PEDESTRIAN-VEHICLE INTERACTIONS AT SEMI-CONTROLLED CROSSWALKS: EXPLANATORY METRICS AND MODELS

by

Yunchang Zhang

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Civil Engineering

Lyles School of Civil Engineering

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Jon Fricker, Chair
> Lyles School of Civil Engineering

Dr. Andrew Tarko
> Lyles School of Civil Engineering

Dr. Konstantina Gkritza
> Lyles School of Civil Engineering

Dr. Robert W. Proctor
> Department of Psychological Sciences

Dr. Vincent Duffy
> Department of Industrial Engineering

Approved by:
> Dr. Dulcy Abraham
> Head of the Graduate Program

ACKNOWLEDGMENTS

I hope to start off by thanking my parents – Wenrui Zhang and Rongfen Sun for continuously encouraging me within this journey. Without your love and support, I would not experience such nice scenes in my master journey. Besides, I also want to show my sincere appreciation to my cousins – Zhifeng Zhang, Zhiwei Zhang and Jing Sun. Thank you for taking care of my family and helping me to deal with family stuffs when I was not there. There have been highs and lows, but you all have been there as always.

In Addition, I want to express my deepest gratitude to my advisor – Dr. Jon D. Fricker who encouraged me to do what I really want to do and kindly supported me to finish my master thesis for more than one and half years. Your passion in teaching, patience, optimism and even jokes impressed me a lot. I cannot wait to work with you in my Ph.D. journey. I would like to thank my committee members – Dr. Konstantina Gkritza and Dr. Andrew Tarko for their guidance on statistical modeling parts of my thesis. I would also thank Dr. Vincent Duffy and Dr. Robert Proctor for your incomparable suggestions in human factor aspects in my thesis. I'm really looking forward to cooperating with you all in my future research.

Finally, I would like to thank my Purdue friends – Yu Qiao, Yuntao (Ross) Guo, Christos Gkartzonikas and Theodora Konstantinou for your suggestions and guidance in my thesis stuffs. I would also express my deepest gratitude to my best friends - Boyu Yang and Minhao Xiao for their endless support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GroupSize	The number of pedestrians in the curb area, including the subject pedestrian
AgeRange	Estimated age range for subject pedestrian(s) (1: 0-10; 2: 10-30; 3: 30-50; and 4: >50).
Sex	Sex of subject pedestrian
Hesitation	Does the pedestrian slow down or wait at curb?
Distraction	Does a pedestrian approach and/or cross while using a cellphone or talking?
FlowWith	The number of pedestrians already crossing in the crosswalk in the same direction when subject pedestrian arrives at curb area
FlowAgainst	The number of pedestrians already crossing in the crosswalk in the opposite direction when subject pedestrian arrives at curb area
FlowOn	Total number of pedestrians already crossing in the crosswalk when an interaction occurs (FlowWith + FlowAgainst).
SameDirec	The number of pedestrians present in the curb area crossing in the same direction as the subject pedestrian
DiffDirec	The number of pedestrians present in a curb area with crossing direction opposite of the subject pedestrian
PedWait	Total number of pedestrians waiting in the curb areas when an interaction occurs (SameDirec + DiffDirec)
ApprSpeed	The approach speed of interacted vehicles when a pedestrian enters the curb area. (mph)
SlowsDown	Does a vehicle slow down or stop on the approach to the crosswalk when a pedestrian enters the curb area?
CloseFollow	Does the interacted vehicle have a close follower when an interaction occurs?
AdjVeh	Is a vehicle already present in the adjacent lane when a motorist begins to interact with a pedestrian?
Distance	The distance of interacted vehicle(s) to subject pedestrians when interaction begins. (in feet)
NoF	Is pedestrian entering curb area on the near side or far side of the approaching vehicle's lane?
Pedestrian Outcomes	Cross: Y = 1; Wait/Yield: Y = 0
Vehicle Response	Level of vehicle deceleration when pedestrians enter crosswalks (3 = stops; 2 = slows down; 1 = Does not slow down).

ABSTRACT

Author: Zhang, Yunchang. MS
Institution: Purdue University
Degree Received: May 2019
Title: Pedestrian-Vehicle Interactions at Semi-Controlled Crosswalks: Explanatory Metrics and
      Models.
Committee Chair: Jon Fricker

A large number of crosswalks are indicated by pavement markings and signs but are not signal-controlled. In this study, such a location is called "semi-controlled". In locations where such a crosswalk has moderate amounts of pedestrian and vehicle traffic, pedestrians and motorists often engage in a non-verbal "negotiation" to determine who should proceed first.

In this study, 3400 pedestrian-motorist non-verbal interactions at such semi-controlled crosswalks were recorded by video. The crosswalk locations observed during the study underwent a conversion from one-way operation in Spring 2017 to two-way operation in Spring 2018. This offered a rare opportunity to collect and analyze data for the same location under two conditions.

This research explored factors that could be associated with pedestrian crossing behavior and motorist *likelihood of decelerating*. A mixed effects logit model and binary logistic regression were utilized to identify factors that influence the likelihood of pedestrian crossing under specific conditions. The complementary motorist models used generalized ordered logistic regression to identify factors that impact a driver's *likelihood of decelerating*, which was found to be a more useful factor than *likelihood of yielding to pedestrian*. The data showed that 56.5% of drivers slowed down or stopped for pedestrians on the one-way street. This value rose to 63.9% on the same street after it had been converted to 2-way operation. Moreover, two-way operation eliminated the effects of the presence of other vehicles on driver behavior.

Also investigated were factors that could influence how long a pedestrian is likely to wait at such semi-controlled crosswalks. Two types of models were proposed to correlate pedestrian waiting time with various covariates. First, Survival models were developed to analyze pedestrian wait time based on the first-event analysis. Second, multi-state Markov models were introduced to

correlate the dynamic process between recurrent events. Combining the first-event and recurrent events analyses addressed the drawbacks of both methods. Findings from the before-and-after study can contribute to developing operational and control strategies to improve the level of service at such unsignalized crosswalks.

The results of this study can contribute to policies and/or control strategies that will improve the efficiency of semi-controlled and similar crosswalks. This type of crosswalk is common, so the benefits of well-supported strategies could be substantial.

Keywords: Crossings; Pedestrian-Motorist Interaction; Pedestrian Wait Behavior

CHAPTER 1.　　INTRODUCTION

1.1　　Background and Problem Statement

The National Highway Traffic Safety Administration (NHTSA, 2016) reported 5,987 pedestrians killed in traffic crashes in the United States. Compared with the number of pedestrian fatalities (5,495) in 2015, thus is a 9% increase. Pedestrian improper crossings and driver failure to yield right-of-way were two main factors contributing to pedestrian collisions. "State Law Yield to Pedestrian Within Crosswalk" signs (Figure 1(a)) are commonly used at unsignalized pedestrian crosswalks where pedestrian-motorist interaction frequently occurs. Pedestrians using crosswalks with "State Law Yield to Pedestrian Within Crosswalk" signs have priority over approaching vehicles. Nevertheless, observations confirm that confusion exists among pedestrians and motorists, because the sign's message is subject to varying interpretations. Sometimes a motorist stops and lets pedestrians standing at the curb cross the street, and sometimes drivers fail to yield to pedestrians entering the crosswalk. In many cases, a non-verbal "negotiation" takes place between pedestrians and motorists, to determine who should proceed first. No dangerous situations were observed in this study. The result is usually delay to pedestrian and/or motorist. However, if misunderstandings during "negotiations" happen, there could be safety issues.

(a)　　　　(b)　　　　　　　　　　(c)

Figure 1 Semi-Controlled Crosswalk

Video recordings were made of pedestrians using crosswalks at locations with 'State Law …' signs. I characterize such crossing sites as "semi-controlled", because the crosswalks are marked (Figure 1(a)), the 'State Law …' signs are present, but no signals are installed there. The videos were examined, looking for patterns and relationships that may exist during the interactions between pedestrians and drivers.

Critical gap is an important parameter in gap acceptance theory when considering pedestrian-motorist interaction. The definition of the critical gap for pedestrian can be defined as the minimum vehicle headway that a pedestrian can accept to undertake a crossing maneuver. However, gap acceptance theory may not by itself be adequate to explain pedestrian behavior at "semi-controlled" crossing locations. This is because of the interactions between pedestrians and motorists, in which non-verbal "negotiations" for priority often take place. Although, in theory, pedestrians can assert their priority to cross in these places, video observation shows that some drivers do not slow down for pedestrians. Consequently, a detailed inventory of pedestrian and motorist behaviors captured on video recordings at crosswalks has been created, making possible analyses that not only supplement gap acceptance methods to model pedestrian behavior, but also analyze factors that influence driver decisions.

1.2   Study Objective and Research Questions

Video recordings at crosswalks were created, making possible analyses that not only supplement gap acceptance methods to model pedestrian behavior, but also analyze factors that influence driver decisions. The primary objective of this project is to establish a framework for investigating pedestrian-motorist non-verbal "negotiations" at semi-controlled locations. This research focuses on the crosswalks on North University Street at Second Street (Figure 1(b) (c)) on the Purdue University campus. This location was chosen because (1) it has a variety of crossing conditions and (2) it was converted from a one-way street in 2017 to two-way operation in 2018. Having a video record of pedestrian-motorist interactions permitted a detailed examination of those interactions. Furthermore, the change from one-way traffic to two-way traffic provided a rare opportunity to study the behavior of a similar population of pedestrians and motorists at a location that underwent a significant change.

In this study, four primary research questions were pursued:

1. What factors can describe and explain the pedestrian-motorist interactions at semi-controlled crossing locations?
2. How will the pedestrian-motorist interaction change if a one-way street is changed to two-way operation?
3. Which characteristics will determine how long a pedestrian waits at a semi-controlled crossing location?
4. How can pedestrian-vehicle interaction and pedestrian waiting behavior at semi-controlled crossing locations be modeled?

CHAPTER 2.   REVIEW OF LITERATURE

2.1   Pedestrian Crossing Behavior

Gap acceptance theory has been commonly used to model pedestrian decision-making. Probability-based approaches and modeling approaches are two main forms of pedestrian gap acceptance studies. Sun et al. (2002) proposed a Pedestrian Gap Acceptance method to model pedestrian decision strategies. They considered the probability of accepting a gap as a random variable that was obtained by fitting distributions to field data. Zhuang and Wu (2011) used statistical methods to analyze pedestrian crossing patterns (eye contact and running) and used gap acceptance models to describe pedestrian crossing behaviors at unsignalized crosswalks in China. Yannis et al. (2013) also employed a probability-based approach (lognormal regression) to test pedestrian gap acceptance in front of approaching vehicles at mid-block crossings. A binary logit model was used to explore the effect of gaps and other related parameters that affected pedestrian decision strategies. Kadali et al. (2014) compared the effectiveness of a non-linear model (artificial neural network) with a linear model (multivariate regression) in establishing a relationship between pedestrian gap acceptance behavior and explanatory factors. These studies listed above were primarily based on pedestrian gap acceptance behavior. However, gap acceptance theory has some limitations and may be inadequate to explain pedestrian crossing behavior. Some researchers have explored the family of discrete choice models to describe pedestrian crossing strategies. Himanen and Kumala (1988) developed a multinomial logit model to interpret the "negotiations" between drivers and pedestrians on crosswalks. Papadimitriou et al. (2012, 2016) designed surveys to investigate the impact of human factors on pedestrian crossing decisions by means of principal component analysis. Furthermore, a theoretical framework to model pedestrian crossing decision making process in urban trips was proposed via different discrete choice models. Lord et al. (2018) designed a questionnaire to understand the relationships between crossing strategies and the perceptions of the elderly, and logistic regression models were applied to explain the observed behaviors.

2.2   Driver Yielding Behavior

Discrete choice models have been widely used to analyze driver behavior, considering several explanatory parameters under different traffic or concurrent conditions at unsignalized crosswalks. Schroeder and Rouphail (2010, 2013) used logistic regression to predict driver yielding behavior at "semi-controlled" crosswalks and roundabouts based on vehicle dynamics, pedestrian characteristics, and environmental conditions. Sucha et al. (2017) studied the communications between pedestrians and drivers with respect to physical gestures (waving and eye contact), then took advantage of logistic regression to explore factors that had an influence on driver yielding behavior. Cloutier et al. (2018) applied a mixed-effects logit model to evaluate factors related to the likelihood of interactions between pedestrians and motorists.

Other researchers explored game theory to explain the interactions between pedestrians and vehicles (Guan et al., 2016; Bjørnskau, 2017; Camara et al., 2018).

2.3   Pedestrian Wait Behavior

Recently, researchers also considered pedestrian wait time as one of the most important performance metrics in pedestrian-motorist interaction. Survival models have been used in transportation studies, especially for travel time and wait time, because of their flexibility in dealing with duration-based data (Washington et al., 2010). Nonparametric, semi-parametric and fully parametric survival models have been utilized to explore the effects of human factors on pedestrian waiting behavior. The Kaplan-Meier estimator (Kaplan and Meier, 1958) and the Lee-Carter method (Lee and Carter, 1992) are two prevailing approaches for non-parametric survival models, which provide practical estimates of survival probabilities and a raw graphical representation of the survival distribution (Washington et al., 2010). In transportation studies, the nonparametric Kaplan–Meier estimator was widely applied to investigate pedestrian wait duration before making unsafe crossings at signalized intersections (Tiwari et al. 2007; Guo et al, 2011). Cox (1972) developed a semi-parametric survival model duration model that included the effects of covariates. Guo et al. (2011) applied the semi-parametric Cox proportional hazard model to analyze influences of personal characteristics and external environment on pedestrian wait duration at signalized intersections in China based on both legal and illegal crossings. Instead of

non-parametric and semi-parametric models, fully parametric models were developed by applying alternative statistical distributions for the baseline hazard function. Fully parametric models are recently popular because their capacities of fitting different types of baseline hazard functions. Guo et al. (2012) further extended their previous research by applying both nonparametric and fully parametric models to explore the effects of human factors on pedestrian waiting behavior at signalized intersections. Li (2013) focused on pedestrian wait time at signalized intersections, and U-shaped distribution of pedestrian wait time was found. Yang et al. (2015) proposed hazard-based duration approach to study the wait time for cyclists and electronic bike riders.

There has been research on how long a pedestrian will wait at unsignalized crosswalks (Hamed, 2001). How long a pedestrian decides to wait reflects how safe he/she perceives it is to cross the roadway. Pedestrians may feel unsafe and tend to wait longer when motorists exhibit aggressive driving behaviors. An investigation of pedestrian crossing behaviors and wait durations at such locations can be useful in developing policies and control strategies to enhance a pedestrian's perceived safety, reduce pedestrian delay, and improve the level of service (LOS) of a unsignalized intersections.

Pedestrian waiting behavior consists of not only a single event, but also recurrent events. Survival models have limitations in dealing with recurrent events in a dataset. Traditional survival models consider only the first event as the critical event. However, the first event analysis only considers the first event, while the subsequent events are ignored. Secondly, the repeated events are treated as identical treatments, which should be modeled differently.

The Anderson-Gill (AG) model (Andersen and Gill, 1982) extended the Cox model to handle recurrent events by applying a counting process. They applied a baseline hazard function to all events. The subject of interest was the number of repeated events, given a specific period. The AG model is widely used in medical science, but it has a strong proportional odds assumption that, in practice, is difficult to be satisfied. Prentice, Williams, and Peterson (PWP) (1981) further recurrent events regression analysis by stratifying the events as ordered series, which allowed separate baseline hazard functions and coefficients to vary across events. Wei, Lin and Weissfeld (1989) also developed a model for a repeated events modeling approach. However, this model is

less efficient than PWP, due to its complicated nature. Although, AG, PWP and WLW are three classic models that have been widely investigated in the repeated events analysis, these three models focused only on the probability of occurrence, rather than the transition process between repeated events. Multi-State Models would be appropriate alternatives for recurrent events analysis.

There are two main research gaps in the existing research literature concerning pedestrian-motorist interaction. First, gap acceptance theory has been the prevailing method to analyze pedestrian behavior, but it may not be adequate at semi-controlled or controlled locations, where pedestrians can assert the priority to cross. Second, most research has studied either pedestrian behavior or motorist behavior separately. The research questions in our study call for an integrated framework that considers the potential relationships between pedestrian behavior and motorist behavior. One the other hand, the existing literature shows that survival models have the potential to model pedestrian waiting behavior at unsignalized semi-controlled crosswalks as a the first-event analysis. Additionally, multi-state models can accommodate the dynamic modeling of pedestrian waiting behavior as recurrent event analysis. In the following sections, both the first-event approach and recurrent events modeling approach are discussed.

CHAPTER 3.   RESEARCH METHODOLOGY

3.1   Study Site Description

Video recordings were made at the unsignalized pedestrian crossing location shown in Figure 1(b).

North University Street at Second Street

North University Street, at the T intersection with Second Street, has two lanes, each 12 ft wide, with a speed limit of 25 miles/h. The two crosswalks are used by students and staff walking between central campus to the east and parking facilities and residences to the west. The first set of videos were made in Spring 2017, when North University Street was a one-way northbound street. By the time the second set of videos were made in Spring 2018, the streets had been converted to two-way operation. This conversion provided a rare opportunity to study pedestrian-motorist interaction at the same site under different conditions. The two sets of video recordings were made at four different time periods (7:40 - 8:20; 12:40 - 13:25; 13:20 - 14:00; 16:20 - 17:00), when moderate traffic volumes and pedestrian flows were observed. The authors recorded approximately 3 hours of video for each set, resulting in a total of 3400 pedestrian-motorist interactions.

3.2   Definition

3.2.1   *Interaction*

The time-synchronized videos were processed in the laboratory. Interaction-based data were extracted to support the development of statistical models to investigate the "negotiation" between pedestrian and motorist. In this study, we define the interaction between pedestrian and motorist as the behavior of either party when in the area of influence of the other. The area of influence is defined by a vehicle being close enough to the crosswalk to affect the pedestrian's crossing decision. We assume (based on behavior seen in the video) that pedestrians make their crossing decisions within the curb area (within 2 meters of the street). From our observations, most pedestrians take a definite look for vehicles within the curb area, and most pedestrians wait within the curb area if drivers do not give an indication of yielding to them. Based on situations seen in the video recordings, an interaction can happen in several ways (Fricker and Zhang, 2019):

1. A pedestrian arrives at the curb and crosses immediately (without delay) while a vehicle accelerates, slows down or stops to avoid a conflict.

2. A pedestrian arrives at the curb and slows down or stops, but a vehicle slows down or stops to yield to the pedestrian.

3. A pedestrian arrives at the curb and slows down or stops, while a vehicle slows down, but does not yield to the pedestrian.

4. A pedestrian arrives at the curb and slows down or stops, while a vehicle keeps a constant speed or accelerates, not yielding to pedestrian.

An interaction does not occur if:

5. A pedestrian arrives at the curb area, but there is no vehicle close enough to the crosswalk to affect the pedestrian's crossing decision.

6. A vehicle approaches the crosswalk, but there is no pedestrian present who is attempting to cross.

3.2.2 *First Event & Critical Event*

Based on the definition of interaction, we aimed to investigate pedestrian wait time when pedestrian-motorist interactions happen. A pedestrian can interact with either one vehicle or multiple vehicles, so that the pedestrian wait time dataset is mixed with single interaction event and recurrent interaction events. In survival models, the first event is considered as a "critical event". We assume that the first event has the greatest impact on the pedestrian's crossing decision than the other event(s) did. Then the first event is considered as a critical event and the first interacted vehicle is called critical vehicle.

3.2.3 *Recurrent Events*

Multi-state semi-Markov models allow the estimation of the instantaneous impact of factors on the probability of transition from different states. By applying multi-state semi-Markov model, we modeled transitions in Figure 2 in terms of three states:

1. A pedestrian reaches the curb area as a vehicle approaches, so that an interaction occurs.

2. Pedestrian rejects the lag (and, if necessary, subsequent gaps).

3. Pedestrian accepts the lag (or gap) and crosses the street.

The potential transition for this model is defined as:

- Transition 1-3: accept the lag directly (vehicle yields).

- Transition 1-2: reject the lag and await recurrent gaps (vehicle fails to yield).

- Transition 2-2: reject following gaps. (This transition is not considered in this thesis)

- Transition 2-3: accept a subsequent gap.

An individual whose first transition was 1-3, is considered simultaneously with the transition as 1-2.
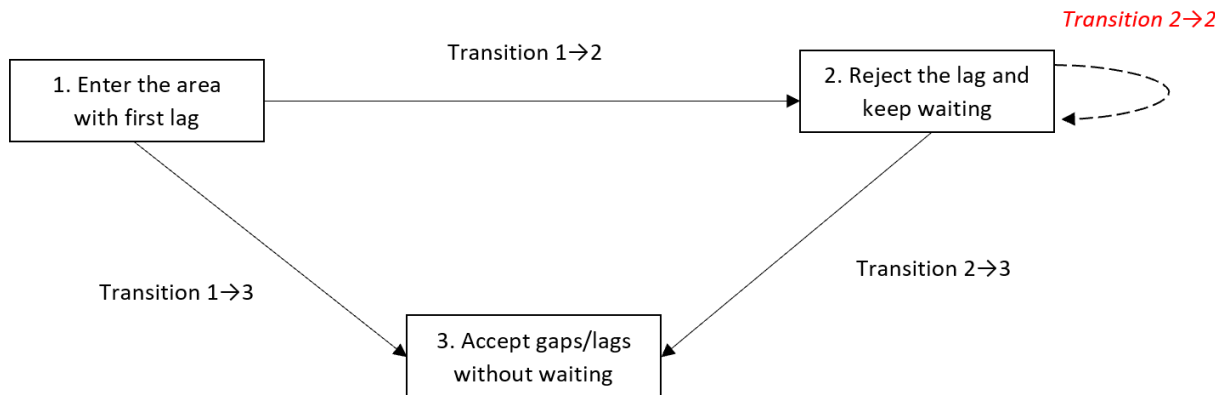
Figure 2 Multi-State Framework for Pedestrian Waiting Behavior

CHAPTER 4.   PEDESTRIAN-VEHICLE INTERACTION

4.1   Explanatory Variables

The work has been published in Fricker, J.D., and Zhang Y. (2019) and it is reprinted here with the permission from Fricker, J.D., & Zhang, Y. (in press). "Modeling Pedestrian and Motorist Behavior at Semi-Controlled Crosswalks: The Effect of a Change from One-Way to Two-Way Street Operation". *Transportation Research Record*.

In order to explore pedestrian-vehicle interaction, all explanatory variables, including pedestrian characteristics and dynamics, vehicle dynamics, and environmental conditions are documented in Table 1. Whenever a pedestrian entered the curb area while a motorist was present in the area of influence, values for all the variables listed in Table 1 were manually recorded.

Table 1 Explanatory Variables

Parameters	Variable Description	Value	Pedestrian Model	Motorist Model
Pedestrian Characteristic and Dynamics				
GroupSize	The number of pedestrians in the curb area, including the subject pedestrian	Integer	√	
AgeRange	Estimated age range for subject pedestrian(s) (1: 0-10; 2: 10-30; 3: 30-50; and 4: >50).	Indicators	√	
Sex	Sex of subject pedestrian	Male=1; Female=0	√	√
Hesitation	Does the pedestrian slow down or wait at curb?	Y=1/N=0	√	√
Distraction	Does a pedestrian approach and/or cross while using a cellphone or talking?	Y=1/N=0	√	√
FlowWith	The number of pedestrians already crossing in the crosswalk in the same direction when subject pedestrian arrives at curb area	Integer	√	
FlowAgainst	The number of pedestrians already crossing in the crosswalk in the opposite direction when subject pedestrian arrives at curb area	Integer	√	

Table 1 continued

FlowOn	Total number of pedestrians already crossing in the crosswalk when an interaction occurs (FlowWith + FlowAgainst).	Integer		√
SameDirec	The number of pedestrians present in the curb area crossing in the same direction as the subject pedestrian	Integer	√	
DiffDirec	The number of pedestrians present in a curb area with crossing direction opposite of the subject pedestrian	Integer	√	
PedWait	Total number of pedestrians waiting in the curb areas when an interaction occurs (SameDirec + DiffDirec)	Integer		√
Vehicle Dynamics				
ApprSpeed	The approach speed of interacted vehicles when a pedestrian enters the curb area. (mph)	Float	√	√
SlowsDown	Does a vehicle slow down or stop on the approach to the crosswalk when a pedestrian enters the curb area?	Y=1/N=0	√	
CloseFollow	Does the interacted vehicle have a close follower when an interaction occurs?	Y=1/N=0	√	√
AdjVeh	Is a vehicle already present in the adjacent lane when a motorist begins to interact with a pedestrian?	Y=1/N=0	√	√
Environmental Characteristics				
Distance	The distance of interacted vehicle(s) to subject pedestrians when interaction begins. (in feet)	Float	√	√
NoF	Is pedestrian entering curb area on the near side or far side of the approaching vehicle's lane?	Near=0; Far=1	√	√
Response Behavior				
Pedestrian Outcomes	Cross: Y = 1; Wait/Yield: Y = 0	Indicators	√	
Vehicle Response	Level of vehicle deceleration when pedestrians enter crosswalks (3 = stops; 2 = slows down; 1 = Does not slow down).	Indicators		√

For Hesitation parameter, the 75% percentile of pedestrian wait time for non-hesitation behavior is 1.57s (one-way) while the 25% percentile pedestrian wait time for hesitation behavior is 1.735s. The numbers for two-way case are 1.80s and 2.51s separately. We re-examine the 25% overlapped wait time for Hesitation and Non-Hesitation behavior through watching the videos repeatedly.

For CloseFollow parameter, we define that the object vehicle has close follower (CloseFollow =1), if the vehicle has a follower at a short headway of approximately 2-4 seconds, which has been defined similarly in former research literature (Schroeder and Rouphail, 2011).

4.1.1 *Pedestrian Behavior*

Based on the recorded interactions between pedestrians and motorists, a predictive model of pedestrian crossing behavior might be developed. There are two potential outcomes that describe pedestrian behavior:

- Pedestrian Crosses (Y=1): the motorist in the interacted vehicle provides an opportunity for the pedestrian to cross.
- Pedestrian Yields (Y=0): a pedestrian offers the motorist an opportunity to pass through the crosswalk first in an interaction.

The variables that seemed appropriate to use in a model of pedestrian behavior are indicated by check marks in the "Pedestrian Model" column of Table 1.

4.1.2 *Motorist Behavior*

Based on numerous recorded interactions, a driver's likelihood to decelerate was determined to be a key factor in the negotiation between pedestrian and motorist. A driver's likelihood to decelerate had better explanatory power than likelihood to yield, because, in an interaction, a motorist could slow down initially, but have the pedestrian wave to the motorist to go first. In this situation, the driver's action to decelerate is considered an important element of the interaction, even if the motorist did not eventually yield to the pedestrian. Consequently, by assigning levels of deceleration to each motorist, the potential outcomes for a motorist in an interaction are:

- Level 1. Keep a constant speed or accelerate: a motorist does not slow down, and the interaction does not cause delays for the motorist.
- Level 2. Decelerate but do not fully stop: a motorist decelerates during an interaction but does not fully stop and incurs some delay.
- Level 3. A motorist stops to accommodate a pedestrian and incurs a delay that is usually greater than in Level 2.

4.2   Descriptive Statistics

We examined descriptive statistics to look for general trends in the data. In videos of one-way University Street, there were 1,759 interactions, involving 1,133 pedestrians and 498 motorists. (Some pedestrians interacted with more than one motorist, and vice versa.) Of the total interactions, 1,240 (70.5 percent) resulted in pedestrians crossing, while 519 (29.5 percent) of total interactions were of the "Pedestrian Yield to Motorist" type. Furthermore, in 993 out of 1,759 cases (56.5 percent), motorists chose to slow down or stop for pedestrians. When University Street was in two-way operation, the number of interactions was 1,574 (involving 933 pedestrians and 506 motorists). Of the total interactions, 1,061 (67.4 percent) had pedestrians crossing, while 513 (32.3 percent) of total interactions were of the "Pedestrian Yield to Motorist" type. Moreover, in 1,005 out of 1,574 cases (63.9 percent), motorists chose to slow down or stop for pedestrians during interactions on the two-way street.

In Table 2, descriptive statistics of all explanatory variables are shown. The asterisks in the "Mean" columns indicate the level of significance of the difference between the mean value of variables for one-way operation and two-way operation, found using t tests. Overall, the data showed a significantly higher percent of pedestrians hesitating (Hesitation, 54.9%) on the one-way street than with two-way traffic (49.9%).

In one-way cases, 29.7% of vehicles have a vehicle following closely behind (CloseFollow) when they are involved in an interaction. However, the value of CloseFollow for two-way streets is 50.6%. In addition, in one-way cases, 43.4% of vehicles arrived at the study area with a vehicle present in the adjacent lane. However, this number for the two-way street is 52.1%. Finally, the

average value of the distance from vehicle to crosswalk is 74.6 feet on the one-way street, which is significantly different from the two-way case (64.7 feet).

Table 2 Descriptive Statistics

Variables	One-way		Two-way	
	Mean	Std.Dev.	Mean	Std.Dev.
GroupSize	2.53***	2.331	2.054***	1.483
AgeRange	2.193***	0.425	2.321***	0.542
Sex	0.429*	0.495	0.471*	0.499
Hesitation	0.549**	0.498	0.499**	0.5
Distraction	0.146	0.353	0.163	0.37
FlowWith	1.229	2.049	1.151	1.847
FlowAgainst	0.875	1.581	0.792	1.607
FlowOn	2.103	3.015	1.943	2.8
SameDirec	1.017	1.402	0.931	1.204
DiffDirec	0.629**	1.123	0.525**	1.003
PedWait	1.646	2.005	1.457	1.725
AppSpeed	8.543	6.939	8.355	7.794
SlowsDown	0.565***	0.496	0.639***	0.481
CloseFollow	0.297***	0.457	0.506***	0.5
AdjVeh	0.434***	0.496	0.521***	0.5
Distance	74.592***	54.729	64.673***	49.12
NoF	1.505	0.5	1.502	0.5
Pedestrian Outcomes	0.705	0.456	0.674	0.469
Vehicle Response	1.851**	0.837	1.945**	0.812
* p<.05; ** p<.01; *** p<.001				

4.3   Modeling Approach

4.3.1   *Pedestrian Model*

A pedestrian's Cross or Yield behavior has a binary outcome: Cross (Y=1) or Yield (Y=0). Commonly, a binary logistic regression model is applied to estimate the probability that a particular choice happened, based on a series of explanatory variables. Using this method, a linear model was built with explanatory variables by transforming the outcomes into Prob{Y=1}. The logistic regression model assumes that, for every explanatory property (Harrell, 2015),

$$\text{logit}(Y=1\mid X) = \log\left[\frac{1-P(Y=1)}{P(Y=1)}\right] = \sum_{i=1}^{I} \beta_i X_i + C \tag{1}$$

where C is the intercept and β_i is the change in the log odds per unit change in X_i, while all other variables are unchanged. Equation (2) can be used to describe the correlates between odds (Y) and variables (Harrell, 2015),

$$\text{odds}(Y=1\,|\,X)=\exp(X\beta) \tag{2}$$

The regression parameters can also be written in terms of odds ratios. The odds that $Y = 1$ when X_j is increased by d, divided by the odds at X_j is

$$\text{odds ratio} = \frac{\text{odds}\left\{Y=1\,|\,X_1,X_2,...,X_j+d,...,X_k\right\}}{\text{odds}\left\{Y=1\,|\,X_1,X_2,...,X_j,...,X_k\right\}}=\exp\left[\beta_jX_j+\beta_jd-\beta_jX_j\right]=\exp\left(\beta_jd\right) \tag{3}$$

Mixed effects logit model has been widely used in transportation safety research due to its flexibility in model structure. Compared with binary logistic regression, the mixed-effects logit model considers the probability as the integral of the standard logit model over a density distribution of a parameter (Ye et al., 2014). The mixed effects logit model can be written as:

$$P(Y=1)=\int\frac{\exp(\alpha_k+\beta_kX)}{\sum_{\forall k}\exp(\alpha_k+\beta_kX)}f(\beta\,|\,\theta)d\beta \tag{4}$$

The θs in the model are normally distributed in both one-way case and two-way case. Estimated values are shown in Table 3. Normally, the mixed effects logit model is compared with binary logistic regression together and AIC is a critical indicator for model selection, which balance the fitness and model complexity. The AIC can be expressed as (Akaike, 1987):

$$AIC=-2\ln(\text{likelihood})+2k \tag{5}$$

where k is the number of parameters. In this part of the study, both binary logistic regression and mixed effects logit model were tried. The model that best represented the data was chosen based on the AIC. The model results are shown in Table 3.

Table 3 Binary Logistic Regression Results for Pedestrian Models

Variables	One-way		Two-way		Interacted Coefficients of Combined Data	
	Logistic Model	Mixed-Effects Logit	Logistic Model	Mixed-Effects Logit	Logistic Model	Mixed-Effects Logit
GroupSize	—	—	0.443**	0.452**	0.424*	0.424*
AgeRange	—	—	—	—	—	—
Sex	—	—	—	—	—	—
Hesitation	-4.767***	-7.878***	-3.699***	-3.788***	—	—
Distraction	—	—	—	—	—	—
FlowWith	0.331***	0.522**	—	—	-0.456***	-0.573***
FlowAgainst	0.28**	0.425*	0.227*	0.232*	—	—
SameDirec	—	—	—	—	—	—
DiffDirec	—	—	0.384**	0.391**	—	—
AppSpeed	-0.147***	-0.224***	-0.111***	-0.114***	—	—
SlowsDown	2.593***	4.352***	3.03***	3.091***	—	—
CloseFollow	—	—	—	—	—	—
AdjVeh	-0.546**	-1.033**	—	—	0.634*	0.865*
Distance	0.042***	0.069***	0.036***	0.036***	—	—
NoF	-0.689***	-0.932**	-0.65**	-0.665**	—	—
Constant	4.604***	6.647***	—	—	3.209***	3.972***
Log Likelihood	-336.817	-320.589	-328.075	-328.008	-658.987	-648.043
θ	-	Esti: 2.750 std: 0.508	-	Esti: 0.0025 std: 3.806	-	Esti: 1.47 std: 0.253
AIC	693.6333	663.1773	676.1501	678.0157	1383.973	1364.087
BIC	748.3583	723.3748	729.7638	736.9908	1585.657	1571.882
Pseudo R^2	0.6843	-	0.6698	-	0.6805	-
Observations	1759		1574		3333	
* p<.05; ** p<.01; *** p<.001						

4.3.2 *Pedestrian Model Discussion*

From Table 3, it can be seen that significant variables were estimated similarly in both models. For one-way data, mixed logit works slightly better based on the AIC and BIC. However, for two-way operation, binary logistic regression performs better. Due to the similarity of results from two models, to avoid confusion, we mainly discuss results and findings based on the binary logistic regression model.

The significant variables in the logistic regression model (Table 3) suggest that, in a pedestrian-motorist interaction, pedestrians are more likely to decide to cross …

- if a pedestrian is assertive without hesitation (Hesitation). Pedestrian-motorist interactions were compared with only the Hesitation variable changing, while keeping other variables equal to their average values. If a pedestrian slows down at the curb while interacting with a motorist, the coefficient indicates that the probability of pedestrian crossing with Hesitation = 1 is,

$$\sum_{i=1}^{I} \beta_i X_{i,1} + C = -4.767*1 + 0.331 * 1.229 + 0.28 * 0.875 - 0.796*2.19 - 0.147*8.543 + 2.593*0.566 \tag{6}$$
$$- 0.546 * 0.434 + 0.042*74.592 - 0.689*1.505 + 4.604 = 0.81633$$

$$P(Y = 1 \mid Hesitation = 1) = \frac{e^{\sum_{i=1}^{I} \beta_i X_{i,1} + C}}{1 + e^{\sum_{i=1}^{I} \beta_i X_{i,1} + C}} = 0.693 = 69.3\% \tag{7}$$

The probability of a pedestrian crossing under the same conditions, but with Hesitation = 0, is

$$\sum_{i=1}^{I} \beta_i X_{i,2} + C = -4.767*0 + 0.331 * 1.229 + 0.28 * 0.875 - 0.796*2.19 - 0.147*8.543 + 2.593*0.566 \tag{8}$$
$$- 0.546 * 0.434 + 0.042*74.592 - 0.689*1.505 + 4.604 = 5.5833$$

$$P(Y = 1 \mid Hesitation = 0) = \frac{e^{\sum_{i=1}^{I} \beta_i X_{i,2} + C}}{1 + e^{\sum_{i=1}^{I} \beta_i X_{i,2} + C}} = 0.996 = 99.6\% \tag{9}$$

- if a driver decelerates (SlowsDown). If a driver slows down during an interaction, the probability of a pedestrian crossing is 98.4%, while the probability for a non-slowing down event is 81.9%. This is a clear indication of a motorist yielding to a pedestrian during the interaction. The effects of the Hesitation and SlowsDown variables are shown in Figure 3 in terms of Distance.
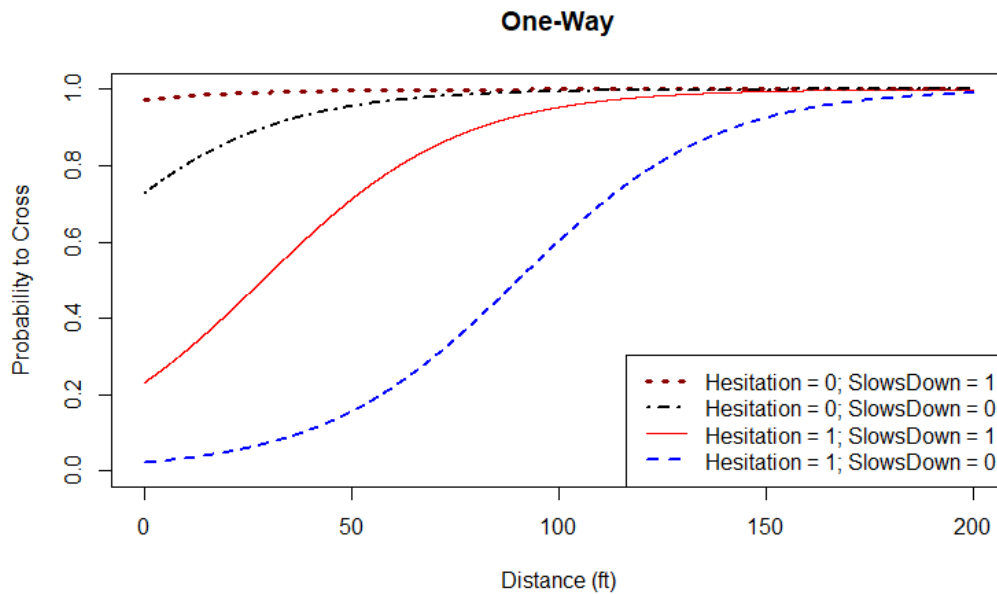
Figure 3 Effects of Hesitation and SlowsDown on Pedestrian Crossing Behavior

- if a car is approaching at a lower speed (AppSpeed). If the approach speed of a vehicle is 10 mph, the probability of pedestrian crossing is 71%, while the probabilities are 36.0% for 20 mph and 11.4% for 30 mph. Other studies (e.g., Brüde and Jörgen, 1993; Leaf and Preusser, 2006) found similar relationships.

- if the distance (in feet) between a pedestrian and motorist is great (Distance). If the interaction distance between vehicle and object crosswalk is 20 feet, the probability of pedestrian crossing is 66.5%, while the probabilities are 91.4% for 60 feet and 98.3% for 100ft. The effects of Distance and AppSpeed variables can be seen in Figure 4.
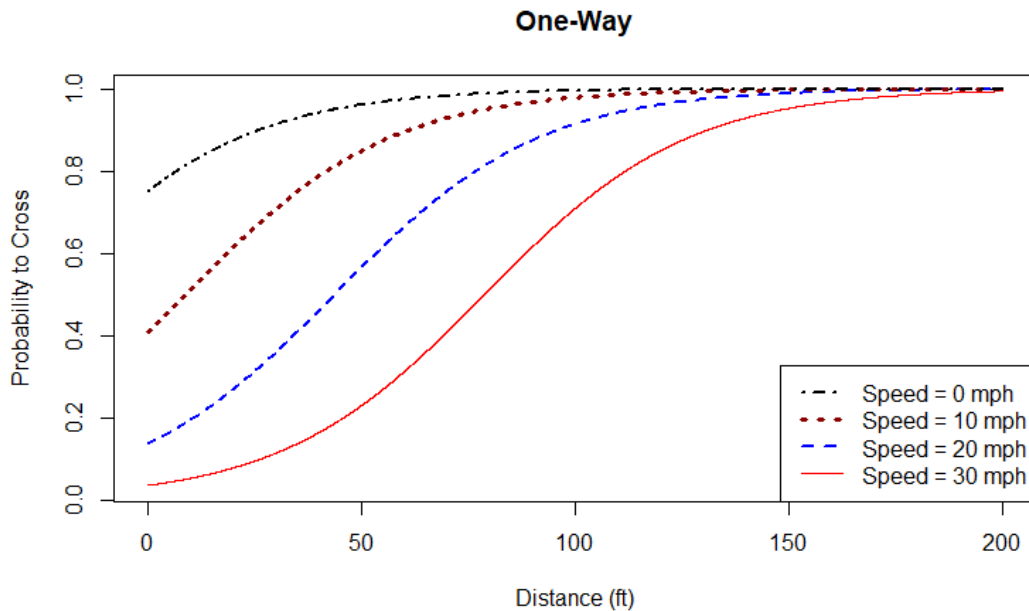
Figure 4 Effects of Distance and Speed on Pedestrian Crossing Behavior

- if there is no vehicle in the adjacent lane (AdjVeh). If pedestrian has to interact with two different vehicles in different lanes, the probability of a pedestrian crossing is 93.5%, compared with the case in which there is no adjacent vehicle (the object pedestrian interacts with only one vehicle in any lane) (96.1%). Schroeder and Rouphail (2011) found a similar relationship.

- if a vehicle is in the near lane (NoF=0). The probability of a pedestrian deciding to cross is 96.5% when a far lane interaction occurs, to 98.2% with a near-lane interaction. A plausible explanation is that, all else being equal, a pedestrian is more confident when crossing before a vehicle in the near lane arrives, compared with the longer crossing distance to the far lane and the risk of being trapped in the crosswalk while waiting for a far lane vehicle to yield or proceed. Effects of AdjVeh and NoF variables can be seen in Figure 5 in terms of Distance.

One-Way

Figure 5 Effects of AdjVeh and NoF on Pedestrian Crossing Behavior

- if other pedestrians are using the crosswalk (FlowWith and FlowAgainst). The effects are shown in Figure 6.
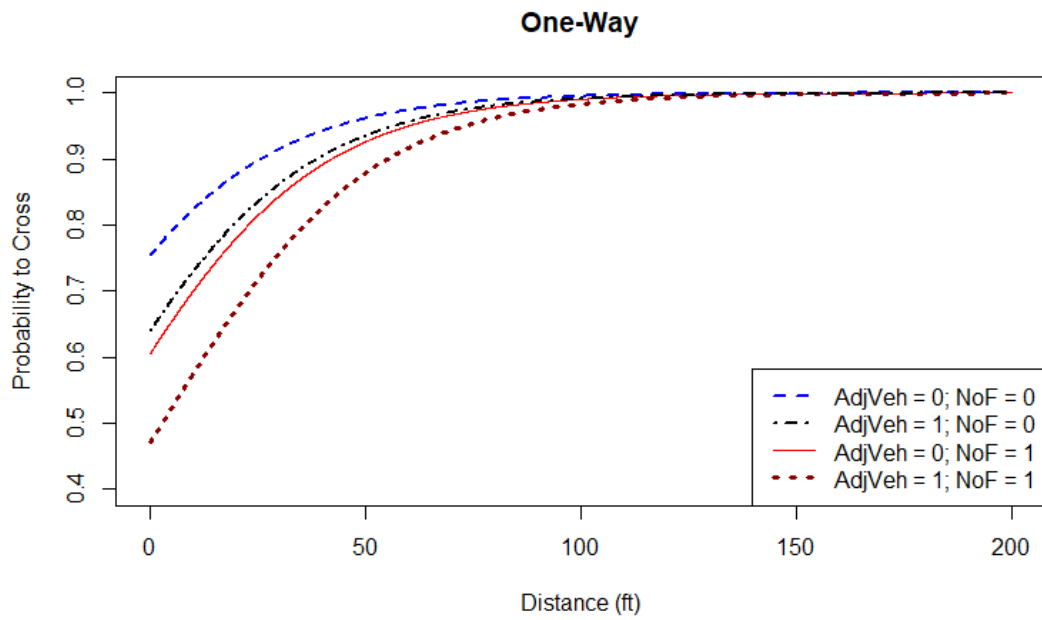
One-Way
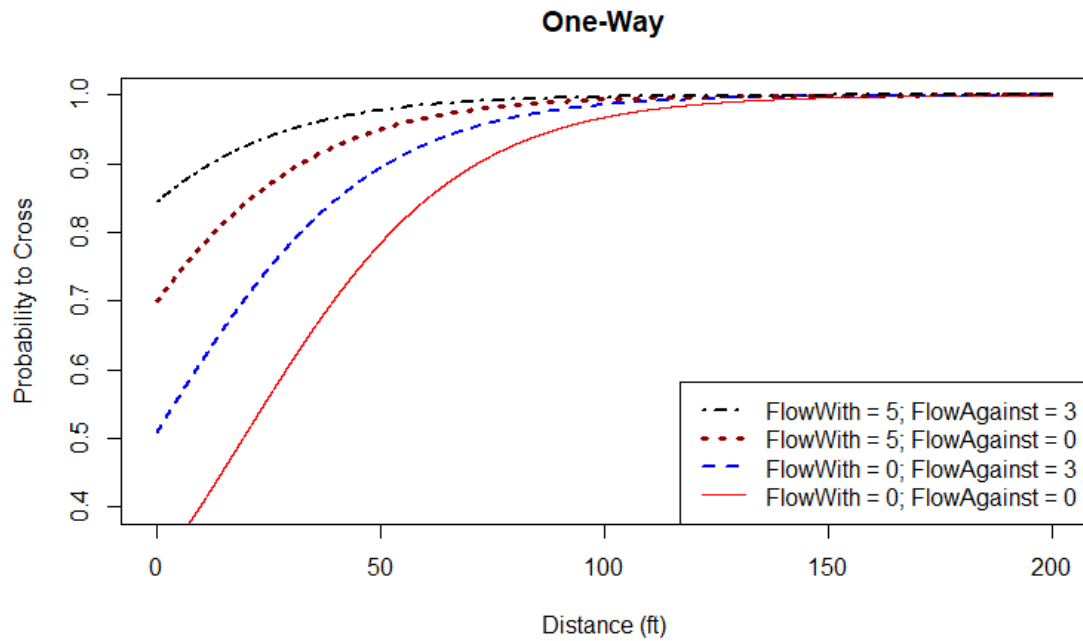
Figure 6 Effects of FlowWith and FlowAgainst on Pedestrian Crossing Behavior

- if other pedestrians are using the crosswalk (FlowWith and FlowAgainst). The effects are shown in Figure 6.

Some of the findings in this study are similar to individual findings found in other studies; other findings in this study represent new contributions. All of the findings are plausible. This gives the model credibility, subject to a more careful look at the model's values, after examining related models.

Based on the data for two-way University Street, the binary logistic regression model (Table 3) suggests that, in a pedestrian-motorist interaction, pedestrians are more likely to cross …

- if pedestrian acts without hesitation (Hesitation=0).
- if a driver decelerates (SlowsDown=1). See Figure 7.
- if a vehicle approaches at a lower speed (ApprSpeed).
- if the distance between a pedestrian and motorist is greater (Distance). See Figure 8.
- if the other pedestrians are crossing in the same direction as the pedestrian being observed (FlowWith).
- if an interacted vehicle is in the near lane (NoF=0).
- if there is a pedestrian waiting on the opposite side of the street (DiffDirec) See Figure 9.
- if a pedestrian is grouped with other people (GroupSize) See Figure 9.
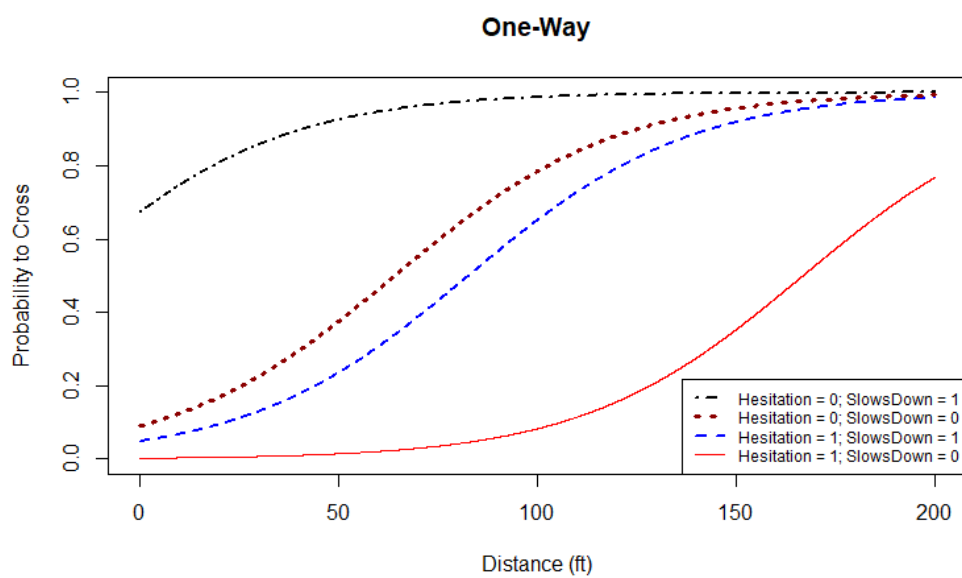
Figure 7 Effects of Hesitation and SlowsDown on Pedestrian Crossing Behavior
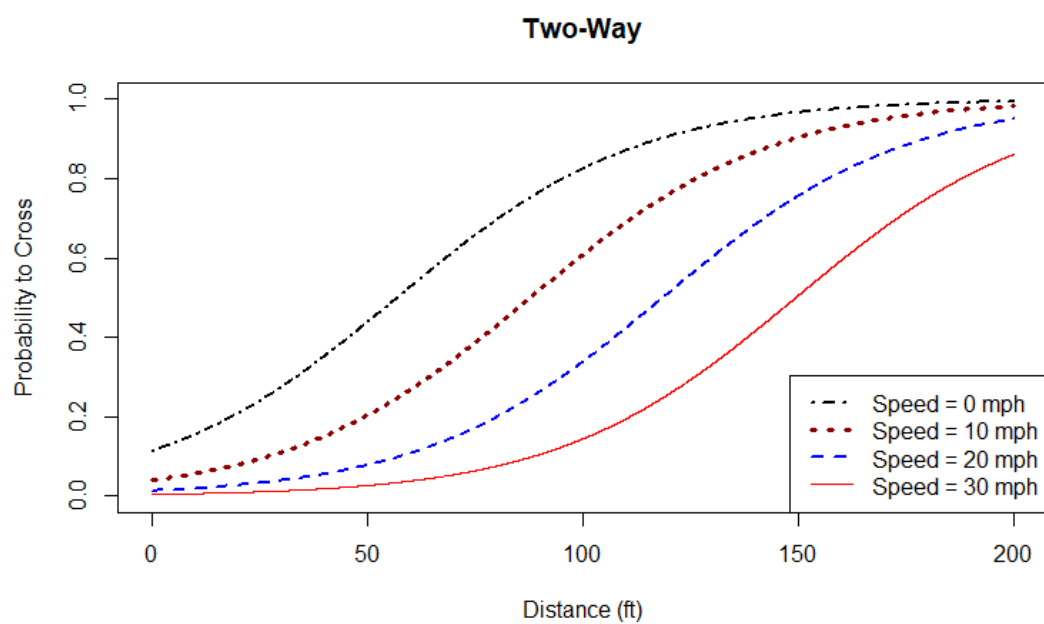
Figure 8 Effects of Distance and Speed on Pedestrian Crossing Behavior
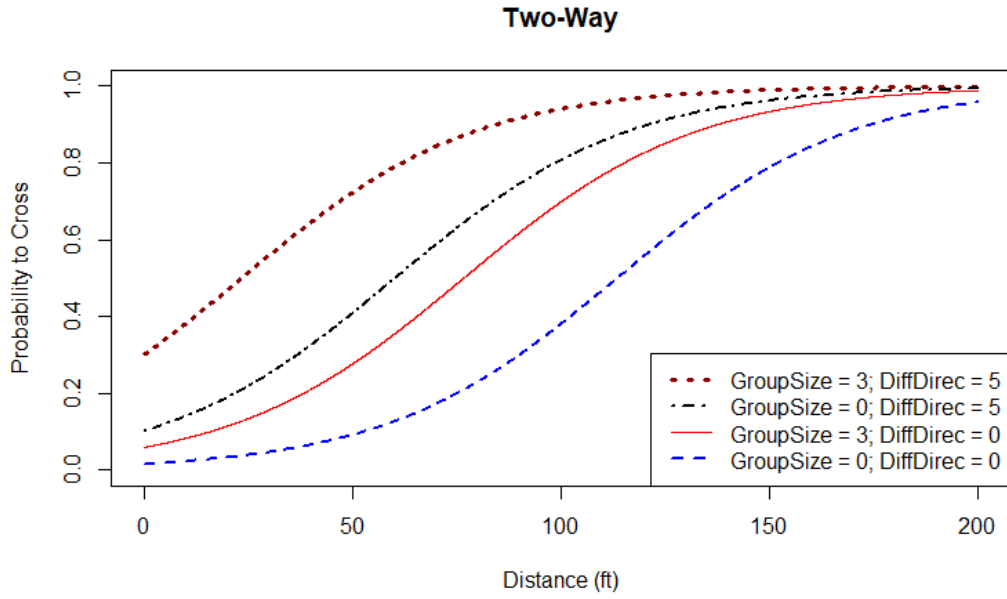
Figure 9 Effects of DiffDirec and GroupSize Variables on Pedestrian Crossing Behavior

After University Street was converted to two-way operation, fewer parameters were statistically significant. The first four variables listed above for two-way operation were also significant for the one-way data. They can be interpreted in the same way as in one-way case.

In an attempt to analyze the effect of the change on University Street from one-way to two-way, the data were combined and a group dummy variable (one-way case = 0; two-way case = 1) was introduced. The binary logit regression was run on the combined data, using an interaction term that pairs group dummy variables with independent variables, as shown in Equation 7 (Cross Validate, 2018):

$$\text{logit}(Y=1|X) = \beta_0 + \beta_1 * \text{indepVar} + \beta_2 * \text{groupDummy} * \text{indepVar} + \beta_3 * \text{groupDummy} \tag{7}$$

where

- β_1 is the vector of coefficients for the one-way case.
- β_2 is the vector that measures the difference in the coefficients between the two separate models (one-way and two-way).
- β_3 shows the differences in intercepts between the separate models.

The *second* rightmost column in Table 3 shows the results for the coefficients of interest in the binary logistic regression model and represents elements of β_2 in Equation 7. Consequently, one can test whether each element in β_2 is significant, to show if the change from one-way to two-way has caused a significant change in the pedestrian crossing models. An example is that the DiffDirec estimated coefficients in one-way case is 0.133 with 95% confidence interval [-0.044450, 0.311147], while the number is 0.384** with 95% confidence interval [0.1317962, 0.6501222]. The 95% confidence interval for the estimated coefficients are overlapped so that in the second right most column in Table 3 is not reported as statistical significance. Other results show that:

- The coefficient for pedestrian arriving group size (GroupSize) has changed significantly. In the one-way case, it was negative and not statistically significant; for two-way case, it is positive and statistically significant (0.424*** in Table 3).

- There is a significant change in the effect of the FlowWith factor (-0.573*** in Table 3). In the one-way case, if there are already pedestrians in the crosswalk, pedestrians are more likely to cross. However, on a two-way street, the FlowWith factor had no significant impact on pedestrian crossing behavior.

- The impact of the presence of an adjacent vehicle is significant in the one-way case, but disappears in the two-way case, because the coefficient changes significantly (-0.685** in Table

4.3.3 *Motorist Model*

By state law, at a semi-controlled crossing, a motorist is supposed to yield to a pedestrian who is "within the crosswalk". In both the one-way and two-way Pedestrian Model, a major factor in a pedestrian's decision to cross was the deceleration of the vehicle(s) during an interaction. Whether a vehicle slows down is a vital factor to study in the negotiations between pedestrians and motorists. In this section, we focus on the parameters that may have significant impacts on drivers' slowing down behavior. By assigning levels of deceleration intensity to each motorist (Level 1 – Motorist does not slow down; Level 2 -- Motorist slows down but does not stop; and Level 3 -- Motorist

stops), an ordered logistic regression can be used to analyze the probability of a particular response level for a series of given parameters (Williams, 2016):

$$\text{Prob}(Y_i > j) = O(X\beta) = \frac{\exp(\alpha_j + X_i\beta_j)}{1 + \exp(\alpha_j + X_i\beta_j)}, \quad j = 1, 2, \dots M - 1 \tag{8}$$

where Y_i is the response variable (motorist action in this application), M denotes the number of ordinal dependent variables ($M = 3$ levels here), and β_j are the same for all categories, but α_j are not necessarily the same among categories. A critical assumption associated with the ordered logit model is the *proportional odds assumption*, which imposes the restriction that regression parameters (except constants) are the same across different dependent levels. However, for deceleration intensity, it is not clear whether distances between adjacent deceleration levels are equal. Considering that the *proportional odds assumption* may be violated by only a subset of variables, a generalized ordered logistic regression (GOLR) partial proportional odds model was adopted. Compared with the ordered logistic regression model, the GOLR model relaxes the proportional odds assumptions for some explanatory variables, while maintaining them for the variables that satisfied the proportional odds assumption (Williams, 2016). The model could be further revised using Equation 9 as:

$$\text{Prob}(Y_i > j) = g(X\beta) = \frac{\exp(\alpha_j + X_{1i}\beta_1 + X_{2i}\beta_{2j})}{1 + \exp(\alpha_j + X_{1i}\beta_1 + X_{2i}\beta_{2j})}, \quad j = 1, 2, \dots M - 1 \tag{9}$$

where β_1 is the vector of variables that are subject to the proportional odds assumption. Explanatory variables X_{2i} that do not satisfy this assumption need the addition of coefficients β_{21} to relax the proportional odds assumption. The results are shown in Table 4.

Table 4 Generalized Ordered Logistic Regression Model Results

Variables	One-way Coefficient Between		Two-way Coefficient Between		Tests of equality of Coefficients (p-value)	
	1 and 2	2 and 3	1 and 2	2 and 3	1 and 2	2 and 3
Sex	—	—	—	—	—	—
Distance	—	—	0.007***	—	0.0035**	—
FlowOn	0.137***	0.137***	0.101**	0.101**	—	—
PedWait	—	—	—	—	—	—
AppSpeed	-0.304***	-1.248***	-0.302***	-0.933***	—	—
CloseFollow	-0.355*	-0.355*	—	—	—	—
AdjVeh	-0.569***	-0.569***	—	—	0.0001***	0.0001***
Hesitation	-0.384*	-0.384*	-1.518***	—	0.0000***	0.0098**
Distraction	—	—	0.451*	0.451*	—	—
NoF	—	—	-0.3*	-0.3*	0.0093**	—
Cutoffs	4.44***	4.046***	4.114***	2.232***	—	—
Log Likelihood	-784.588		-754.676		—	—
Pseudo R²	0.5855		0.5626		—	
Observations	1759		1574		—	
* p<0.05; ** p<0.01; *** p<0.001; "— = not applicable"						

In Table 4, the first column of coefficients can be interpreted in terms of Equation 9, where the dependent variable is recoded as Prob $(Y_i>1)$, which is equivalent to the probability that motorist deceleration Levels 2 and 3 occur, i.e., j>1. The second column of coefficients can be interpreted in terms of Equation 9, where the dependent variable is recoded as Prob $(Y_i>2)$, which is equivalent to the probability that motorist deceleration Level 3 occurs (Williams, 2016). This model has been widely used in traffic crash analysis to analyze the relationship between the severity of injury and associated variables. Furthermore, in some literature (e.g., Wang et al., 2008; Michalaki et al., 2015), marginal effects were used to measure the effect that a change in an explanatory variable has on the predicted probability of a specific category. The marginal effects are shown in Table 5.

Table 5 Marginal Effects

Variables	One-way			Two-way		
	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
Sex	—	—	—	—	—	—
Distance	—	—	—	-0.0007***	0.001***	-
FlowOn	-0.0176***	0.0146***	0.003***	-0.0099**	0.0041**	0.0058**
PedWait	—	—	—	—	—	—
AppSpeed	0.0393***	-0.0115***	-0.0278***	0.0296***	0.0237***	-0.0533***
CloseFollow	0.0459*	-0.038*	-0.0079*	—	—	—
AdjVeh	0.0735***	-0.0609***	-0.0127***	—	—	—
Hesitation	0.0496*	-0.0411*	-0.0085*	0.1488***	-0.1602***	—
Distraction	—	—	—	-0.0442*	0.0184*	0.0258*
NoF	—	—	—	0.0294*	-0.0123*	-0.0171*
* p<0.05; ** p<0.01; *** p<0.001; "— = not applicable"						

4.3.4 *Motorist Model Discussion*

For University Street in its one-way operation, the generalized ordered logistic regression model suggests that, in a pedestrian-motorist interaction:

- A motorist is more likely to slow down if the driver's approach speed is lower. In the GOLR model, the coefficients for both Level 1 and Levels 2 and 3 are negative (-0.304 and -1.248) and marginal effects also suggest that a higher approach speed leads to a higher likelihood of non-slowing down behavior (0.0393*** in Table 5).

- A driver is more likely to slow down if there are no other vehicles present in the adjacent lane (AdjVeh Coef. = -0.569; p-value = 0.000 in Table 4). Furthermore, the marginal effect of this factor for Not Slowing Down (Level 1) is 0.0735*** in Table 5, which is a significant increase. This means that, for the one-way case, the behavior of a vehicle in the adjacent lane will cause significant effects on a motorist's decision.

- A motorist is less likely to decelerate if there is a close follower behind him/her (CloseFollow Coef. = -0.355* in Table 4). The marginal effects reflect that a driver will be more aggressive without slowing down (0.0459* in Table 5) and less likely to brake or stop (-0.038* and -0.0079* in Table 5) if another driver closely follows him/her.

- If a pedestrian slows down or stops at the curb during an interaction (Hesitation=1, Coef. = -0.384***), the marginal effects indicate that a driver will be more likely to continue without slowing down (+0.0496*in Table 5) and less likely to slow down (-0.0411* in Table 5).

- if an interaction occurs when there is a greater number of pedestrians already in the crosswalk (FlowOn), a driver is more likely to slow down (Coef. = 0.137*** in Table 4), because in Table 5, marginal effects indicate a positive impact on slowing down (0.0146**) and on stopping (0.003**), with negative impacts on non-deceleration (-0.0176***). It is intuitive that more pedestrians already in the crosswalk will lead to drivers slowing down.

For two-way University Street, the model shows the same effects as in the one-way case for variables FlowOn, AppSpeed, and Hesitation. Some other variables became significant.

- For the NoF variables in Table 4, their coefficients are -0.3*, which means that the probability of a motorist slowing down is lower when the pedestrian is on the far curb (NoF=1), not the near curb, which is proved by the marginal effects (0.0294* for Level 1; -0.0123* for Level 2; and -0.0171* for Level 3 in Table 5).

- Distance (Coef. Distance = 0.007*) has significant impact on the driver's decision to slow down. With the increase of distance, the marginal effects for this parameter have a positive influence on a driver slowing down (0.001***) while having a negative effect on not-slowing down behavior (-0.0007***).

- For the distraction variable (Distraction), a driver is more likely to decelerate if a pedestrian in the interaction uses a cellphone or talks to others (Distraction, Coef. = 0.451* in Table 4). In Table 5, the marginal effects for this parameter have a positive influence on a driver slowing down and stopping (0.0184* and 0.0258*), while having a negative effect on not-slowing down behavior (-0.0442*).

4.3.5   *Summary*

As with the pedestrian model, we test whether the coefficients in the one-way and two-way cases are equal. The two rightmost columns in Table 4 show the p values for the hypothesis tests, which indicated that variables Hesitation, AdjVeh, NoF, and Distance change significantly. Meanwhile, compared with one-way street operation, some variables (CloseFollow and AdjVeh) were no longer significant.   One interesting result is that, for one-way operation, driver behavior is influenced greatly by both pedestrian characteristics (Hesitation and FlowOn) and vehicle dynamics (AppSpeed, CloseFollow, and AdjVeh). For two-way operation, driver behavior is significantly determined by more pedestrian characteristics factors (Hesitation, Distraction, and FlowOn), fewer vehicle dynamics factors (AppSpeed) and more environmental characteristics factors (Distance and NoF), when an interaction occurs.

There are limitations in models. Endogeneity is introduced when we use SlowDown parameter in the pedestrian model. Moreover, we explored pedestrian behavior and motorist behavior separately rather than study the complex netiogiations between pedestrians and motorists, which required a more complicated framework.

CHAPTER 5.   PEDESTRIAN WAIT TIME – SURVIVAL ANALYSIS

5.1   Accelerated Failure Time (AFT) Model

Based on the definitions of interaction and critical event in Chapter 3, we aimed to investigate pedestrian wait time when pedestrian-motorist interactions happen. To investigate the impacts of critical vehicles on pedestrian behavior, we coded information for the critical vehicle: vehicle type, driving in the near or far lane, distance to pedestrian, and approach speed at the time when pedestrian reaches the curb.

Table 6 Explanatory Variables and Descriptive Statistics

Variable	Description
Explanatory Variable	
Pedestrian Wait Time	Duration (in seconds) between the time a pedestrian reached the curb area and the time the pedestrian started crossing (One-way: mean=2.67s, sd=2.61s; Two-way: mean=3.02s, sd=3.08s)
Independent Variable	
Pedestrian characteristics	
Sex	1 if the pedestrian is Male (51.1%), 0 if the pedestrian is Female (48.9%).
Estimated Age Category	Young for pedestrians that appear to be younger than 30 years old (77%); Mid-age for pedestrians between 30 and 50 years old (21%); Elderly for pedestrians older than 50 years old (2%).
Cellphone Indicator	1 if pedestrian is using cellphone when waiting at the curb (9%), 0 otherwise (91%).
Talking Indicator	1 if pedestrian is talking to others when waiting at the curb (8.4%), 0 otherwise (91.6%).

Table 6 continued

Traffic Condition	
Vehicle Arrival Rate	Number of vehicles driving past the crosswalk per minute (mean=8.72 veh/min, sd=3.95 veh/min)
Near Side Indicator	1 if the critical vehicle is in the near lane (53%), 0 if the critical vehicle is in the far lane (47%).
Bus/Truck Indicator	1 if the critical vehicle is bus or large truck (16.7%), 0 otherwise (83.3%).
Veh-to-Ped Distance	Distance (in ft) between pedestrian and the first approaching vehicle (mean=61.8ft, sd=51.7ft).
Approaching Speed	Speed (ft/s) of the first approaching vehicle when the pedestrian arrives at the curb (mean=12.4ft/s, sd=11.2ft/s).
Adjacent Vehicle Indicator	1 if there is one or more vehicles presenting in the adjacent lane within the area of influence when the motorist begins to interact with a pedestrian (32.3%), 0 otherwise (67.7%).
Vehicle Close Follower Indicator	1 if there is at least one vehicle closely following the critical vehicle (31.9%), 0 otherwise (68.1%).
Other Pedestrians	
Group Size	Number of people in the pedestrian group (45.7% pedestrians came alone, 54.3% came in a group; mean=2.31, sd=2.03)
Nr Ped Waiting	Number of pedestrians waiting at the curb as a pedestrian arrives at the curb (52.8% cases of no pedestrians waiting, 47.2% cases of at least one pedestrian was waiting at the curb; mean=0.84, sd=1.21)
Nr Ped Crossing	Number of pedestrians crossing the street as a pedestrian arrives at the curb (52.8% cases of no pedestrians crossing, 47.2% cases of at least one pedestrian crossing in crosswalk; mean=1.27, sd=1.99)

Note: mean=average value; sd=standard deviation

5.1.1 *Log-Linear Model*

First, a multivariate model was developed to analyze the relationship between pedestrian wait time and explanatory variables. At semi-controlled locations, pedestrian wait time is always non-

negative. As a result, we transform the pedestrian wait time into log format. The log-linear model is:

$$\log(y_i) = \beta_0 + \sum_{\forall k} \beta_k * x_{i,k} + \varepsilon_i \qquad (10)$$

where y_i is the wait time and $x_{i,k}$ is the k-th variable for pedestrian i; β_0 is the estimated constant; β_k is the coefficient estimated for the k-th variable, and ε_i is the error term.

5.1.2 *AFT Model Structure*

Hazard-based duration models are widely used with duration-related datasets. Studies related to accident analysis (Nam and Mannering, 2000), travel activity behavior (Yang et al., 2015), and queueing theory (Paselk and Mannering, 1993) usually applied hazard-based duration models due to their flexibility with time-dependent dataset.

The duration analysis primarily focused on the length of time that elapsed from the starting state of an event (in this project, we call the event an interaction) until the ending state of an event. Duration analysis is also interested in the likelihood that an event would end in the next short period of time, given its current state (Nam and Mannering, 2000). The hazard function at time (t) can be expressed as a density function $f(t)$, and its cumulative distribution function $F(t)$.

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} = \lim_{dt \to 0} \frac{\Pr(t \leq T \leq t + dt)}{dt * S(t)} = -\frac{S'(t)}{S(t)} \qquad (11)$$

To better understand pedestrian waiting duration, some external variables, such as pedestrian features and vehicle dynamics, should be introduced as covariates into the hazard function. We also included the effects of covariates by multiplying the covariates $g(X; \beta)$ with the hazard function $h(t)$:

$$h(t; X) = h(t)g(X; \beta) \qquad (12)$$

in Equation (12). The term $g(X; \beta)$ can be expressed as any relationship between pedestrian waiting duration and covariates *X*.

In this section, we discuss the fully parametric approach to investigate pedestrian waiting duration at semi-controlled places. Different from non-parametric risk (hazard) functions, parametric functions need to be given a probability distribution $h(t)$. Typical probability distributions such as exponential, Weibull, log-logistic, log-normal, or Gompertz were investigated as alternatives in fully-parametric hazard-based functions. The exponential distribution assumes that the hazard function is constant over time. Weibull or Gompertz distributions both assume that the hazard function is decreasing or increasing overtime non-monotonically. The log-logistic distribution is a widely used probability distribution in hazard-based duration analysis, because of its flexibility in dealing with non-monotonical relationships. These alternatives were tested in this study, and their model performances were compared using the Akaike information criterion (AIC). The AIC is an estimator of the relative quality of statistical models, which provides a means of model selection. The AIC can be calculated using Equation (13), which represents a trade-off between model fit and model complexity (Akaine, 1987).

$$AIC = -2*\log(likelihood) + 2(p+k) \tag{13}$$

where p is the number of parameters, k=1 for the exponential model, and k=2 for the Weibull, log-logistic, and log-normal models (Klein et al., 1997). A lower AIC value indicates a higher likelihood. In this part, the log-logistic model was found to provide better model performance than other distributions in modeling pedestrian wait time. In addition, Figure 10 shows the close match between the observed survival curve and the log-logistic distribution fitted with the AFT model.
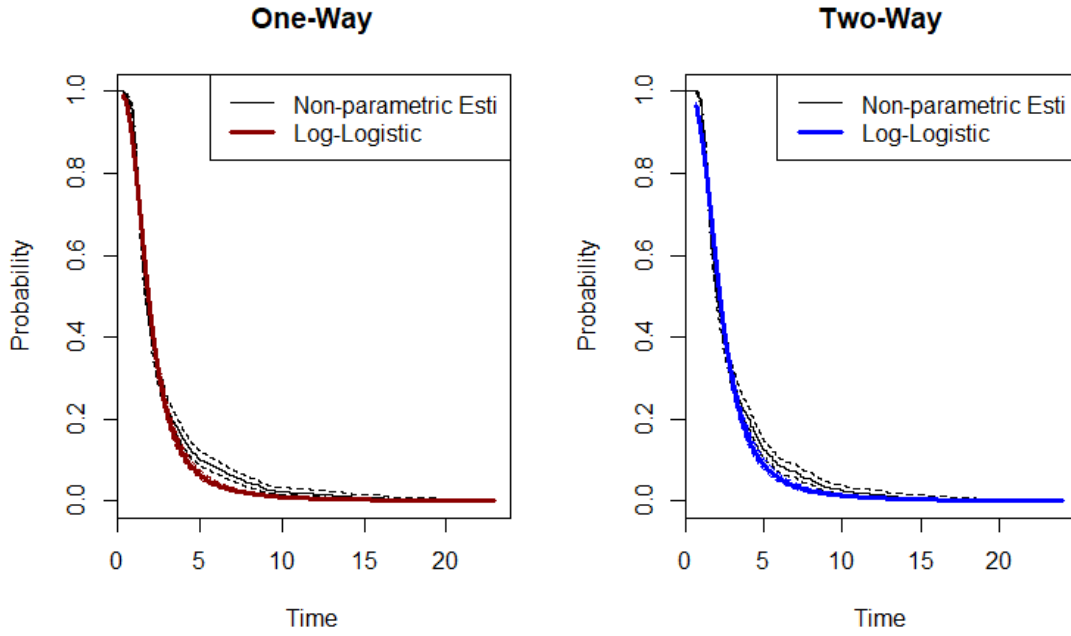
Figure 10 Non-Parametric Survival Estimations and Fitted Distributions

The hazard function h(t), survival function S(t), and survival (wait) time T for a log-logistic survival model are shown in Equations (14) to (16).

$$h(t) = \frac{\lambda p t^{p-1}}{1 + \lambda t^p} \tag{14}$$

$$S(t) = \frac{1}{1 + \lambda t^p} \tag{15}$$

$$T = \left(\frac{1}{S(t)} - 1\right)^{\frac{1}{p}} \frac{1}{\lambda^{1/p}} = \left(\frac{1}{S(t)} - 1\right)^{\frac{1}{p}} \exp(\beta_0 + \beta X) \tag{16}$$

5.2   Model Discussions

The estimation results of the log-linear regression model and the survival models are presented in Table 7. The estimation results of the two models are mostly consistent with each other in terms of variable significance and sign consistency. The interpretation of the estimation results of the log-linear regression model is straightforward through the marginal effect: an influential variable with a positive sign indicates an increase on the wait time, meaning a one unit increase of an

influential variable with coefficient β_i leads to a β_i*100 percent increase in pedestrian delay. For example, the coefficient estimated for Male is -0.0673, meaning that the wait duration of a male pedestrian is 6.73% shorter than the duration of a female pedestrian, when all other factors are the same.

Table 7 Log-Linear Regression

Log-Linear Model	One-way			Two-way		
Number of pedestrians	1132 obs			927 obs		
Explanatory Variable	β	t	P-value	β	t	P-value
(Intercept)	-0.3294	0.149	0.027*	0.69	0.1692	<0.001***
Sex	-	-	-	-0.0673	0.0383	0.079.
Age	0.1844	0.0413	<0.001***	0.0782	0.037	0.035*
Distraction	-	-	-	0.1409	0.0529	0.008**
Group Size	-0.0228	0.0958	0.017*	-0.0495	0.0145	0.002**
Near or Far Side	-	-	-	-0.0974	0.0389	0.012*
Vehicle Type	0.1487	0.0413	<0.001***	-	-	-
Distance	0.0017	0.0011	0.1256	-	-	-
(Distance)^2	-1.87E-05	5.76E-06	0.0012**	-	-	-
Approaching Speed	0.0226	0.0055	<0.001***	0.0106	0.0056	0.056.
(Approaching Speed)^2	-0.00047	0.0002	0.079.	-	-	-
Adjacent Vehicle	0.2845	0.0416	<0.001***	0.2345	0.0427	<0.001***
Close Follower	0.2751	0.0419	<0.001***	0.1895	0.0149	<0.001***
Nr Ped. Waiting	0.0496	0.0113	<0.001***	-	-	-
Nr Ped. Crossing	-0.0542	0.0071	<0.001***	-0.0541	0.0085	<0.001***
R^2	0.2418			0.1635		
F-statistic:	25.45			12.73		
Degree of freedom	11			10		
Significance	< 2.2e-16			< 2.2e-16		

Signif. codes: '***': 0.001; '**': 0.01; '*': 0.05; '.': 0.1

Table 8 AFT Model Estimated Results

AFT Model	One-way			Two-way		
Number of pedestrians	1132 obs			927 obs		
Explanatory Variable	β	Std.Err	P-value	β	Std.Err	P-value
(Intercept)	-0.294	0.174	0.09.	0.528	0.192	0.006**
Sex	-	-	-	-0.0663	0.039	0.089.
Age	0.199	0.0448	<0.001***	0.09	0.0413	0.029*
Distraction	-	-	-	0.104	0.0563	0.065.
Group Size	-	-	-	-0.0394	0.0148	0.008***
Near Side or Far Side	-	-	-	-0.104	0.0448	0.021*
Vehicle Type	0.111	0.0611	0.07.	-	-	-
Distance	0.00131	0.0012	0.287	-	-	-
(Distance)^2	-1.67E-05	5.88E-06	0.0046*	-	-	-
Approaching Speed	0.024	0.00681	<0.001***	0.0142	0.00637	0.025*
(Approaching Speed)^2	-0.000363	0.00022	0.1000.	-	-	-
Adjacent Vehicle	0.243	0.0602	<0.001***	0.227	0.0504	<0.001***
Close Follower Indicator	0.214	0.059	<0.001***	0.184	0.053	<0.001***
Nr Ped. Waiting	0.052	0.0129	<0.001***	-	-	-
Nr Ped. Crossing	-0.0494	0.00765	<0.001***	-0.05	0.00871	<0.001***
Log Likelihood	-1766.8			-1581.6		
Degree of freedom	11			10		
Significance	< 2.2e-16			< 2.2e-16		

Signif. codes: '***': 0.001; '**': 0.01; '*': 0.05; '.': 0.1

The log-linear and AFT models showed similar results. In the AFT framework, the exponential of the estimated coefficient is called the accelerated factor (AF), which measures, for each variable, the increased pedestrian delay associated with an increase in the value of that variable. For example, the exponential of a positive coefficient, such as Age Indicator in the one-way model, is AF = exp (0.199) = 1.22, which means that with the increase of Age indicator by 1, it shows an increase of 22% probability of waiting. Conversely, the exponential of a negative coefficient, such as Group Size, is exp (-0.104) = 0.901. The interpretation is that a pedestrian is likely to wait about 0.901 times as long (9.9% shorter) when the group size increases by 1 person, while keeping all the other variables unchanged. Generally, a coefficient greater than zero (or, equivalently, an exponent parameter greater than 1.0) indicates that an increase in the explanatory variable results in increased pedestrian delay, and vice versa.

5.2.1 *Before and After Studies*

5.2.1.1 Distance and Speed

In both the regression model and the AFT duration model, the squared term of the vehicle-to-pedestrian distance and the squared term of vehicle speed were highly significant in the one-way case, indicating that there exists a non-monotonic relationship between pedestrian delay and the two variables. Figure 11 shows that pedestrian wait time is greatest when the interacted vehicle is 39 - 46 ft from the crosswalk, approaching at an average speed. The pedestrian wait time is smaller when the vehicle is closer than 39 - 46 ft, because the pedestrian is content to let the vehicle pass before crossing the street with increases as the speed increases and as the distance decreases only up to certain thresholds, after which the relationship becomes the opposite. Such a non-monotonic relationship (See Figure 11) has not been identified in past studies.

However, only speed term showed significant impact on pedestrian waiting time in two-way operation.
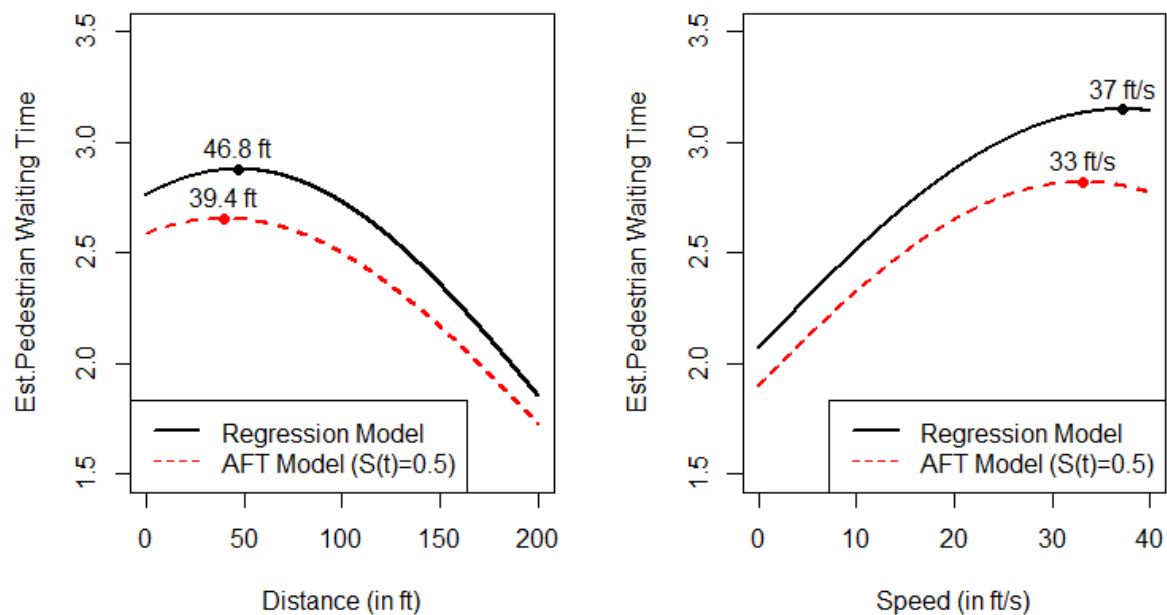

Figure 11 Relationships between Pedestrian Delay and Distance & Speed (One-Way Case)

5.2.1.2   Pedestrian Characteristics

Pedestrian Characteristics (Sex, Distraction) have significant impacts on pedestrian waiting behavior in the two-way case. *Male* showed significant lower waiting durations than females (See Figure 12). Moreover, *distraction* (Talking and Cellphone Using) will result in a longer waiting time (See Figure 12).
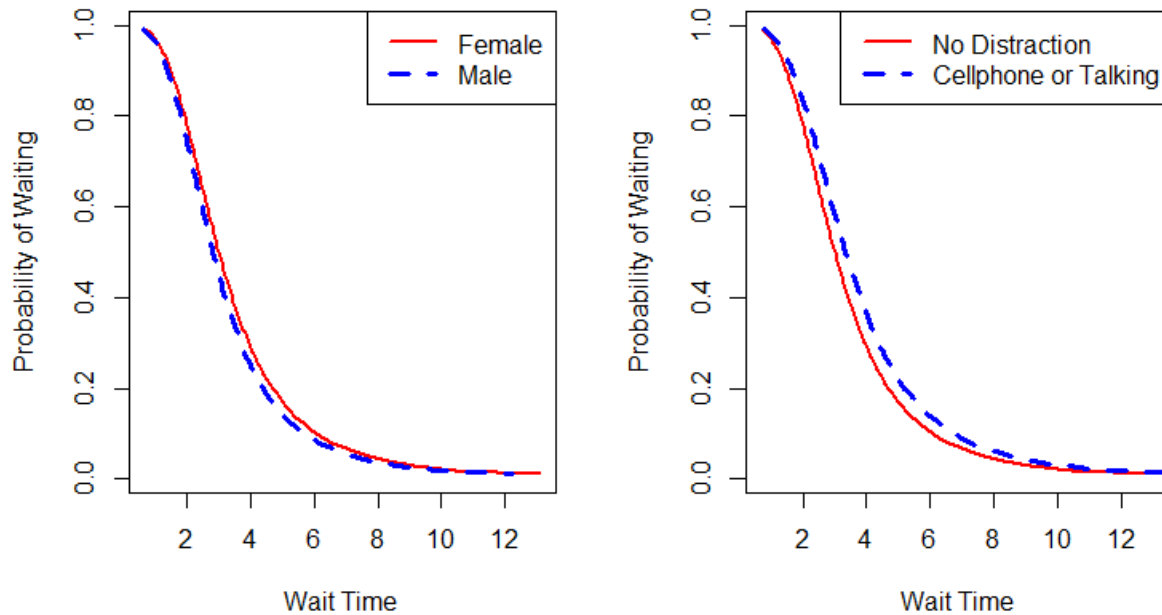

Figure 12 Effects of Pedestrian Characteristics (Two-Way)

5.2.1.3   Environmental Factors

The variables *near side* and *far side* showed significant impacts on pedestrian waiting durations in the two-way case. If the interacted vehicle is in far lane, the model indicates that pedestrians will have a lower wait time. Furthermore, with the increase in group size, the subject pedestrian will wait less (See Figure 13).
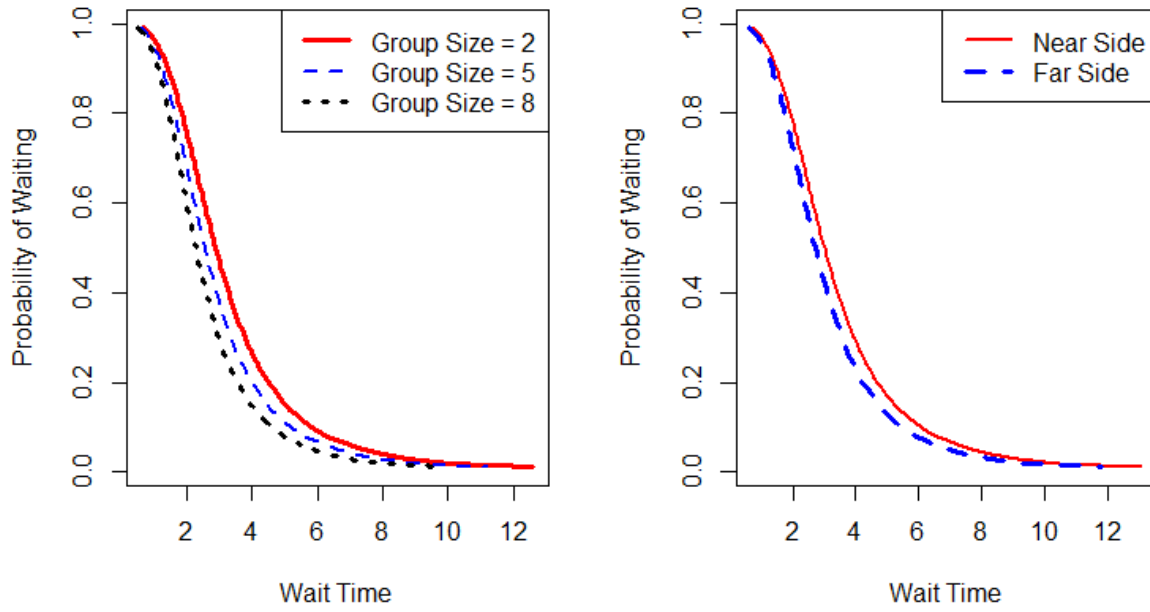
Figure 13 Effects of Environmental Factors (Two-Way)

5.3   Model Performance

5.3.1   *Mean Absolute Percentage Error (MAPE)*

To compare the model performance between the regression models and the AFT duration models, the mean absolute percentage error (MAPE) is used. The MAPE is a summary measure widely used for evaluating the accuracy of prediction results. It can be calculated using Equation (17).

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{O_i - P_i}{O_i} \tag{17}$$

where O_i is the observed waiting duration for the i-th pedestrian, P_i is the predicted wait duration for the i-th pedestrian, and n is the number of pedestrians included in the model.

A lower MAPE value indicates a higher accuracy of the prediction model. In this study, the MAPE value was calculated as 47.3% for the log-linear model and 37.6% for the log-logistic AFT duration model in the one-way case; 46.6% for the log-linear model and 36.4% for the log-logistic AFT

duration model in the two-way case. In safety studies related to human factors, the MAPE value range from 21% to 50% is at a reasonably accuracy level (Chung, 2010).

5.3.2 *Error Tolerance*

Another measure of model prediction accuracy used in duration modeling is related to a certain tolerance of the actual durations (Chung, 2010; Yang et al., 2015). In this part, we defined the percentage error as the percentage difference between the observed and predicted value. The prediction accuracy under certain error tolerance is calculated as the ratio of the predicted durations with percent errors smaller than the given error tolerance to the total number of prediction points. Figure 14 presents the prediction accuracy under error tolerance from 0% to 100% for the two estimation models and in the one-way and two-way cases. The plots show that the log-logistic AFT duration model outperformed the log-linear model at each tolerance level in term of the prediction accuracy for both the one-way and two-way cases.
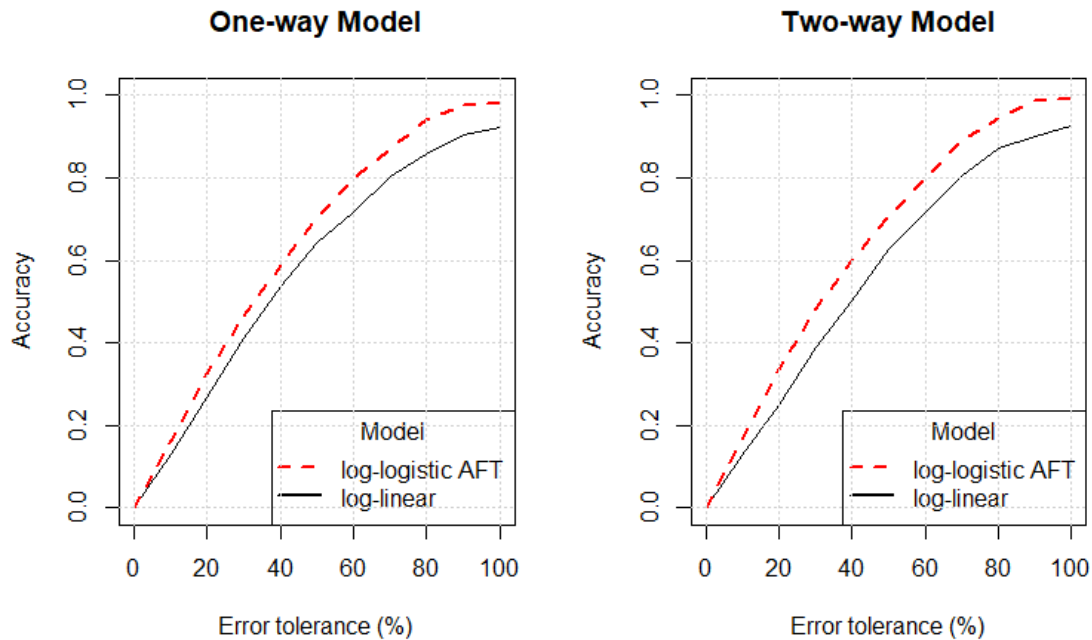

Figure 14 Prediction Accuracy under Different Error Tolerance

5.4   Summary

In this chapter, we use the first event analysis to estimate the pedestrian waiting behavior. The occurrence of the first event (interaction) is considered to have the most critical impact on pedestrian wait durations. Log-linear model and AFT model are utilized to investigate the effects of covariates on pedestrian waiting behavior. Based on the results, parameters distance and speed are shown to have non-monotonic relationship on pedestrian wait durations in one-way case. The peak values for distance (39.4 ft in AFT model and 46.8 ft in Log-Linear) and speed (33 ft/s in AFT model and 37 ft/s in Log-Linear) are revealed to cause the longest pedestrian delay at semi-controlled crosswalks.

As for the two-way street, a pedestrian is less likely to wait more when an interacted vehicle is on the far side. Based on our observations, compared with one-way operation, pedestrians have different crossing strategies because they have to look for vehicles from different directions. If the object vehicle is in the far lane, pedestrian is more assertive to cross even if the far lane vehicle does not yield. This is because there's one lane buffer for pedestrians and they may feel comparatively safe towards the far line interacted vehicles.

CHAPTER 6.   PEDESTRIAN WAIT TIME – MARKOVIAN APPROACH

6.1   Model Formulation

Consider a Markov renewal process (J_n, T_n), where $T_0 < T_1 < \ldots < T_n < \infty$ are the successive times of entry to states $J_0, J_1, \ldots, J_n$. If $S_n = T_n - T_{n-1}$ is the sojourn time (gap time or lag time), the Markov renewal kernel $Q_{hj}(d)$ is a cumulative distribution function of time:

$$Q_{hj}(d) = P(J_{n+1} = j, \ S_{n+1} \le d \mid J_0, J_1, \ldots, J_n = h, \ S_1, S_2, \ldots, S_n)$$
$$= P(J_{n+1} = j, \ S_{n+1} \le d \mid J_n = h) \tag{18}$$

$J_0, J_1, \ldots, J_n$ is an embedded homogeneous Markov chain taking values in a finite state space with transition probability:

$$p_{hj} = P(J_{n+1} = j \mid J_n = h) = \lim_{t \to \infty} Q_{hj}(t), \ n \in N \tag{19}$$

We define the distribution function of the sojourn time in state h by:

$$H_h = \sum_{j=1}^{s} Q_{hj}(t), \quad \forall t \in \square \tag{20}$$

The probability distribution function of sojourn time (gap time or lag time), through the transition probabilities of the embedded Markov chain in terms of conditional probability, is:

$$F_{hj}(d) = P(\ S_{n+1} \le d \mid J_{n+1} = j, J_n = h) = \frac{Q_{hj}(d)}{p_{hj}} \tag{21}$$

$F_{hj}(d)$ is a cumulative probability distribution and is called a sojourn time in state h if the next state will be j. Based on equation (21), we can write the probability density function as $f_{hj}(d)$. The hazard function α_{hj} of $F_{hj}(d)$ will be:

$$\alpha_{hj} = \lim_{\Delta d \to 0} \frac{\Pr(d < S_{n+1} \le d + \Delta d \mid J_{n+1} = j, J_n = h, S_{n+1} > d)}{\Delta d} = \frac{f_{hj}(d)}{S_{hj}(d)} = \frac{f_{hj}(d)}{1 - F_{hj}(d)} \tag{22}$$

6.1.1 *Distribution of Durations*

For the semi-Markov process, we need to first assume that sojourn time (gap time or lag time) belongs to a specific parametric distribution. The sojourn time, given any state *h* to state *j*, is modeled as a random variable from the best fitted distribution. The *SemiMarkov* package in R software (Listwon-Krol and Saint-Pierre, 2015) offers three distributions -- Exponential, Weibull and Exponential Weibull. Based on maximum likelihood estimation, the Weibull distribution (Weibull, 1951) was chosen to model the sojourn time from state *h* to state *j*. The Weibull distribution is defined as the probability density function:

$$f(x \mid \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1} e^{-\left(\frac{x}{\lambda} \right)^k} \tag{23}$$

According to Equations (22) and (23), the hazard ratio for the Weibull distribution is:

$$\alpha_{hj} = \frac{f_{hj}(d)}{1 - F_{hj}(d)} = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1} \tag{24}$$

Table 9 Weibull Distribution Duration Parameters

Duration Parameters in Weibull Distribution								
	One-Way				Two-Way			
Transition	λ		k		λ		k	
	Estim.	SE	Estim.	SE	Estim.	SE	Estim.	SE
1→2	2.864	0.12	1.364	0.06	3.021	0.09	1.785	0.07
1→3	2.014	0.04	1.843	0.04	2.229	0.06	1.794	0.05
2→3	1.881	0.11	1.329	0.07	1.67	0.11	1.075	0.05

Table 10 Wald Test of Weibull Distribution

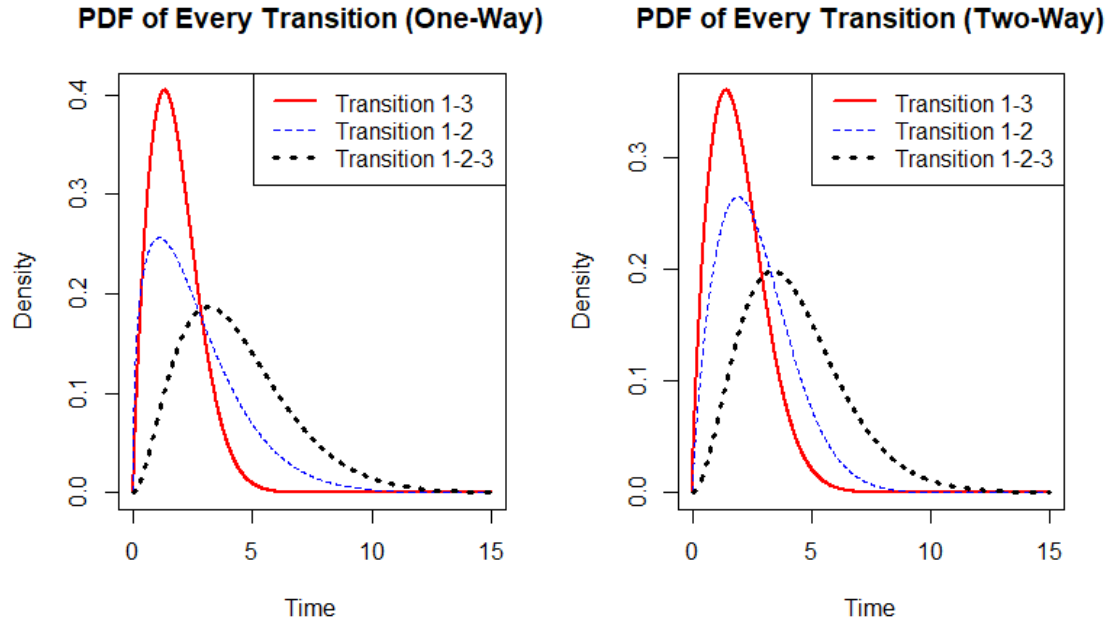| One-Way			Two-Way		
Transition	Wald Test	P-value	Transition	Wald Test	P-value
1→2	43.53	<0.0001	1→2	128.59	<0.0001
1→3	357.06	<0.0001	1→3	259.26	<0.0001
2→3	19.84	<0.0001	2→3	1.89	0.1692

Figure 15 Density Functions Between Different Transitions.

Figure 15 shows the probability density functions. The Transition 1-2-3 (reject the lag, then accept the next gap) can be expressed as the convolution product $f_{1-2-3} = \int_0^x f_{1-2}(u)f_{2-3}(x-u)du$, where x = the total wait time during the Transition 1-2-3. The probability density function $f_{1-2}(u)$ permits the calculation of the probability that the Transition from state 1 to state 2 takes place in μ time units. The probability density function $f_{2-3}(x-u)$ leads to the probability that the transition from state 2 to state 3 takes place in the remaining x- μ time units. Table 11 shows the most likely wait times for each distribution.

Table 11 The Most Likely Wait Times

The most likely wait times	One-Way	Two-Way
Transition 1-2	1.087s	1.906s
Transition 1-3	1.317s	1.415s
Transition 1-2-3	3.15s	3.38s

On the one hand, Transition 1-2-3 includes recurrent events -- rejecting the first lag (the first vehicle does not yield), then accepting the following gap (second interacted vehicle yields). Although there is information about the two events, the survival models in Chapter 5 only used information about the first event (Transition 1-2). On the other hand, when a pedestrian experiences Transition 1-2, it means that he/she rejects a gap and his/her accepted wait time is greater. The semi-Markov model considers state 2 as a right-censored state and has the potential to estimate the pedestrian actual accepted wait time continuously through the Markovian renewal process until the last observation is observed.

6.1.2   *Parameterization*

In order to illustrate the influence of covariates associated with the semi-Markov process, the Cox proportional model (Cox, 1972) was used. Let Z_{hj} be a vector of explanatory variables related to the transition from *h* to *j* and $\beta_{h,j}$ be the vector of estimated regression parameters. By the Cox proportional model:

$$\alpha_{hj}(d \mid Z_{h,j}) = \alpha_{hj}(d)\exp(\beta_{h,j}{}^{T}Z_{h,j}) \tag{25}$$

According to Equation (25), we took advantage of transition-specific variables $Z_{h,j}$ defined in Table 6. Using the *Semi-Markov* package in R, significant factors were retained. The estimated regression coefficients are shown in Table 12.

Table 12 Parametric Effects of Multi-State Model

Coefficients	One-Way			Two-Way		
	1→2	1→3	2→3	1→2	1→3	2→3
Near or Far Side	-	-	-	-	-	-
Group Size	-	-	-	0.123.	0.078**	0.199*
Nr. Ped. Crossing	-0.319***	0.105***		-0.148***	0.082***	-
Close Follower Indicator	-0.459***	-	-	-0.184.	0.196*	
Adjacent Vehicle	-	-0.196*	0.450*	-	-	-
Sex	-	-	0.325*	-	-	0.353*
Age	-	-0.355***	0.312*	-	-0.371***	-
Nr. Ped. Waiting	-	-0.157***	-	0.134**	-	-
Hesitation	-1.364*	-1.287***	-1.316***	-0.351.	-1.165***	-0.877***
Vehicle Type	-0.173***	-	-	-	0.318*	-
Distance	-0.635***	0.303***	-	-0.456***	-	-
Approaching Speed	0.057***	-0.031***	-	0.034***	-0.011*	-
Distraction	-	-	-	-	-	-0.312.
Log-Likelihood	-1128.8095 vs -1318.376 (Null LL)			-1097.4645 vs -1191.7082 (Null LL)		
. p<0.1; * p<.05; ** p<.01; *** p<.001						

Table 12 illustrates the effects of covariates on the sojourn time in each transition. Positive coefficients denote the increasing risk or accelerating factors. while negative coefficients demonstrate decreasing risk. We will discuss the effects of significant variables in the following sections.

6.1.3   *Hazard of Semi-Markov Process*

The hazard rate of a semi-Markov process is defined as the probability of transition towards state j between the time d and $d+\Delta d$, given that the process is in state h for a duration d.

$$\gamma_{hj}(d) = \lim_{\Delta d \to 0} \frac{\Pr(J_{n+1} = j, d < S_{n+1} \leq d + \Delta d \mid J_n = h, S_{n+1} > d)}{\Delta d}$$

$$= \begin{cases} \dfrac{q_{hj}}{1 - H_h(d)} = \dfrac{p_{hj} f_{hj}(d)}{1 - H_h(d)} = \dfrac{p_{hj}(1 - F_{hj}(d))\alpha_{hj}}{1 - H_h(d)} & \text{if } p_{hj} > 0 \quad \text{and} \quad H_h(d) < 1 \\ 0 & \text{otherwise} \end{cases} \quad (26)$$

Note that Equation (24) and Equation (26) demonstrate two different hazards. To better understand the differences, we defined the hazard in Equation (24) as the hazard given transition from state h to state j. The hazard defined by Equation (26) is the hazard of the semi-Markov process, which represents the immediate probability of going to state j given state h in a small-time interval [d, d+Δd] (Dominicis and Manca, 1984). Therefore, for the state space $I = \{1, 2, 3\}$, we can use Equation (26) to calculate the "staying" probability for the case $h=j$:

$$p_{11}(d) = e^{-\int_0^d \gamma_{12}(\tau)+\gamma_{13}(\tau)\,d\tau}$$
$$p_{22}(d) = e^{-\int_0^d \gamma_{23}(\tau)\,d\tau} \tag{27}$$
$$p_{33}(d) = 1$$

Consequently, we can calculate the probabilities of each transition in a Markov chain as:

$$p_{1-2}(d) = \int_0^d p_{11}(\tau)\gamma_{1-2}(\tau)p_{22}(d-\tau)d\tau$$
$$p_{1-3}(d) = \int_0^d p_{11}(\tau)\gamma_{1-3}(\tau)p_{33}(d-\tau)d\tau$$
$$p_{2-3}(d) = \int_0^d p_{22}(\tau)\gamma_{2-3}(\tau)p_{33}(d-\tau)d\tau \tag{28}$$
$$p_{1-2-3}(d) = \int_0^d p_{11}(\tau)\gamma_{1-2}(\tau)p_{2-3}(d-\tau)d\tau$$

An interpretation will be needed to illustrate the transition $P_{1\text{-}2\text{-}3}(d)$ in Equation (28). The term $p_{11}(\tau)\gamma_{1-2}(\tau)$ denotes the transition from state 1 to 2 (pedestrian rejects the first lag, or the first vehicle doesn't yield) in τ duration time. $p_{2-3}(d-\tau)$ indicates the probability of transferring from state 2 to state 3 (pedestrian accepts the next gap) in the remaining time d-τ.

$P_{1\text{-}3}$ and $P_{1\text{-}2\text{-}3}$ in Equation (28) are the total waiting behavior of pedestrians in the curb area, because the total number of transitions 1-3 + 1-2-3 is 966 out of 1132 for the one-way case. Therefore, we can use the transition probability of the semi-Markov process $P_{1\text{-}3} + P_{1\text{-}2\text{-}3}$ to explain the variables in Table 12.

## 6.2   Model Discussions

For a one-way operation:

6.2.1 *Number of Pedestrians Impacts*

(A). Figure 16 (a) shows that the number of pedestrians already on the crosswalk will speed up the Transition 1-3 (-0.319*** in Table 12). Pedestrians on a crosswalk is an indication that it is safe to cross. Nevertheless, the number of pedestrians on crosswalks will cause much more delay for Transition 1-2-3 (0.105*** in Table 12 for Transition 1-2). This indicates that, if there are many pedestrians already on the crosswalk and a pedestrian chooses to wait, he or she should wait for a longer time for 1-2-3.

(B). Figure 16 (b) indicates that the number of pedestrians waiting on curb will result in a delay for subject pedestrian to cross for Transition 1-3 (-0.157*** in Table 12).
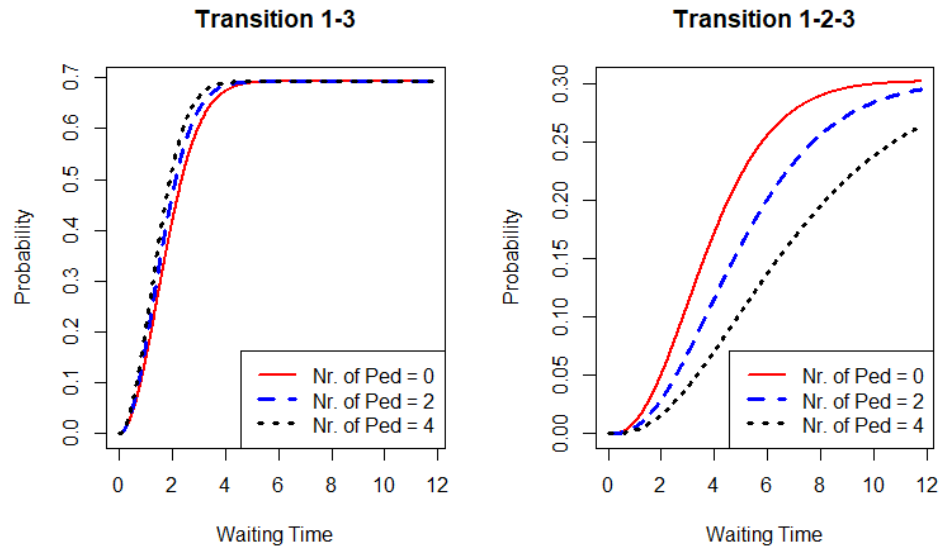
(a) Nr. of Pedestrians on Crosswalks
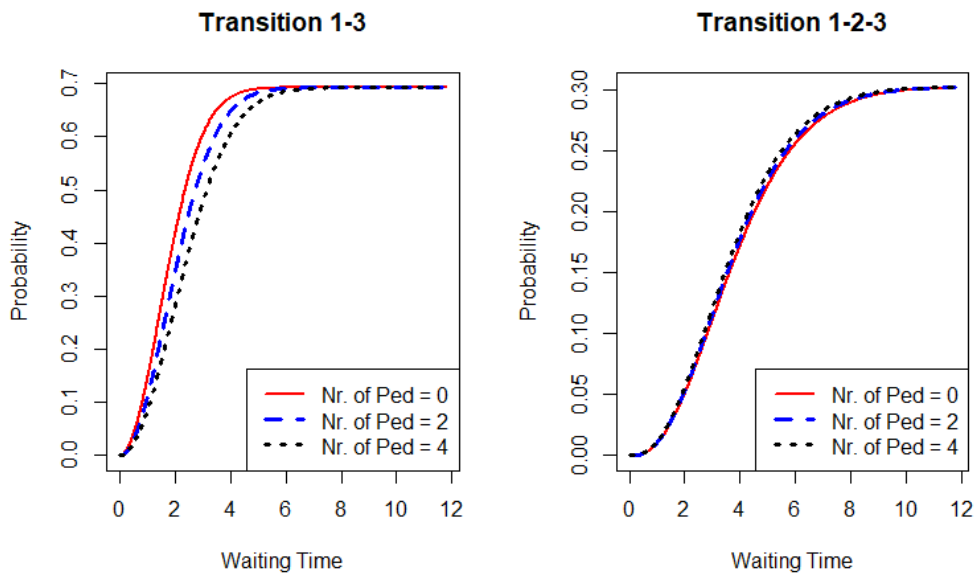
(b) Nr. of Pedestrians on Curb Waiting

Figure 16 Pedestrian Impacts

6.2.2 *Vehicle Dynamics*

Figure 17(a) demonstrates that a platoon of vehicles has little impact on the decision-making process for Transition 1-3 because pedestrians only "negotiate" with the leading vehicle while making a decision. The close follower indicator has effects on delay on Transition 1-2-3 (-0.459*** for Transition 1-2 in Table 12), however. For Transition 1-2-3, if a pedestrian chooses to wait for a platoon of vehicles, he or she can expect to wait for a longer time compared with individual vehicle.
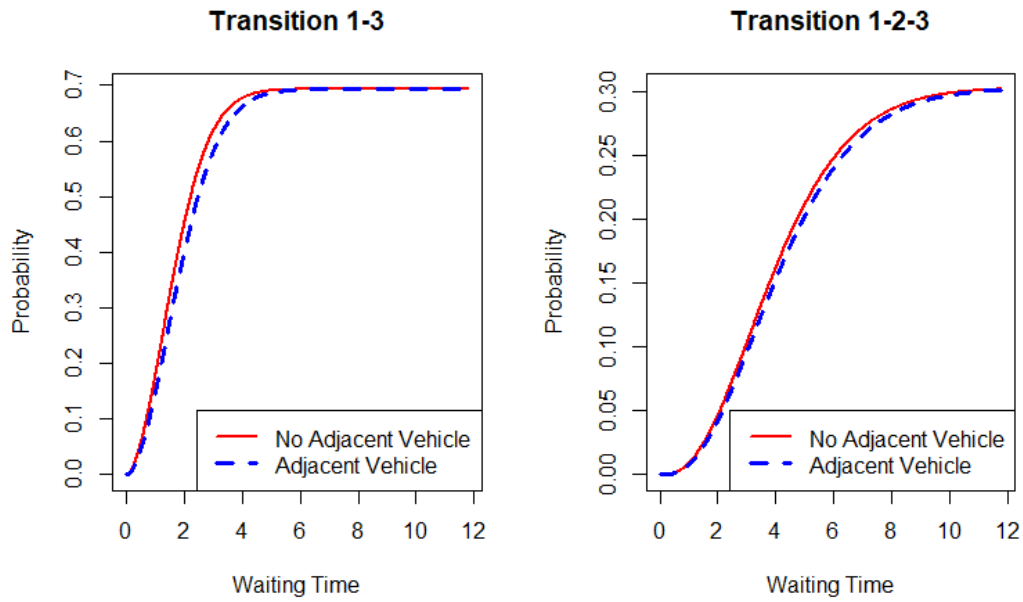
The adjacent vehicle indicator has a negative impact on Transition 1-3 (-0.196* in Table 12) because pedestrians have to "negotiate" with two different vehicles in different lanes. See Figure 17(b).
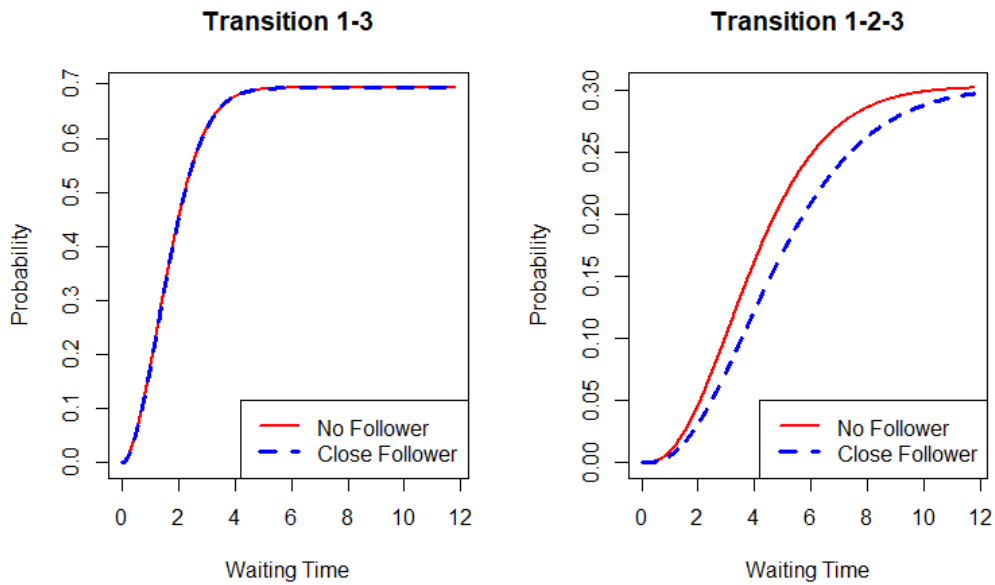
(a) Close Following Vehicle Impact

Figure 17 Multiple Vehicle Effects

Figure 17 continued

(b) Adjacent Vehicle Impact

6.2.3   *Hesitation*

The *Hesitation* parameter (-1.364* for Transition 1-2; -1.286*** for Transition 1-3; and -1.316*** for Transition 2-3 in Table 12) has effects on delay on pedestrian waiting behavior. This is intuitive because, if a pedestrian hesitates in the curb area, the misunderstanding between pedestrians and motorists increases. This will delay the pedestrian's time to cross. See Figure 18.
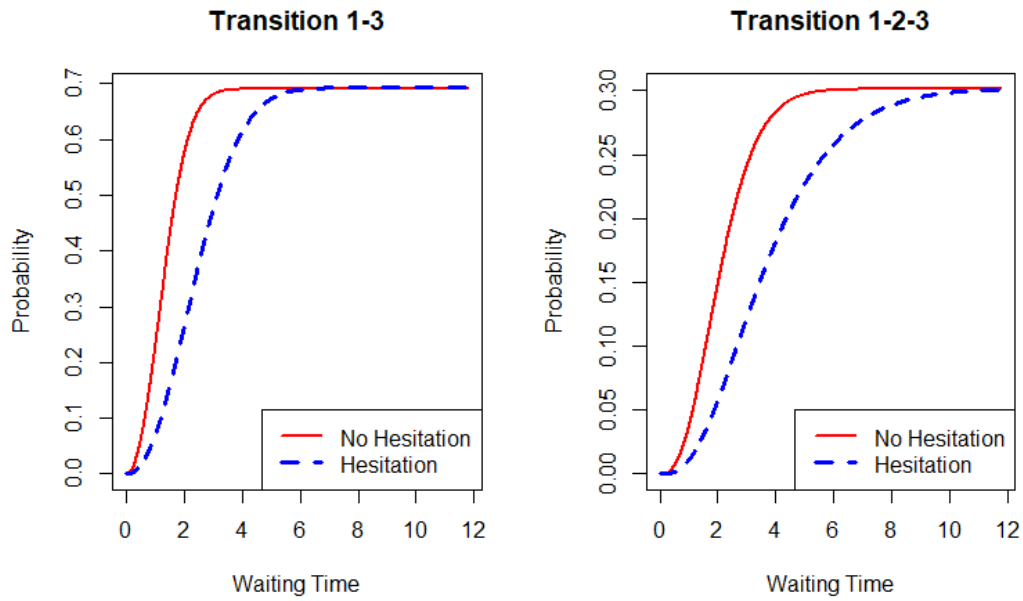
Figure 18 Hesitation Parameter Effects

6.2.4 *Pedestrian Characteristics*

Figure 19 illustrates the effects of pedestrian characteristics on pedestrian waiting process. Compared with other groups, *young* pedestrians are more assertive for Transition 1-3 (-0.355*** in Table 12) because they have lower wait durations than do other groups. See Figure 19(a). Besides, males are less likely than females to wait for subsequent gaps (0.325* in Table 12). See Figure 19(b).
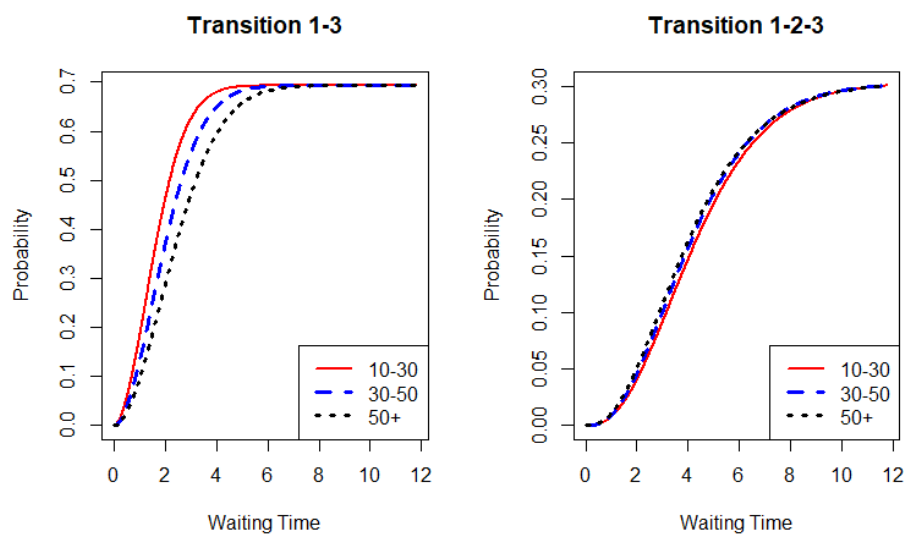
(a) Age
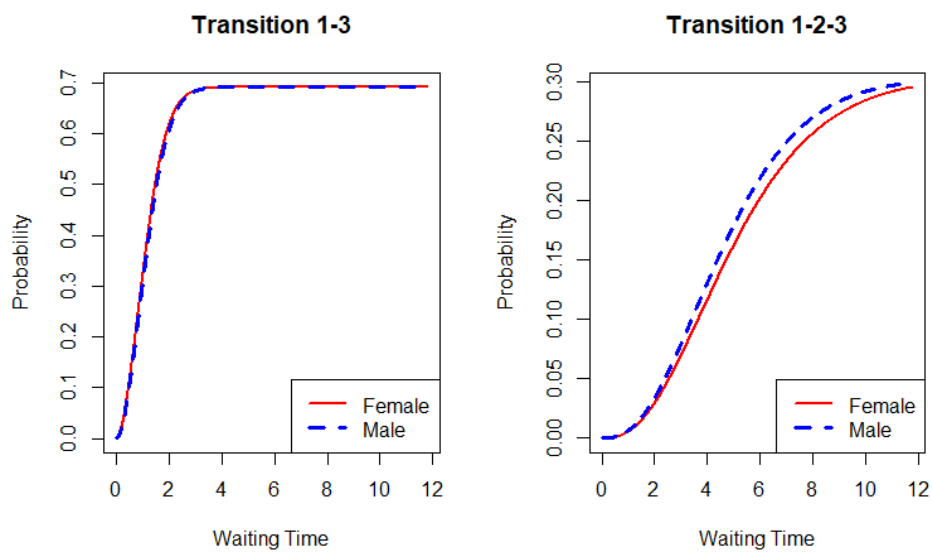
(b) Sex

Figure 19 Pedestrian Characteristics Impact

6.2.5 *Distance and Speed*

(A). Distance has significant impact on pedestrian waiting behavior. For Transition 1-3 (0.303***
in Table 12), with the increase of distance to vehicle, pedestrian wait time is reduced. However, it
has inverse effects on the Transition 1-2-3. A shorter distance to the first interacted vehicle will
result in a faster Transition 1-2 (-0.635*** in Table 12) and then speed up the waiting process for
Transition 1-2-3. The closer the vehicle is to the crosswalk, the more uncertain and unsafe a
pedestrian feels. "Let the car go first" will be a safe crossing strategy for pedestrians, if the
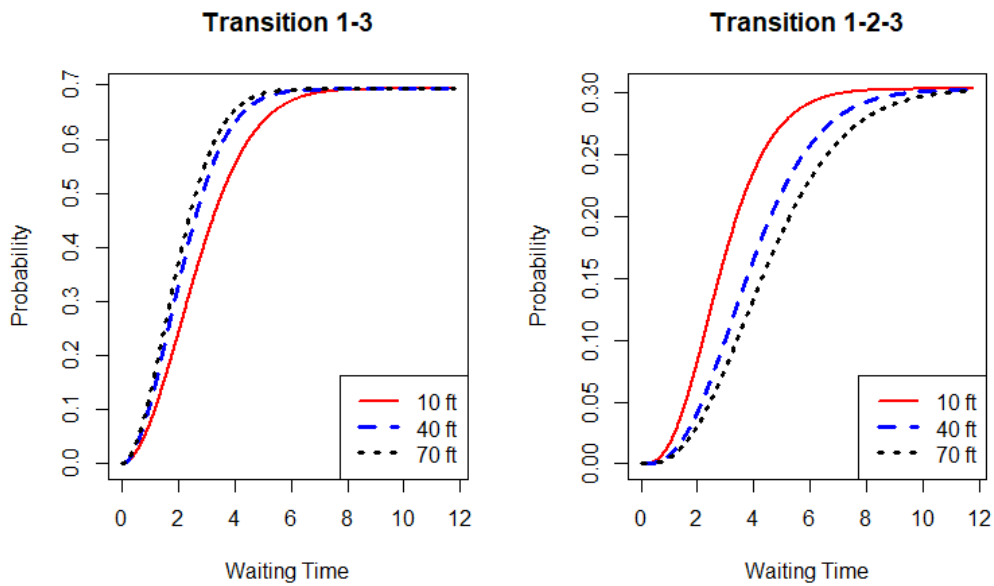interacted vehicle is too close to yield. See Figure 20.

Figure 20 Distance Parameter Effects

(B). The effects of vehicle speed on pedestrian decision making are shown in the following figure.
We can see that, with the increase in speed, pedestrian wait time of Transition 1-3 (-0.031*** in
Table 12) is increasing, while the pedestrian wait time of Transition 1-2-3 is decreasing. The first
interacted vehicle will pass the area quickly with a higher speed, which result in a faster Transition
1-2 (0.057*** in Table 12) so that it reduces the delay for Transition 1-2-3. "Let the car go first"
will also be a safe crossing strategy for pedestrian to cross, if the interacted vehicle is too fast to
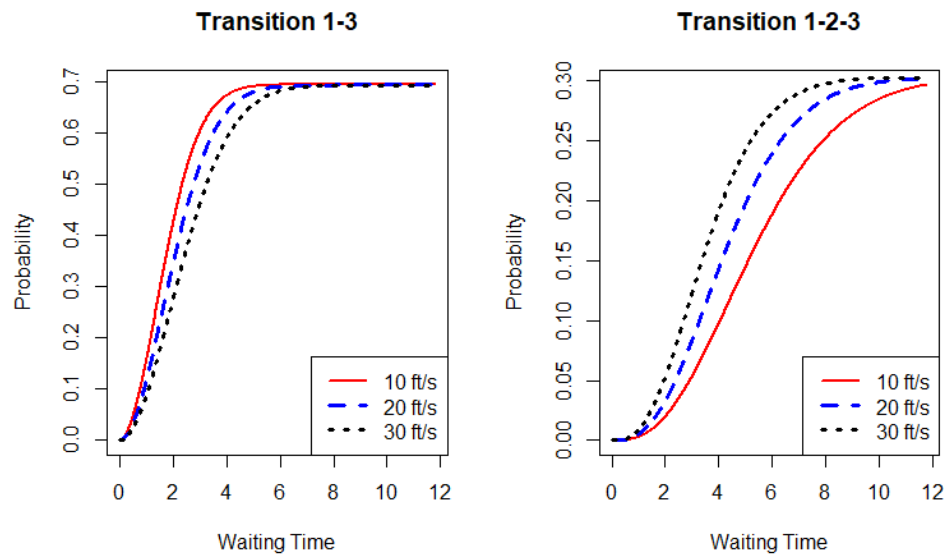yield. See Figure 21.

Figure 21 Speed Parameter Effects

6.3   Before and After Studies

6.3.1   *Group Effects*

The Markovian model shows results similar to the survival model. The number of pedestrians in a group significantly decreases the pedestrian delay in Transitions 1-3 (0.078** in Table 12) and Transition 2-3 (0.199* in Table 12). See Figure 22.
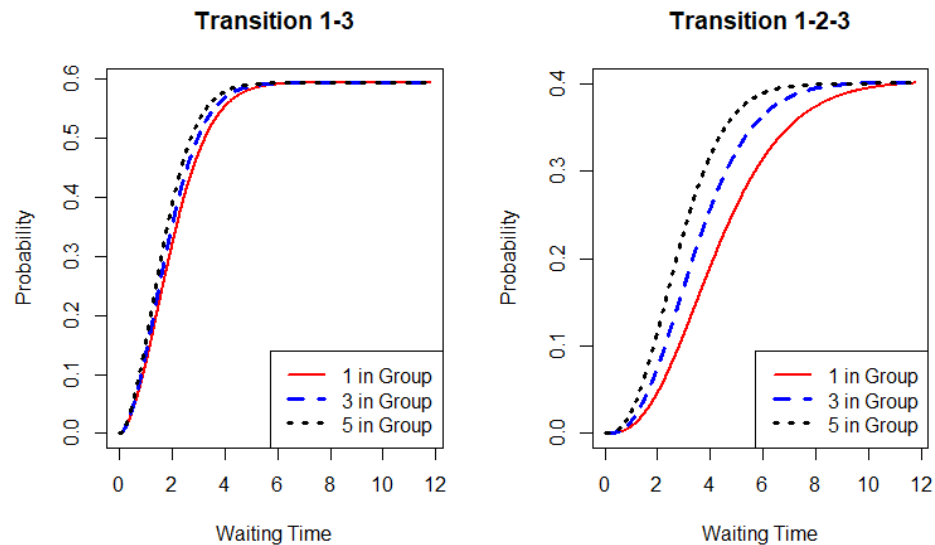

Figure 22 Group Effects on Two-Way Case

6.3.2   *Using Cellphone or Talking*

While using a cellphone or talking, pedestrians have a lower risk moving from state 2 to state 3 (-0.312. in Table 12), which results in a higher delay. See Figure 23.
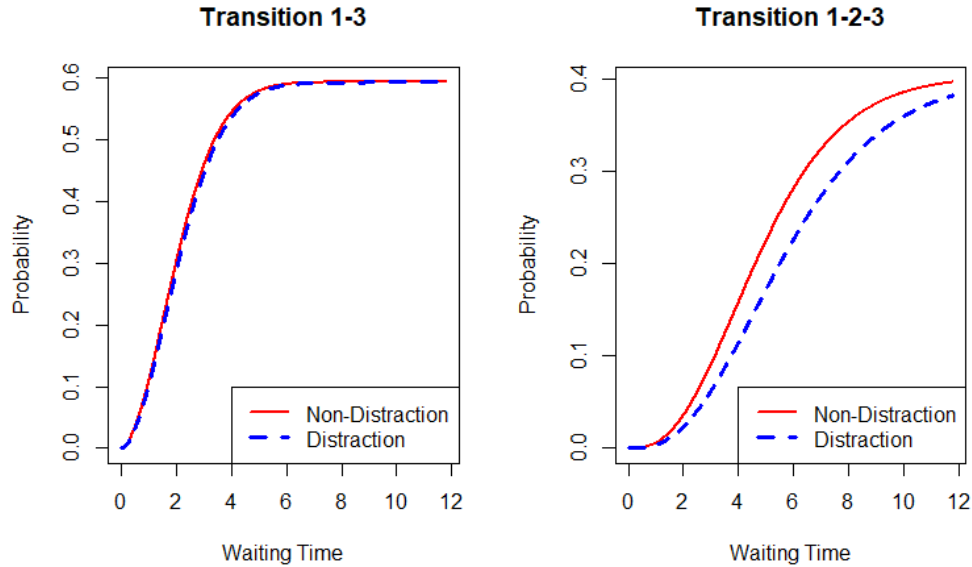

Figure 23 Distraction Effects on Two-Way Case

6.3.3   *Vehicle Type*

Vehicle Type affects delay in Transition 1-2-3 with one-way street operation period. This indicates that, if a pedestrian chooses to yield to a bus (Transition 1-2 -0.173*** in Table 12), he/she has to wait longer. However, after conversion to a two-way street, the crossing probability for Transition 1-3 is increasing (0.318* in Table 12). After the conversion from one-way to two-way operation, Lafayette CityBus was required to remove most bus routes from University Street. Therefore, in the two-way case, there were fewer buses.

6.3.4   *Adjacent Vehicle*

Adjacent vehicle indicator has no effects on pedestrian wait time on two-way cases. This means that whether a pedestrian waits or not is less likely to be affected by an adjacent vehicle when a pedestrian-motorist interaction occurs. We found the similar results in the analysis of motorist behavior in Section 4.3.4.

CHAPTER 7.   CONCLUSIONS

7.1   Pedestrian-Motorist Interaction

Table 13 summarizes the variables (defined in Table 1) that were found to be significant in explaining the pedestrian and motorist behavior seen in the recorded video at the University Street crosswalks. Some variables were found to be significant in both the one-way and two-way street cases. Combining the one-way and two-way data, we were able to identify factors that changed significantly between the two cases. These are shown with check marks in the "Significant difference" columns of Table 13.

Table 13 Summary of Model Results

Variable	Pedestrian Model		Motorist Model	
	Significant in both cases	Significant difference	Significant in both cases	Significant difference
ApprSpeed	√	—	√	—
Hesitation	√	√	√	√
FlowWith/On	—	√	√	—
SlowsDown	√	—	—	—
Distance	√	—	—	√
AdjVeh	—	√	—	√
NoF	√	—	—	√
GroupSize	—	√	—	—
— = not applicable				

The model results are consistent with expectations in terms of the direction of influence. Some examples are described below.

    1. A pedestrian is more likely to cross during an interaction at the semi-controlled crosswalk if the approaching vehicle is moving at a slow speed, is slowing down, or is far enough away from the crosswalk. However, what this research offers is a more quantitative assessment of the pedestrian response to these and other factors, as well as highlighting the importance of a pedestrian's actions with respect to hesitation. While the findings of the pedestrian model are

largely behavioral, there are some practical aspects to the findings. For example, the *approach speed* finding above can be translated into a speed that would lead to a desired likelihood of pedestrians choosing to cross, all else being equal.

2. The study found factors that affect an approaching driver's behavior, which focused on a driver's likelihood of slowing down for pedestrians, rather than the likelihood of yielding. Examples of these findings with respect to two-way vehicle traffic operation are:

A. A greater number of pedestrian characteristics factors (Hesitation, Distraction and FlowOn) have a significant impact on a driver's willingness to decelerate in the two-way case than in one-way operation.

B. Except for the speed variable (AppSpeed), variables concerning vehicle dynamics and characteristics become insignificant (CloseFollow and AdjVeh), when compared with one-way operation. This means that a driver on the two-way street is less likely to be affected by a close-following vehicle or by an adjacent vehicle when a pedestrian-motorist interaction occurs.

C. Environmental characteristics factors (Distance and NoF) became significant in a driver's decision to slow down, compared to one-way operation.

A driver's decision is mainly influenced by interacted pedestrian behavior and the environmental characteristics when an interaction occurs. The change of one-way to two-way operation removed the effects of interaction between vehicles (CloseFollow and AdjVeh) on a driver's decision and led drivers to react more to the interacted pedestrian.

7.2   Pedestrian Waiting Time

This research describes the data and models used to analyze the wait durations of pedestrians when they interact with vehicles at "semi-controlled" crosswalks. The variables for 2059 pedestrian wait durations were carefully defined and measured from video recordings.

Survival models and multi-state semi-Markov models were developed and compared. For the survival models, a first-event analysis was conducted, which suggested non-monotonic relationships of distance and speed on pedestrian waiting behavior in one-way operation. In multi-state semi-Markov models, a recurrent events analysis was undertaken, which examined transition-specific covariates on pedestrian waiting behavior. The two models showed different results.

The survival model revealed different waiting behaviors in terms of pedestrian characteristics, vehicle dynamics and environmental factors with one-way and two-way operation. Multi-state semi-Markov model suggested consistent pedestrian waiting behaviors when one-way convert into a two-way operation. Compared with survival models,

1. Multi-state semi-Markov models can better explain pedestrian delay, because they incorporate the yielding behavior of motorists in the determination of waiting transitions (Transition 1-3 vs Transition 1-2-3).

2. Multi-state semi-Markov models help explain non-monotonic relationships of speed and distance versus pedestrian wait time found in the survival model.

There are two limitations for multi-state semi-Markov models,

1. Multi-state semi-Markov models can be further improved by including random effects to correlate in-group observations. Furthermore, it can be flexible if more sojourn time distributions are tested.

2. Another limitation for Markovian model is that existing statistical software cannot deal with the Transition 2-2. This lost the information when pedestrians rejected multiple gaps. We hope to solve this problem in my future research.

In Chapter 5, Survival models reveal non-monotonic relationships between distance, speed versus pedestrian wait time in one-way case (Figure 11). In Chapter 6, we analyze the effects of motorist yielding behavior on pedestrian wait durations, which provides explanations for the non-

monotonic relationships (See Figure 20 and 21). The non-monotonic relationships are surprising results that have a plausible explanation: Firstly, if a vehicle is too close to the crosswalk and moving too fast to yield, the normal pedestrian choice is to "let the vehicle go first", and it will cause little delay to the pedestrian (See Transition 1-2-3 in Figure 20 and 21). And if a vehicle is too far or too slow, then normal pedestrian choice is to cross directly without any hesitation because pedestrian will feel safe (See Transition 1-3 in Figure 20 and 21). However, if the vehicle is neither too far nor too close with an approach speed around 22 mph (around peak points in Figure 11), negotiations between pedestrian and motorist will be more complicated and more pedestrian delay would be incurred.

In addition, both survival models and Markovian models reveal that pedestrian will wait less when the object pedestrian is in a group in two-way case (See Figure 13 and 22). Traffic engineers can consider geometry designs such as curb designs to indirectly increase pedestrian arriving groups and reduce pedestrian delay.

Future research should focus specifically on the areas where the most complicated negotiations between pedestrians and motorists occur. Additionally, simple model structures may not explain the complicated interaction adequately and the most advanced frontier of modeling approaches such as Markov switching models, multivariate models etc. can be further explored.

LIST OF REFERENCES

Akaike, H. (1987). Factor analysis and AIC. In Selected Papers of Hirotugu Akaike (pp. 371-386). Springer, New York, NY.

Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. The annals of statistics, 1100-1120.

Bjørnskau, T. (2017). The Zebra Crossing Game–Using game theory to explain a discrepancy between road user behaviour and traffic rules. Safety science, 92, 298-301.

Brüde, U., & Larsson, J. (1993). Models for predicting accidents at junctions where pedestrians and cyclists are involved. How well do they fit?. Accident Analysis & Prevention, 25(5), 499-509.

Chung, Y. (2010). Development of an accident duration prediction model on the Korean Freeway Systems. Accident Analysis & Prevention, 42(1), 282-289.

Cox, D. R. (1972). Regression models and life- tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2), 187-202.

Cloutier, M. S., Lachapelle, U., d'Amours-Ouellet, A. A., Bergeron, J., Lord, S., & Torres, J. (2017). "Outta my way!" Individual and environmental correlates of interactions between pedestrians and vehicles during street crossings. Accident Analysis & Prevention, 104, 36-45.

Cross Validated, Test the Significant Differences Between the Coefficients Obtained by Logistic Regression on Two Data Sets, https://stats.stackexchange.com/questions/20330/test-the-significant-differences-between-the-coefficients-obtained-by-logistic-r, Accessed December. 17, 2018.

De Dominicis, R., & Manca, R. (1984). An algorithmic approach to non-homogeneous semi-Markov processes. Communications in Statistics-Simulation and Computation, 13(6), 823-838.

Fricker, J.D., & Zhang, Y. (in press). Modeling pedestrian and motorist behavior at semi-controlled crosswalks: the effect of a change from one-way to two-way street operation. Transportation Research Record.

Hamed (2011), M. M. Analysis of pedestrians' behavior at pedestrian crossings. Safety Science. 2011. 38(1), 63-82.

Harrell, F. E. (2015). Multivariable modeling strategies. In Regression modeling strategies (pp. 63-102). Springer, Cham.

Himanen, V., & Kulmala, R. (1988). An application of logit models in analysing the behaviour of pedestrians and car drivers on pedestrian crossings. Accident Analysis & Prevention, 20(3), 187-197.

Guan, J., Wang, K., & Chen, F. (2016). A cellular automaton model for evacuation flow using game theory. Physica A: Statistical Mechanics and its Applications, 461, 655-661.

Guo, H., Gao, Z., Yang, X., & Jiang, X. (2011). Modeling pedestrian violation behavior at signalized crosswalks in China: A hazards-based duration approach. Traffic injury prevention, 12(1), 96-103.

Guo, H., Wang, W., Guo, W., Jiang, X., & Bubb, H. (2012). Reliability analysis of pedestrian safety crossing in urban traffic environment. Safety Science, 50(4), 968-973.

Kadali, B. R., Rathi, N., & Perumal, V. (2014). Evaluation of pedestrian mid-block road crossing behaviour using artificial neural network. Journal of traffic and transportation engineering (English edition), 1(2), 111-119.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282), 457-481.

Leaf, W. A., & Preusser, D. F. (1999). Literature review on vehicle travel speeds and pedestrian injuries. US Department of Transportation, National Highway Traffic Safety Administration.

Lee, R. (2000). The Lee-Carter method for forecasting mortality, with various extensions and applications. North American actuarial journal, 4(1), 80-91.

Li, B. (2013). A model of pedestrians' intended waiting times for street crossings at signalized intersections. Transportation research part B: methodological, 51, 17-28.

Lord, S., Cloutier, M. S., Garnier, B., & Christoforou, Z. (2018). Crossing road intersections in old age—With or without risks? Perceptions of risk and crossing behaviours among the elderly. Transportation research part F: traffic psychology and behaviour, 55, 282-296.

Michalaki, P., Quddus, M. A., Pitfield, D., & Huetson, A. (2015). Exploring the factors affecting motorway accident severity in England using the generalised ordered logistic regression model. Journal of safety research, 55, 89-97.

Nam, D., & Mannering, F. (2000). An exploratory hazard-based analysis of highway incident duration. Transportation Research Part A: Policy and Practice, 34(2), 85-102.

NHTSA, U. (2016). Department of Transportation Releases Policy on Automated Vehicle Development.

Ouhbi, B., & Limnios, N. (2003). Nonparametric reliability estimation of semi-Markov processes. Journal of Statistical Planning and Inference, 109(1-2), 155-165.

Papadimitriou, E. (2012). Theory and models of pedestrian crossing behaviour along urban trips. Transportation research part F: traffic psychology and behaviour, 15(1), 75-94.

Papadimitriou, E., Lassarre, S., & Yannis, G. (2016). Introducing human factors in pedestrian crossing behaviour models. Transportation research part F: traffic psychology and behaviour, 36, 69-82.

Paselk, T. A., & Mannering, F. L. (1994). Use of duration models for predicting vehicular delay at a US/Canadian border crossing. Transportation, 21(3), 249-270.

Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. Biometrika, 68(2), 373-379.

Salamati, K., Schroeder, B. J., Geruschat, D. R., & Rouphail, N. M. (2013). Event-based modeling of driver yielding behavior to pedestrians at two-lane roundabout approaches. Transportation research record, 2389(1), 1-11.

Schroeder, B. J., & Rouphail, N. M. (2010). Event-based modeling of driver yielding behavior at unsignalized crosswalks. Journal of transportation engineering, 137(7), 455-465.

Sucha, M., Dostal, D., & Risser, R. (2017). Pedestrian-driver communication and decision strategies at marked crossings. Accident Analysis & Prevention, 102, 41-50.

Sun, D., Ukkusuri, S. V. S. K., Benekohal, R. F., & Waller, S. T. (2003, January). Modeling of motorist-pedestrian interaction at uncontrolled mid-block crosswalks. In Transportation Research Record, TRB Annual Meeting CD-ROM, Washington, DC.

Tiwari, G., Bangdiwala, S., Saraswat, A., & Gaurav, S. (2007). Survival analysis: Pedestrian risk 21 exposure at signalized intersections. Transportation research part F: traffic psychology and 22 behaviour, 10(2), 77-89.

Wang, X., & Abdel-Aty, M. (2008). Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accident Analysis & Prevention, 40(5), 1674-1682.

Washington, S. P., Karlaftis, M. G., & Mannering, F. (2010). Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC.

Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American statistical association, 84(408), 1065-1073.

Williams, R. (2018). GOLOGIT2: Stata module to estimate generalized logistic regression models for ordinal dependent variables.

Yang, X. , Huan, M. , Abdel-Aty, M. , Peng, Y. , & Gao, Z. . (2015). A hazard-based duration model for analyzing crossing behavior of cyclists and electric bike riders at signalized intersections. Accident Analysis & Prevention, 74, 33-41.

Yannis, G., Papadimitriou, E., & Theofilatos, A. (2013). Pedestrian gap acceptance for mid-block street crossing. Transportation planning and technology, 36(5), 450-462.

Ye, F., & Lord, D. (2014). Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit and mixed logit models. Analytic methods in accident research, 1, 72-85.

Zhuang, X., & Wu, C. (2011). Pedestrians' crossing behaviors and safety at unmarked roadway in China. Accident analysis & prevention, 43(6), 1927-1936.

PUBLICATIONS

Fricker, J.D., & Zhang, Y. (in press). Modeling pedestrian and motorist behavior at semi-controlled crosswalks: the effect of a change from one-way to two-way street operation. Transportation Research Record.