Describe the Research Methodology in Non-Technical Language: (*Explain what will be done with or to the research participant.)*

In this research, data will be collected through an instrumented vehicle experiment as well as cameras positioned at critical sidewalks. Participants will be recruited and asked to drive through two or three pre-specified routes. We will record their speed throughout the network, their car-following and lane changing behavior, as well as their yielding and braking behavior in the vicinity of crosswalks as a function of the presence of pedestrians. Throughout the experiment (after the completion of each route) we will ask participants to stop and respond to a series of questions regarding the portion of the route they just drove. We will also conduct a survey at the end of the entire experiment.

Describe Potential Benefits:

The objective of this project is to understand and model driver behavior along urban streets, as well as yielding behaviors in the vicinity of pedestrian crosswalks. The study will consider driver characteristics and attitudes, crosswalk characteristics, and pedestrian characteristics. There are currently no such models available to facilitate urban street design and improve vehicular as well as pedestrian operations.

Describe Potential Risks: (*If risk of physical, psychological or economic harm may be involved, describe the steps taken to protect participant.)*

No risk is anticipated in the experiments. Each participant in the instrumented vehicle experiment would be provided the driving routes in advance (these would vary by participant), and will be accompanied by one or two researchers during driving. The research team will confirm that each participant has a valid drivers' license. Each participant will also be provided the opportunity to drive the instrumented vehicle in advance of the experiment, to familiarize themselves with its controls.

The researchers will minimize the interaction with participants during driving to avoid distracting them.

Describe How Participant(s) Will Be Recruited:

Participants will be selected based on age, gender, driving experience and vehicle ownership. A prescreening questionnaire has been developed to help identify qualified participants. Advertisements for the recruitment will be posted on campus and around the Gainesville area, and publicized on Craigslist websites. The participant recruitment will target residents of the Gainesville area who are generally familiar with the area.

Other criteria for the participant recruitment are set as:
1. Must be at least 25 years old;
2. Must be a regular driver with driving experience no less than one year;
3. Should indicate a willingness to participate in the experiment.

All participants will be asked to fill out a background survey form before the beginning of the experiments. The background survey form provides information related to the driving habits of the participants and their perceived degree of aggressiveness.

Maximum Number of Participants (to be approached with consent)	**30**	**Age Range of Participants:**	**25 to 60**	**Amount of Compensation/ course credit:**	**$50**

Describe the Informed Consent Process. (Attach a Copy of the Informed Consent Document. See http://irb.ufl.edu/irb02/samples.html for examples of consent.)

The informed consent form explains and clearly states that participation is optional and that the outcome will be summarized in a manner that does not identify any participants. A separate copy of the informed consent form

(attached) will be used to advise potential participants and obtain voluntary agreement at the beginning of the experiments. *The informed consent form describes the data collection process and informs the participants of their expected duties during the experiment. The associated risks and benefits, as well as the expected duration of the survey after finishing driving and the compensation of the participants are also stated. If the vehicle is not working properly during the experiment, then the participant should stop at the nearest possible location to terminate the experiment. Risks associated with distractions during driving task will be minimized as the researchers will only communicate with the participants to inform them (if necessary) of upcoming driving maneuvers.*

(SIGNATURE SECTION)		
Principal Investigator(s) Signature:		**Date:**
Co-Investigator(s) Signature(s):		**Date:**
Supervisor's Signature (if PI is a student):		**Date:**
Department Chair Signature:		**Date:**