Work Zone Safety Assessment and Throughput Analysis for High Volume Highways Using Random Parameter Models

by

Seyedata Nahidi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Said Easa
Professor, Dept. of Civil and Environmental Engineering
Ryerson University

Supervisor(s): Susan L. Tighe
Professor, Dept. of Civil and Environmental Engineering,
University of Waterloo

Internal Member: Hassan Baaj
Professor, Dept. of Civil and Environmental Engineering,
University of Waterloo

Internal Member: Chris Bachmann
Professor, Dept. of Civil and Environmental Engineering,
University of Waterloo

Internal-External Member: Alex Penlidis
Professor, Dept. of Chemical Engineering,
University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Highways and freeways could be considered the most important transportation infrastructure in North America; these vital routes are necessary for the efficient haulage of huge amounts of goods and services. Several factors such as the high volume of heavy truck traffic as well as harsh winters in this region could result in a faster deterioration rate of the transportation infrastructures, specifically pavements. Transportation agencies, under the supervision of municipalities, are responsible to maintain, preserve, and reconstruct these segments. Applying the proper care results in a significant reduction in the number of observed conflicts and collisions on high-volume highways.

Washington State Department of Transportation defines a work zone as,

"...an area of a roadway with construction, maintenance, or utility work activities. A work zone is typically marked by signs, channelizing devices, barriers, pavement markings, and/or work vehicles."

Based on previously conducted studies, work zones can significantly interrupt the regular traffic flow on highways. These interruptions can have adverse effects on the safety of the roads and increase the likelihood of undesirable conflicts and collisions. To avoid any unexpected work zone related safety concerns, Departments of Transportation in the US, as well as Ministries of Transportation in Canada, encourage agencies to propose detailed plans to minimize the queuing period and injury severity of work zone collisions; the most common strategy is to set up work zones at nighttime. Independent reports by the Ministry of Transportation Ontario (MTO) also identified that predicting the throughput, and the queuing length, as well as the queuing period, can significantly improve the planning stage, reduce the user delay costs, and increase work zone safety for workers and motorists.

Statistical analyses and modelling are methods used to acquire information from historical data sets and gain a more realistic insight into future events with an acceptable confidence level. This research involves the statistical evaluation of work zones' safety and performance, along with comprehensive analyses of work zones' throughput in North America. To evaluate the different strategies, Multiple Linear Regression (MLR) and Negative Binomial Regression (NBR) models were developed to identify the critical historical factors which affect the traffic throughput of work zones. For safety assessment of work zones, innovative random parameter approaches were adapted in combination with ordered probability models to produce robust and realistic results. Furthermore, the practicality and applicability of random parameter models were discussed to clarify the advantages of using these models. Random Parameter Negative Binomial (RPNB) and Random Parameter Ordered Logit (RPOL) models developed in this study were found to be the most accurate

models for throughput and safety analysis, respectively. Also, the implementation of k-fold cross validation proved that the model predictions correlated well with historical data. Finally, a new approach for Random Parameter prediction was proposed which considers the similarity level between a potential event and historical data.

Based on these evaluations, the overall feasibility of each strategy was examined. The results denoted several practical recommendations to decrease traffic congestion and create safer work zones. The random parameter negative binomial model for throughput analysis showed that to avoid queuing in work zones where there are two or more obstructed lanes, multiple short (less than 3 km) work zones are more efficient than longer ones; this factor increases the frequency of passing vehicles by 177 per hour per lane. Besides, weekend nights are found to be the most appropriate time to set up work zones. It is observed that weekend nighttime work zone set-ups increase the number of passing vehicles by 493 vehicles per hour per lane compared to other scenarios. In general, nighttime closures, occurring on any day during the week, are found to have a higher discharge rate in comparison with daytime closures. On highways with more than 20% truck traffic, it is expected to have 102 fewer vehicles passing through work zones due to the induced congestion. Similarly, random parameter ordered probability models identified several factors which are shown to have a statistically significant impact on work zone collisions' injury severity level. As an example, aggressive driving behaviours, e.g. failing to keep in the proper lane, running other drivers off the road, and tailgating, increase the major injury and fatal collisions' likelihood by 78%. The installation of traffic control devices, specifically warning signs, reduce the probability of fatalities by 14%. Moreover, alcohol and drug consumption increase the probability of fatal and major injury collisions by 36% based on random parameter ordered Logit model, so by enforcing strict laws many lives can be saved.

After analysis, common practices and the author's recommendations for each significant factor in the selected models are discussed. Primarily, the prohibition of truck traffic, designing efficient detours, and installation of extra and more innovative traffic control devices prior to the work zones are recommended. It was also concluded that the most efficient way to have a safe and comfortable environment in work zones on high-volume highways is to encourage government, engineers, and motorists to collaborate. Collaboration could take the form of the public awareness campaigns, setting and enforcing effective laws and regulations, and assuring the proper implementation of existing guidelines.

Last but not least, the accurate prediction of work zone throughput frequency at queuing time provided an appropriate context for better work zone planning to reduce the possible user delay cost. The outcome of this research was the development of a novel planning and decision-making tool ('smart form') to help engineers and contractors to evaluate the work zone safety of high-volume highways in North America.

Acknowledgements

I was actually mulling it over about how to start my "acknowledgements" section. This is the only part of my thesis where I can reveal my inner personality and mindset; therefore, I have decided to start it with a quote from Steven D. Levitt[1], who said:

"Most of us want to fix or change the world in some fashion. But to change the world, you first have to understand it."

I believe every single person has a role to play in this world and it is my belief that if someone can conduct research which can potentially save human lives, that could be a great role with a great impact. Throughout my research, I came across fatality statistics and discovered that annually 1.3 million people lose their lives in traffic-related collisions. This fact inspired me to look for a solution for this issue, and that inspiration developed into my passion, turned to be the thing that I wanted to change in this world. Based on what Prof. Levitt mentioned in his quote, I decided to understand it better and deeper. Throughout this journey many people helped me and I would like to thank them personally.

Firstly, I would like to express my sincere gratitude to my advisor Prof. Susan L. Tighe for her continuous support during my PhD study, for her patience, for being the source of my motivations, and her immense knowledge. Her guidance enlightened my path during this period and I could not have imagined having a better advisor and mentor for my PhD study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Hassan Baaj, Prof. Chris Bachmann, Prof. Alex Penlidis, and Prof. Said Easa for their insightful comments and encouragement, but also for the challenging questions which heartened me to widen my research from various perspectives.

I thank my fellow colleagues in CPATT for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the past three years.

Last but not the least, I would like to thank my family: my parents and my sister for supporting me unconditionally and spiritually throughout writing this thesis and my life in general.

[1] Steven D. Levitt is a well-known economist and writer of the New York Times Best Seller Book "Freakonomics"

Dedication

This is dedicated to my beloved parents

Jalal Nahidi and Simin Hamidini

and to my lovely sister

Asal

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Maintenance, rehabilitation, and preservation of high-volume Ontarian highways are one of the most important responsibilities of the Ministry of Transportation Ontario (MTO). The 400-series highways inside and around the Greater Toronto Area (GTA) are responsible for transporting vast numbers of people as well as goods and resources [1]. Thus, setting up the work zones on these highways should be well planned to minimize and ideally eliminate possible traffic flow interruptions and safety issues.

High-volume highways are generally confronted with many issues such as rapid pavement and utility deterioration. These issues urge the MTO to frequently set up work zones to maintain the serviceability of these roads. Setting up work zones in high-volume areas considerably increases the user delay cost, and collision frequency and severity. To combat these concerns, current MTO policies and regulations can be altered to lessen the consequences of work zone.

Several Departments of Transportation within the United States and Ministries of Transportation in Canada have examined inventive and traditional methods to deal with the issues inside and around the work zones. Installing portable barriers, the prohibition of heavy vehicles, decreasing the speed limit, and strict traffic enforcement were some of the strategies that were examined to improve safety. It was observed that the presence of the law enforcement officers at the start of the work zones reduced the violation frequency. Also, to capture the influence of trucks and other heavy vehicles on lane blockage and secondary crashes, these vehicles were prohibited in the work zones [2].

Generally, agencies prefer to schedule their work zone set-ups and activities at nighttime. Higher traffic during the daytime and a longer period of off-peak hours during the nighttime encourage ministries of transportation to set up the work zones toward the end of the day.

This arrangement is to avoid excessive traffic and queuing on major highways; this action also lessens the inconvenience for motorists during the daytime undoubtedly [3].

This study aims to investigate whether the current work zone strategies and guidelines are appropriate for traffic throughput, and if not, which additional factors should be accounted for to improve the safety and traffic of the work zones. The gaps in previous studies will be identified and a comprehensive study will be conducted to find and recommend solutions for them. To that end, high-performance statistical models will be used to, first, predict the possible range of throughput in these sites at the time of the queuing. Secondly, based on historical data, factors with significant effects on severity of collisions will be identified. Finally, by combining the results derived from these approaches, the most important countermeasures will be presented to improve the safety and performance of the work zones.

1.1 Research Background

Previously, the MTO has conducted two studies between 2007 and 2011 to evaluate the work zone throughput pattern in time of the queuing in high-volume highways. In these studies, several work zones were visited and throughput data was collected. The data set was analyzed and several multiple linear regression models were developed to predict the possible future throughput. These generic models had the benefit of providing broader insight into the factors which play a role in work zone throughput at the queuing condition. However, these studies did not focus on the safety of the work zones [4][5].

Over the last few decades, the highway design, vehicle safety, and traffic guidelines have been advanced, but the number of collisions which resulted in fatalities and injuries are still unacceptably high. Transport Canada [6] reported that around 155,000 collisions occurred in 2017 which resulted in serious injuries for drivers, passengers, and pedestrians. Based on the same report, almost 1700 of the collisions resulted in fatalities. However, there was no clear evidence whether these fatalities were related to the work zones or not. Work zones hypothetically could be considered as high-risk locations. Specifically for nighttime work zones, factors such as low visibility, more impaired drivers, and higher speed increase the likelihood of experiencing more severe injuries and fatalities. Lack of attention to the work zone safety and the importance of accurately estimating the throughput (for later planning purposes and capacity analysis) were the major motivations of this thesis.

1.2 Research Hypotheses

The main hypotheses for this research are as follows:

- Work zone configurations, location, and time of closures are likely to affect the queuing in the work zones.
- With the collaboration of the government, engineers, and motorists, the safety of the work zones can be improved.
- Random parameter Negative Binomial model provides the most reasonable and logical results due to its compatibility with the nature of the response variable and its luxury to account for unobserved heterogeneity.
- Different distributions for factor estimations alter the performance of ordered probability models.
- The recommended methodology based on the prediction of the random parameter models defines the relationship between work zones and injury-severity of collisions.
- The designed smart form works as a system, which summarizes all of the underlying safety models, converts work zone information into the probability of occurrence of each injury severity class.

1.3 Research Scope and Objectives

This research is focused on the work zone throughput and safety analysis; the final goal is to address the traffic and safety-related concerns, e.g. satisfying the 'vision zero' goals, and reducing queuing time. Multiple statistical models will be developed to investigate the influential factors on queuing and collision severity in these locations. The results from of this study will help MTO to overcome the mentioned issues in work zones and provide a safe environment for motorists and workers.

The random parameter models which are proposed for this research can account for unobserved heterogeneity and present more accurate and detailed predictions for throughput and injury-severity level of collisions.

The main objectives of this research are as follows,

- Investigation of the possible influential factors by developing more comprehensive statistical models and detailed data sets that could potentially lessen the congestion and improve road safety in or around the work zones located at the high-volume highways.
  - Development of fixed and random parameter multiple linear regression and negative binomial models for throughput analysis.
  - Development of random parameter ordered probability [Probit, Logit, Arctangent] models for safety analysis.
  - Prediction and performance assessment of the developed models for both throughput models and safety models; and recommend the best models for each purpose.
- Recommendation of an innovative methodology to predict the injury severity level of future events.
  - Suggestion of a new parameter which can quantify the similarity level of the upcoming event with the historical data.
  - Proposing a new methodology to predict with random parameter models by considering their ability to change across the observations.
  - Conducting the 10-fold cross validation methodology to have more robust results.
- Designing a smart form to determine the probability of each injury severity class in future work zone events.
  - Preparing a user-friendly form which can be easily used in industry.
  - Converting the conventional deterministic approach to innovative probabilistic approach.

Chapter 2 discusses the fundamentals associated with this research through a comprehensive literature review. The identified gaps in current research are explored and discussed as related to the objectives of this thesis. Chapter 3 discusses the methodological framework and data collection procedure. This chapter presents the concepts related to the various statistical and econometric models adopted for this thesis; the compatibility of the collected data with recommended methodologies is also evaluated. Chapter 4 explains the detailed findings related to work zone throughput; the results of this chapter were presented at the

"2018 Canadian Society of Civil Engineers Conference" and a comprehensive report of the findings of this chapter was also presented to Ministry of Transportation Ontario. The data collected for this chapter was gathered in collaboration with CIMA+ and University of Toronto. Chapter 5 deliberates the major safety concerns imposed by high-volume highway work zones. Several solutions are proposed to minimize fatal incidents and major collisions. Results of this section were presented at the "2019 Canadian Society for Civil Engineers Conference". The k-fold cross validation strategy is also discussed as related to the verification of the results of each model. An innovative approach for predicting future events is proposed by considering the similarity level of historical events with present data. Finally, Chapter 6 discusses the conclusions, recommendations, contributions of this research, and future work. Figure 1.1 presents a breakdown of all six chapters.

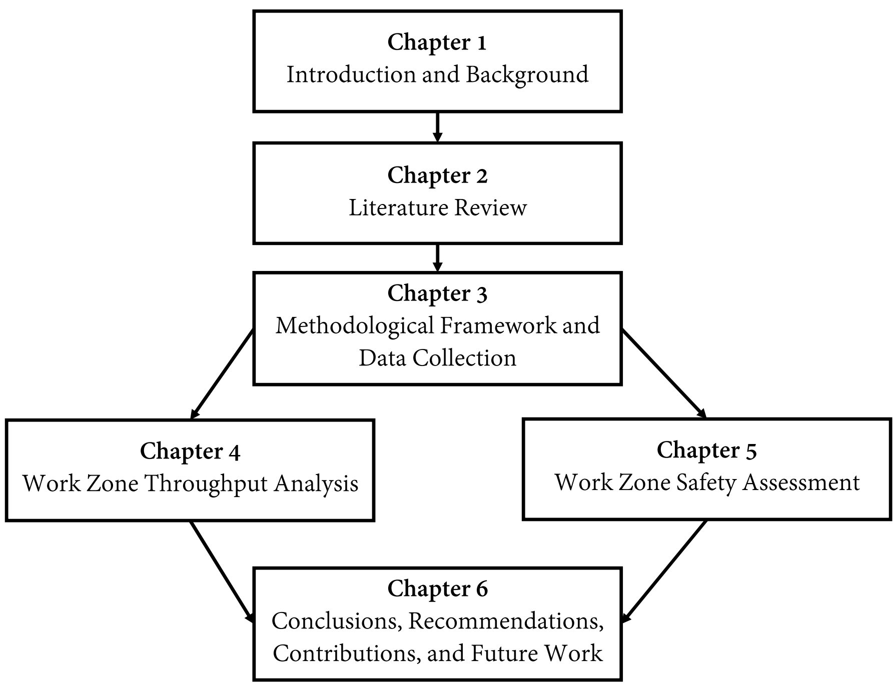


Figure (1.1) Breakdown of Chapters

Chapter 2

Literature Review

This section will present details of the current state of the topics related to the work zone safety and throughput and proposed methodologies of this study.

2.1 Road Classification System

In 1986, the Transportation Association of Canada (TAC) [7] published a manual of geometric design standards for Canadian roads. This manual was the first attempt in Canada which presented a hierarchical categorization of road systems by considering the types and scale of services they offer to the public. This hierarchical categorization in transportation engineering is referred to as "Road Classification System". This fundamental tool could be considered as the first step in urban development and road management. Bringing together roads with similar functional class can have a significant impact on improving transportation planning, road infrastructure design, maintenance, traffic and road operations.

The primary role of classifying the roadways is not only to have better and concise management of the transportation services but to protect society against the adverse impacts of motorized traffic in neighbourhoods. Among various road classes, some allow motorists to drive with higher speed and others, which are the majority of roads, are authorized for lower speeds. This concept could also be expanded in terms of the volume of traffic. Ignoring the hierarchical grouping of roads could beget less efficient transportation network with the associated escalation in the time and cost of public and private transportation, e.g. foot, bike, bus or car, and freights. Besides, eluding the use of this classification could highly worsen the quality of civic life by escalating congestion in the neighbourhoods. Categorizing roads into different groups based on their functionality is one of the most popular ways of road classification systems, and each road would be labelled based on its "Road Functional Class". The city of Toronto divides streets into five groups based on their functionality (not considering public lane-ways),

- Local Roads
  - High accessibility to local residences
  - Low Average Annual Daily Traffic (AADT), less than 2500 veh/day
  - Low speed limit
  - No bus routes (generally)
  - at least side walks
  - Low priority for winter maintenance
  - Truck restriction preferred
- Collector Roads
  - Connection between local residences and traffic movement
  - Low to moderate Average Annual Daily Traffic (AADT) approximately between than 2500 and 8000 veh/day
  - less than 1,500 bus/day
  - both sides have side walks
  - Traffic control devices and sign are required at intersection with arterial roads
  - Medium priority for winter maintenance
  - Truck restriction preferred
- Minor Arterial Roads
  - Mobility and traffic movement are the main goals
  - Moderate Average Annual Day Traffic (AADT) approximately 8,000 to 20,000 veh/day
  - 1,500 to 5,000 bus passenger per day
  - 40 to 60 km/hr speed limit

  - Signalized main intersections (No ”STOP” signs)
  - No truck restrictions
  - High priority on winter maintenance
  - Both sides have side walks

- Major Arterial Roads
  - Mobility and traffic movement are the main goals
  - Access controls are mandatory
  - More than 20,000 veh/hr Average Annual Daily Traffic (AADT)
  - More than 5,000 bus passenger per day
  - High priority on winter maintenance
  - Both sides have side walks

- Expressways/Freeways
  - Mobility and traffic movement are the main goals
  - No property access
  - High speed limits, 80 to 100 km/hr
  - High Average Annual Daily Traffic (AADT), more than 40,000
  - Forbidden for use of pedestrians and cyclists
  - No traffic Signs, Grade-separated intersections
  - High priority on winter maintenance

Federal Highway Administration [8] defined Average Annual Daily Traffic (AADT) as the cumulative volume of vehicle traffic of a road segment in a year divided by 365 days. This measure has been used in several studies in the past, and it simply represents how busy roads are. Figure 2.1 presents some examples for each functional class.

Figure 2.2 demonstrates the relationship between mobility and accessibility based on road functional classes. As was mentioned in each category's specifications, roads with broad accessibility, such as local roads, practically cannot offer high mobility, and vice versa [8].

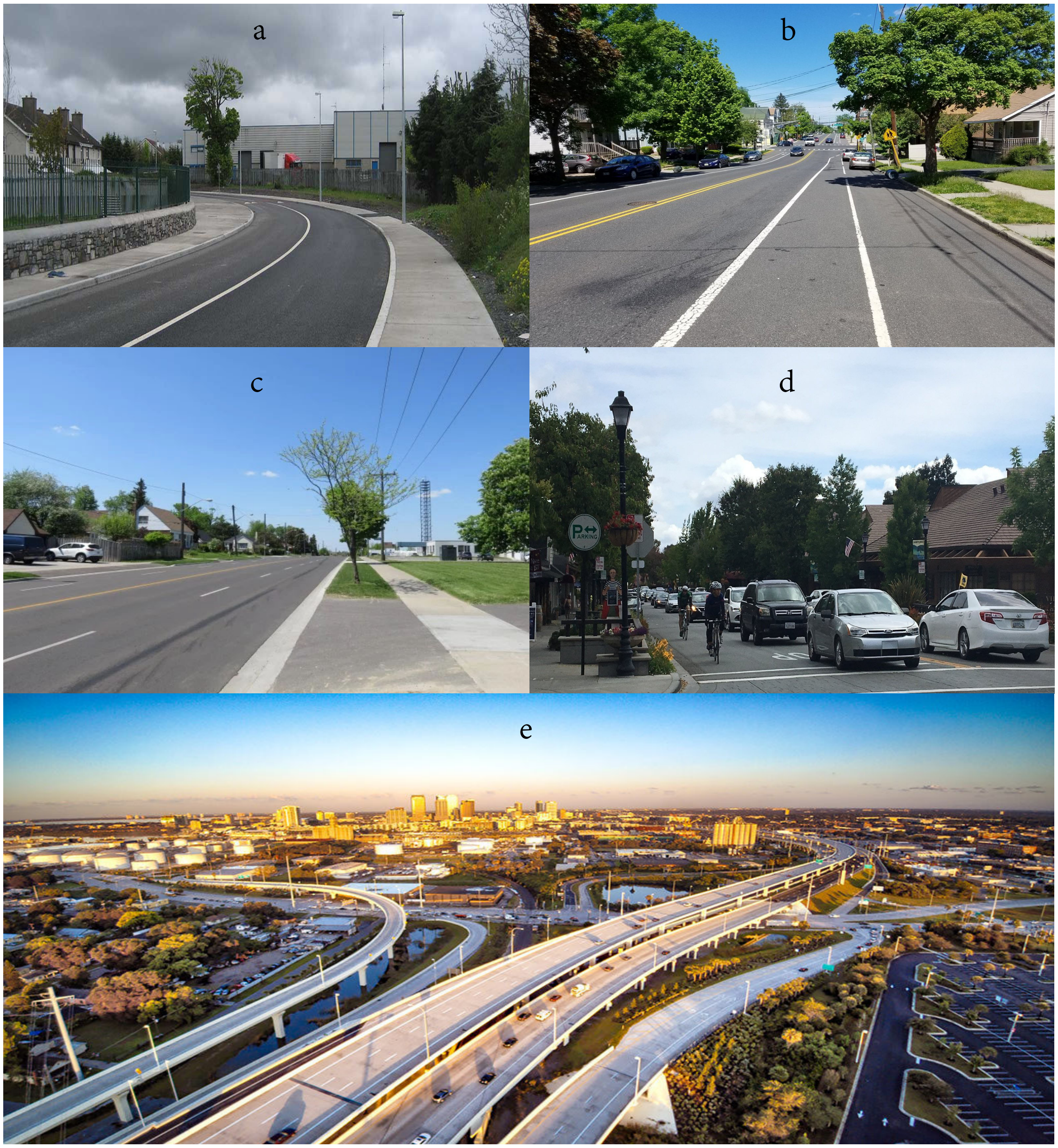


Figure (2.1) Road Functional Classes a: Local Road; b: Collector Road; c: Minor Arterial Road; d: Major Arterial Road; e: Expressway/Highway

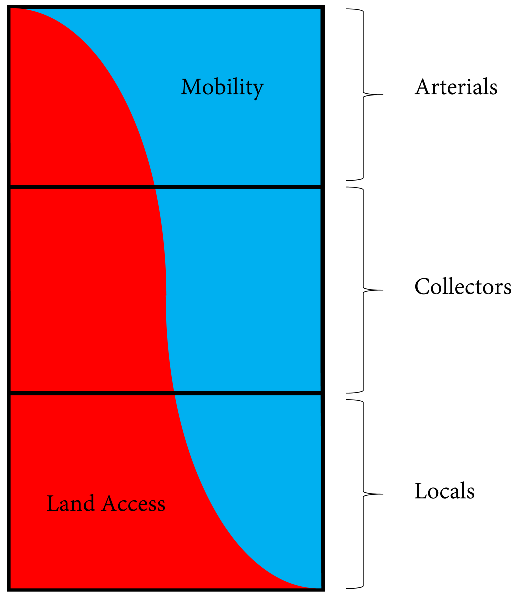


Figure (2.2) Accessibility vs. Mobility for Road Functional Classes

2.2 Maintenance, Preservation, and Rehabilitation of Pavement

One of the goals of this study is to eliminate (or minimize) the observed construction-related safety issues and lessen the delays caused by long queuing in the work zones. These affairs generally increase the user delay cost, number of collisions, and potentially their severity.

A portion of a highway with homogeneous characteristics is referred to as a 'road segment'. In Canada, Due to severe weather condition and the high-volume of heavy trucks on highways (mostly in southern Ontario), some segments could suffer from low serviceability, extensive distress, reduced friction. Various activities could be adapted by decision-makers to improve the infrastructure performance; the most common activities include repair, maintenance, rehabilitation, preservation and replacement of the segments, which each of them will be discussed individually.

Maintenance is an essential activity that ensures the fine physical condition of the roads. In the other words, allocated funds for the maintenance of physical assets will be used to delay or prevent infrastructure failure [9]. All maintenance types —preventive, routine, and corrective —are mostly inexpensive and easy to implement. Agencies generally start routine maintenance after initial construction to postpone deterioration of the infrastructure. Based on the surface type of the highway, there are several activities which can temporarily improve the performance of roadways. The most common activities for roadway maintenance are summarized in Table 2.1 [10]. Also, Figure 2.3 demonstrates some of the maintenance activities presented in Table 2.1.

Table (2.1) Common Maintenance Activities based on the Pavement Surface Type

Surface Type	**Maintenance Activities**
Flexible pavement	Pothole repair
	Shallow patching
	Drainage improvement
Rigid pavement	Partial-depth slab repair with asphalt
	Full-depth slab repair with asphalt
	Drainage improvement
Thin bituminous surfaces	Pothole repair
	Shallow patching
	Drainage improvement
Gravel surfaces	Local grading
	Dust control

Another set of activities which could be utilized to reduce the rate of pavement deterioration in its early ages are pavement preservation activities. Federal Highway Administration (FHWA) in the U.S. defined pavement preservation as a program employing a network level, long-term strategy that enhances pavement performance by using an integrated, cost-effective set of practices that extend pavement life, improve safety and meet motorist expectation [11].

In Canada, usually major maintenance is part of the preservation, and there are some overlaps between these two concepts. Some of the major maintenance activities and minor rehabilitation (non-structural) can be classified as pavement preservation [10] as well. Proactive pavement preservation activities are designed to:

- Prevent premature distresses, and
- Slow down the rate of deterioration

Therefore, pavement preservation is commonly used to extend the pavement's service life. The activities which are mostly known as pavement preservation are listed in Table 2.2. Figures 2.4 demonstrate some of the preservation activities presented in Table 2.2.

Figure (2.3) Roadway Maintenance Activities [1]
a: Pothole Repair on Flexible Pavement; b: Drainage Improvement on Rigid Pavement; c: Partial-depth Slab Repair; d: Shallow Patching on Thin Bituminous Surfaces; e: Dust Control on Gravel Roads

Other than maintenance and preservation, which are mostly focused on non-structural improvements of the pavement, there is another set of activities that generally enhance the structural performance of the pavement; these activities are known as rehabilitation. Rehabilitation activities are also recommended to be used for enhancing the structural

[1] a: http://www.ezstreetasphalt.com/case-studies/pothole-repair/bellmawr-pothole-repair
b: http://www.cedarhilltx.com/1181/Mansfield-Road
c: https://fp2.org/2013/05/21/concrete-repair
d: http://www.watco.co.uk/bitumend-pour-and-restore-black.html
e: http://lymandustcontrol.com/portfolio-item/driveways

Table (2.2) Common Preservation Activities based on the Pavement Surface Type

Surface Type	**Preservation Activities***
Flexible pavement	Crack sealing
	Spray patching
	Seal coat
	Thin asphalt overlay
	Hot in-place recycling
Rigid pavement	Crack and joint sealing
	Diamond grinding
	HMA overlay
	Slab stabilization
	Shot blasting
Thin bituminous surfaces	Spray patching
	Seal coat
	Levelling
Gravel surfaces	Grading
	Drainage improvement

* More activities for each surface type is presented in Pavement Asset Design and Management Guide 2013

strength of the roadways, which were initially designed for lower traffic volumes than the ones they are carrying currently. The rehabilitation process includes a standard procedure, which can be summarized as follows [10],

1. Identify and prioritize the pavement segments with rehabilitation need. Monitoring activities are required to evaluate the functional and structural condition of the segments;

2. Recommend feasible strategies for rehabilitation;

3. Evaluate proposed strategies which results in the most cost-effective options based on specific criteria such as service life, life-cycle costs, and budgetary constraints;

4. Competent measurement of the rehabilitated pavements in order to check the adequacy of performance.

Figure (2.4) Roadway Preservation Activities [2]
a: Crack Sealing on Flexible Pavement; b: Hot in-Place Recycling on Flexible Pavement; c: Diamond Grinding on Rigid Pavement; d: Shot Blasting on Rigid Pavement; e: Spray Patching on Thin Bituminous Surfaces; f: Grading Gravel Road

[2] a: http://www.gardenstatesealing.com/crackandjointsealing/
b: https://canada.constructconnect.com/heat-design-equipment-improves-hot-place-recycling
c: https://en.wikipedia.org/wiki/Diamond-grinding-of-pavement
d: http://www.rovanmachine.com/en/productshow.asp?showid=403
e: https://theasphaltpro.com/articles/how-to-place-an-ultra-thin-lift-pavement-treatment/
f: https://www.youtube.com/watch?v=dz9JyEXfpOI

Some the most common rehabilitation activities are presented in Table 2.3 and Figure 2.5,

Table (2.3) Common Rehabilitation Activities based on the Pavement Surface Type

Surface Type	**Rehabilitation Activities**
Flexible pavement	Resurfacing - structural Cold in-place recycling Bonded concrete overlay Full-depth reclamation
Rigid pavement	Bonded concrete overlay Unbonded concrete overlay Rubblization and resurfacing HMA overlay
Thin bituminous surfaces	Pulverization or scarification and resurfacing Full-depth reclamation
Gravel surfaces	Stabilization Scarification and grading

Figure (2.5) Roadway Rehabilitation Activities [3]
a: Full-Depth Reclamation; b: Rubblization; c: Pulverization; d: Stabilization

[3] a: http://www.midlandasphalt.com/pavement-preservation/fdr/
b: http://www.antigoconstruction.com/
c: https://www.rocksolidstabilization.com/service/asphalt-pulverization/
d: https://www.wirtgen.de/en/news-media/press-releases/article_detail.2564.php

2.3 Work Zones

Work zone could be defined as a section of a roadway with some construction, maintenance, or utility-work activities [12]. Work zones are generally noticeable by their signs, channelling devices, barriers, pavement markings, and work vehicles. Work zones are supposed to begin from the first warning sign or flashing lights on a vehicle to the last sign that shows the 'End of Road Work' sign or the last traffic control device. Based on the type of the work zone and also the type of activities, work zones could be categorized as short duration or long duration, stationary or moving ones.

- Long-term stationary work zones are mostly recommended for highway construction activities such as building a new bridge, adding travel lanes to the roadway,
- Mobile work zones are common for maintenance activities such as striping the roadway, median, and roadside grass mowing/landscaping, and pothole repair.
- Short-term stationary work zones are usually suggested for activities such as repairing electric, gas, or water lines within the roadway.

Traffic management in work zones of high-volume highways could be considered as the most challenging task of the transportation agencies. The most critical concerns that are needed to be addressed are related to the proper traffic management in work zones to reduce the motorists' delay and also reduce the risk of collisions. These problems could be predominantly captured on the segments prior to the work zones in which one or more lanes are dropped. Gipps [13] conducted a study to capture the possible influential parameters on drivers' behavior during the lane changing. He claimed that factors such as the lane selection, the urgency of lane change, entry and departure of non-transit vehicles into and from transit lanes, and heavy vehicles can significantly affect the drivers' lane-changing behavior in the work zones. Moreover, these sections can be described by high traffic turbulence, resulting in enhancing collision risks and longer delay periods [14].

Based on the work zone designs, generally, the likelihood of experiencing conflicts, collisions, and delays is noticeable in the entry section of the work zones. Capacity change and aggressive lane changes to avoid dense traffic because of lane closures are the most significant reasons that negatively affect the traffic performance of the work zones. These sudden manoeuvres at the merging point, as well as the work zone, creates an exceptionally hazardous environment for the motorists and possibly workers in the work zone. Following sections will present some of the methodologies used to overcome the mentioned issues and model the traffic of the roadways and work zones.

2.3.1 Work zone Simulation Approaches

2.3.1.1 Microscopic Simulation Approach

Microscopic simulation approach consists of analyzing the behaviour of every single vehicle as a function of traffic condition in its environment [15] [16] [17]. One of the most well-known microscopic approaches in traffic flow simulation is car-following models. These models are presented with ordinary differential equations, and they aim to describe the dynamics of vehicles' positions and velocities. Generally, three major components are considered in these types of simulations which are,

1. Driving speed
2. Bumper-to-Bumper distance
3. Velocity of leading vehicle

Since the microscopic analysis of the traffic condition considers each vehicle individually in the network, it requires a large quantity of data. This characteristic of the microscopic approach makes it less popular among other approaches.

2.3.1.2 Macroscopic Simulation Approach

In macroscopic simulations, analysis is generally based on the observable factors such as traffic flow, density, and speed. In addition, the aim of the macroscopic models is to evaluate the aggregate behavior of vehicles in the highway network.

Greenshields' macroscopic stream model demonstrates how the change of one traffic flow factor can affect other ones. In this model, the most crucial factors are speed and density. He claimed that the relationship between these factors could be considered as linear. Under this interpretation, considerable reduction of speed causes high density of the vehicles in that specific segment which is known as 'queuing' or 'jam' [18]. This relationship is presented in Equation 2.1,

$$v = u_f - \frac{u_f}{k_j}.k \tag{2.1}$$

where v is the mean speed at density k, u_f is the free-flow speed and k_j is the jam density. Equation 2.1 is commonly known as Greenshields' model. Greenshields' model states that when density is very low (close to zero), the speed could be as high as free-flow speed. Furthermore, the relationship between the three traffic flow variables is given as Equation 2.2,

$$q = u.k \tag{2.2}$$

By combining equation Equation 2.1 and Equation 2.2, the parabolic relationship between density and flow could be derived,

$$q = u_f.k - \frac{u_f}{k_j}.k^2 \tag{2.3}$$

Figures 2.6 is showing how speed changes based on the density of the highway [19]. This figure shows that in order to drive with free-flow speed, the density of the vehicles on the highway should be very low.

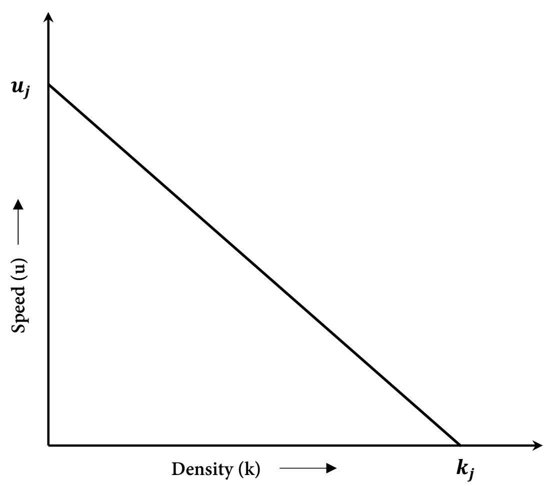


Figure (2.6) Relationship between Speed and Density in Greenshield's model

Figure 2.7 is usually used to estimate the speed at which the optimum flow occurs. In this model, it is assumed that the flow rates are approximately equal in two cases:

1. When the speed is high and the density is lower and;
2. When the speed is low and the density is higher.

These claims could be verified by Equation 2.3 and Figure 2.7 [19].

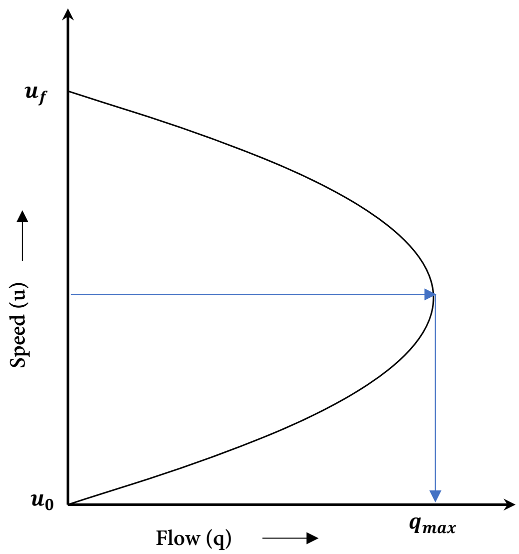


Figure (2.7) Relation between Speed and Flow in Greenshield's model

There are several other macroscopic stream modelling approaches which are assuming a non-linear relationship between speed and density. Table 2.4 describes each method briefly [20]:

Table (2.4) Common Models for Speed and Density Relationship

Models	**Assumed Relationship between Speed and Density**
Greenberg's logarithmic model	$u = u_0 \ . \ ln(k_j/k)$
Underwood's exponential model	$u = u_f \ . \ exp(-k/k_0)$
Pipes' generalized model	$u = u_f \ . \ [1 - (k/k_j)^n]$

All these models are assuming that the relationship between speed and density are constant in all ranges of densities seen in the traffic stream. Thus, these models are commonly known as single regime models. Based on the field observations and previous studies, human behaviour varies, and it is expected to see the discontinuous relationship between speed and density. Multi-Regime models were developed to provide more realistic results based on various traffic conditions; however, analyzing these models is out of the scope of this research.

2.3.2 Statistical Analysis of Work Zones

Although deterministic approaches for work zone performance assessment are vastly used to analyze the traffic condition in these areas, there is always a need to have a detailed perspective regarding the factors that can play a role in the safety and throughput of the work zones. Benefiting from the power of statistical approaches to attain a meaningful trend in historical data could offer this luxury for later assessments. Following sections will discuss studies related to throughput estimation and safety evaluation of highways, respectively.

2.3.2.1 Work zone Throughput Modelling

Recently, several studies were conducted to investigate the factors that can potentially affect the throughput of the work zones. Some of these studies used data from high-volume 400-series highways located in Ontario, Canada [4] [5]. Findings of these studies captured that the presence of law enforcement officers, the number of closed lanes, and type of barriers are the most significant factors for throughput prediction.

Table 2.5 summarizes the mathematical (statistical) models which were developed to estimate the throughput of the work zones. Although econometric and statistical models are highly data-dependent and the magnitude of the coefficients could vary from one model to another, it is expected to capture similar behaviour in some specific factors in most of the models. For example, in most of the studies queuing will increase by closing more than two lanes. Chung [26] have adapted the random parameter modelling to assess the non-recurrent traffic congestion caused by work zones in high-volume highways; he claimed his proposed method could inform travelers to choose other alternatives routes and avoid congestion caused by the work zones. Not many other literature publications were found to use statistical modelling for throughput analysis.

2.3.2.2 Safety Modelling

There have been numerous studies regarding collision analysis, but just a few of them specifically focused on work zones. Predicting accident frequencies [27] or rates [28] and identifying high-risk spots [29] can result in significant improvements in the transportation system; this can eventually help to minimize the number of collisions. To achieve this goal, finding the best models that fit historical data is vital. Past researchers used many modelling methodologies to identify the most influential parameters affecting safety.

Table (2.5) Summary of Previous Work zone Models

Models	Rouphail & Tiwari 1985 [21]	Sarasua et al. 2004 [22]	Krammes & Lopez 1994 [23]	Kim et al. 2000 [24]	Al-Kaisy & Hall 2002 [25]	Hicks 2009 [5]	Mushtaq 2011 [4]
Non-Flagging Site	✓	✓	✓	✓	✓		
Work Activity	✓	✓	✓	✓	✓		
Heavy Vehicle	✓	✓	✓	✓	✓	✓	✓
No. of Open Lanes	✓	✓	✓	✓		✓	✓
Light Conditions					✓		
Lane Width					✓		
Lateral Distance				✓			
Length of Closure			✓				
Driver Population				✓			
Side of Closure				✓	✓		
Ramps			✓		✓		
Weather					✓		
Crossover					✓		
Grade				✓			
Traffic Management		✓	✓				
Day of Week						✓	✓
Barrier Type						✓	✓
Presence of Police			✓		✓	✓	

These parameters are divided into five major groups, namely road geometry (e.g., median width, number of curves, etc.), traffic characteristics (e.g., Average Annual Daily Traffic), pavement condition (e.g., International Roughness Index, Pavement Condition Rating, etc.), weather conditions (e.g., number of rainy days), and human-related factors (e.g., drinking, driving experience). Several statistical methods were also adapted to model the safety-related dependent variables (collision frequency, collision rate, etc.); the nature of the dependent variable could be considered as the primary factor for selection of the most appropriate modelling approach. Accident frequencies can be modelled with Poisson regression [30] [31], or with negative binomial models, depending on the significance of over-dispersion parameter [32] [33] [34] [35] [36] [37] [38]; zero-inflated models can also be used when there is a large number of roadway segments (or locations) with no observed collisions [32] [37] [39] [40]. Researchers such as Maycock and Hall [41] were pioneers in applying count data models such as Poisson and negative binomial regressions to estimate the accident frequency. Around the 1990s, researchers started adapting more advanced statistical models on crash data. They recognized that there were several segments in the roadway (or work zones) without any recorded collisions; this initiated the use of zero-inflated Poisson and negative binomial models. Miaou [32] investigated the relationship between the geometric design of the road segments and truck-related crashes. He also conducted a comprehensive comparison between the results of zero-inflated Poisson and negative binomial models. In the last two decades, more innovative methods with more complicated statistical methodologies have been examined. Bhat et al. [42] developed a new method that accounted for endogenous covariates in count data models. A reliable analytical approximation, as well as the ability to evaluate asymptotic standard errors, were the most highlighted results of his study. Chen [43] investigated the effect of the road geometry factors on various road classes by using bivariate negative binomial models.

In the case of availability of the accident rate data, the Tobit model could be selected as the most appropriate approach. This approach offers the luxury of accounting for the left-censored data [44] [45] [46]. For injury-severity analysis, discrete outcome models are found to be appropriate, such as multinomial logit/probit models [47], ordered logit/probit models [48], and nested logit/probit models [49]. Mannering and Bhat [50] conducted a detailed review of the various statistical modelling approaches in the transportation safety analysis. Their review describes the chronological attempts of researchers in offering the most comprehensive solutions for road safety concerns. Shibata and Fukuda [51] conducted an unconditional multiple logistic regression analysis to investigate the risk factors in fatal collisions. They found that motorcyclist helmet use, using seat belts, and low alcohol level can significantly avoid fatalities in collisions. This research offered a new standpoint toward solving crash-related concerns, and more advanced crash injury severity methodologies have been introduced since then. The findings of this study was later supported by other researchers as well [52]. Several subsets of Logit/ Probit models such as Multinomial Logit [53] [54] [55], Nested Logit [49] [56], Sequential Logit/ Probit [57] [58] have been introduced to crash-injury severity modelling.

Some of the aforementioned modelling approaches can be combined using a multivariate modelling approach [45] [59] [60] [61]. Most of the previous modelling approaches are single response innately; this means that these methodologies could only predict one variable in a run. This limitation of single response models, in some cases, causes a bias resulted by not capturing the effect of other extrinsic factors. The most common multivariate modelling approaches could be listed as,

- Transfer function models
- Vector autoregression (VAR)
- State-space (Kalman Filter) models

Mannering and Bhat [50] published a more detailed review of the use of multivariate modelling approach in transportation safety modelling.

Usually, some of the information needed for modelling is not available or up to date, which causes unobserved heterogeneity issues in the developed models. In literature to avoid unobserved heterogeneity, random parameters modelling approach has been proven to have superior statistical fit in comparison with fixed parameter modelling [35] [55] [62] [63]. The boldest feature of random parameters modelling is its ability to allow the factor alter across the observations; thus, it improves the model's explanatory power. Typically, a random parameters model will yield a larger number of statistically significant explanatory parameters, as compared to its fixed parameters counterpart. Eluru and Bhat [64], as well as Anastasopoulos & Mannering [62], were the innovators who applied the random parameter method on crash-injury severity and accident frequency models, respectively.

Aside from statistical modelling, in the last 50 years, researchers conducted several statistical analyses based on descriptive statistics of the collected data in work zone areas. These analyses can also provide useful insight into the possible influential factors [65]. The summary of these studies is presented in Table 2.6.

Table (2.6) Summary of Work Zone Statistical Analyses

Related Literature	**Work zone Scope**	**WZ sites**	**Crashes**
Nemeth & Migletz 1978 [66]	Construction Zones on Rural Interstate Systems in OH	21	3 years
Hargroves 1981 [67]	Highway Work Zones in VA	statewide	1 year
Rouphail et al. 1988 [68]	Short- And Long-Term Urban Freeway Work Zones in IL	26	6 years
Hall & Lorenz 1989 [69]	Construction-Zone Accidents on Rural State Highways in NM	3	3 years
Pigman & Agent 1990 [70]	Construction and Maintenance Work Zones in KY	20	4 years
Zhao & Garber 2001 [71]	Construction and Maintenance Work Zones in VA Urban Areas	7	4 years
Ha & Nemeth 1995 [72]	Construction and Maintenance Work Zones In OH	60	5 years
Bryden et al. 1998 [73]	Construction Work Zones in NY	statewide	3 years
Daniel et al. 2000 [74]	Work Zones with Fatal Crashes in GA	statewide	3 years
Zhao & Garber 2001 [71]	Work Zones in VA	statewide	4 years
Chambless et al. 2002 [75]	Work Zones in AL, MI, TN	statewide	5,3,2 years

Table 2.6 Cont'd: Summary of Statistical Analyses Studies

Related Literature	**Work zone Scope**	**WZ sites**	**Crashes**
Schrock et al. 2004 [76]	Work Zones with Fatal Crashes in TX	77	15 months
Salem et al. 2006 [77]	Work Zones on Interstate Highways in OH (Fatal & Injury Crashes)	statewide	3 years
Arditi et al. 2007 [78]	Highway Construction Work Zones with Fatal Crashes in IL	statewide	6 years
Ullman et al. 2008 [79]	NY Work Zones & Projects in other CA, NC, OH, WA	NY+64 projects	multi year
Jin et al. 2008 [80]	Road Construction Projects on Different Types of Highways in UT	202	4 years
Li & Bai 2009 [81]	Highway Work Zones with Fatal and Injury Crashes in KS	statewide	13 year
Dissanayake & Akepati 2009 [82]	Work Zones in IA, KS, MO, NE, WI	statewide	5 years
Xing et al. 2010 [83]	Work Zones on an Expressway with Lane Closure in Japan	132	5 years
Ullman et al. 2011 [84]	Highway Work Zones in NY	statewide	6 years
Swansen & Knodler Jr. [85]	Work Zones in the MA	statewide	3 years
Li et al. 2012 [86]	Highway Work Zones in KS	statewide	9 years
El-Rayes et al. 2013 [87]	Work Zones in the IL	statewide	10 years

Crash records could present useful information about the consequences of collisions. These records are generally used to identify crash severity. In the literature, there was no agreement on the severity of the crashes in the work zones. Some researchers believe that work zones cause collisions with higher severity, however, some other researchers claimed that there is no significant difference between the severity of the collisions in the work zones and non-work zone sections of the highways; in other words, they concluded from their studies that injuries and fatalities in work zones do not differ significantly from non-work zone collisions. In Texas, work zones were responsible for 6% of the fatal collisions directly; the indirect effect of work zones was found to be 39% [76]. Also, some other studies reported more severe collisions in the work zones [70] [88]. Based on a survey completed by Benekohal and Shim [89], the majority of 930 truck drivers surveyed perceived that driving through work zones is more hazardous than non-work zone sections of the highways. However, several other researchers concluded that there is no significant difference between work zone crashes and regular crashes in terms of severity [69] [75]. Even some of these researchers believe that work zone collisions are even less severe than non-work zone crashes [67] [68] [71] [74]. Figure 2.8 demonstrates various studies conclusions about the severity of crashes in the work zones.

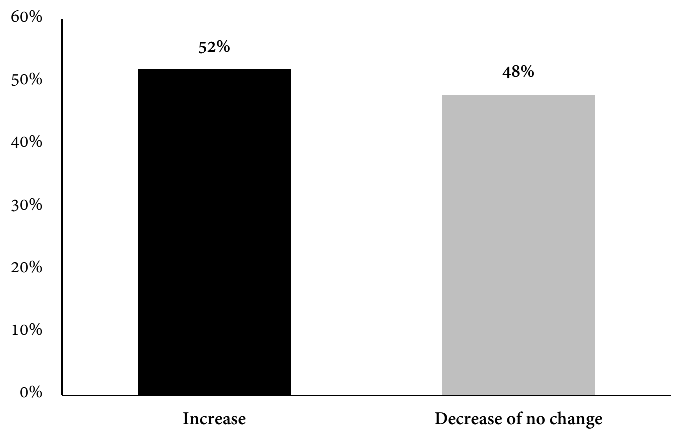


Figure (2.8) Various Conclusions from Literature

Based on the 2017 Transport Canada [6] statistics study: 1,841 fatalities, 9,960 major injury collisions, and 154,886 minor injury collisions occurred on Canadian highways. Generally, work zones located in high-volume highways could be considered as high-risk locations due to more congestion, narrower lanes, etc. In a 2007 report [90] by the same organization concerning the social costs claimed that the exposed cost by motor vehicle collisions consist of highway traffic delays, out-of-pocket expenses, hospital/health care, tow trucks, and Police, Fire and Ambulance Services. Using these factors, the social cost in Ontario for motor vehicle collisions was estimated to be approximately 18 billion dollars. Among all other injury-severity levels, fatalities had the highest share (11 billion dollars). With growth of the traffic volume on Ontario roadways, it is expected that the needs for setting up the work zones to maintain the serviceability of the highways would also increase; and supposedly based on the nature of the work zones (i.e. narrower lanes and temporary changes in the geometry of the highways), it is expected to see more frequent and severe collisions in these vulnerable areas. Correspondingly, the odds to have a positive trend in social costs that is induced by these unpleasant events would be higher as well.

In a study by Khattak et al. [91], it was concluded that the collision rate generally increases in the high-risk work zone sections of highways. Turochy et al. [92] conducted a study to investigate the most influential factors in work zone crashes. They concluded that evening and overnight closures, various manners of collision (head-on, rollover, and angle crashes), and excessive speed significantly affected the injury severity level of the collisions occurring in the work zones of Alabama. Meng and Weng [93] focused on rear-end crashes in work zones, and they stated that rear-end crashes are more likely to occur in expressways compared to arterial roadways. Also, they determined that truck percentage and lane traffic flow are the most influential factors in work zone related crashes. Yang et al. [94] found that the length of the work zones, as well as the traffic volume, increases the likelihood of crash occurrence in the work zones. They also claimed that although there were frequent attempts to determine a comprehensive relationship between work zone crashes and time, weather, and traffic control devices, a meaningful relationship cannot be determined based on the current findings. Lin et al. [95] explored the influence of the speed limit control on the safety of highway work zones. The exciting finding of this research was that not only lower speed improves the safety of the work zones, but also a reduction in the variance of speed significantly improves safety.

Mohan and Gautam [96] categorized the work zone collisions into two groups: worker-based and motorist-based collisions. Based on their analysis, 30% of the recorded work zone collisions workers were involved, and the rest were concerning the second group who were motorists. They also determined that driver error is the main reason for these collisions. An earlier study by Bryden and Andrew [97] deduced that 15 percent of all severe injuries and over 40% of the fatal crashes involved pedestrian workers. Factors that significantly affected the injury-severity of the crashes occurred in the work zone were identified by Wong et al. [98]. They categorized these factors as the location of the accident or work zone, duration of the work zone, time of day and type of activity in the work zone. Another factor that researchers believe that could decrease the crash frequency is the use of traffic control devices in work zones [99]. Although these studies emphasized that using traffic control devices reduce the crash frequency, in some cases, it was found that they can also be hazardous for drivers, as well as pedestrians and workers [91]. This demonstrates that there is not a clear finding that constitutes the effective use of these devices, and further studies could be undertaken to draw more robust conclusions about their effects [100] [73]. Hall and Lorenz [69] analyzed before and during-construction period of more than 1500 construction zones and they captured that rear-end collisions significantly increased due to construction zones. They recognized that almost 36 percent of the collisions in the work zones were rear-ended. However, they also concluded that other crash types were less frequent in these zones.

2.4 Summary

This chapter evaluated the literature concerning several concepts that will be discussed and analyzed in this research. Initially, the fundamentals related to pavement deterioration and the methods that are commonly used in pavement and transportation engineering to avoid or minimize them are discussed. It was identified that maintenance and preservation activities are usually appropriate for non-structural improvements; however, rehabilitation activities are more aiming to tackle the structural issues of the infrastructures. The necessity of maintaining roads and bridges encourages agencies such as Ministry of Transportation Ontario to set up frequent work zones. It is vital to note that although work zones improve the quality of the ride, and sustains the serviceability of the infrastructures, the presence of them on roads, especially high-volume highways, cause traffic interruption; the most important consequences of setting up work zones are identified as queuing and collisions.

Several researchers attempted to propose solutions for congestion imposed by work zones in the highways. Microscopic simulation methods used the driving speed, bumper to bumper distance and velocity of the vehicles to model the traffic condition in its environment; however, its dependency on large databases made it less popular throughout the time. Macroscopic simulation methods intended to find a relationship between velocity, capacity, and density. Greenshield's model could be identified as the most traditional model in this area. Greenberg, Underwood, and Pipes offered more complicated and advanced version of these models. Several years later in 2007, a fixed parameter linear regression model was developed by Hicks [5] to find a relationship between work zone configuration and frequency of passing vehicles in high-volume highways. The major concerns regarding this model were that, first, this model was not taking account the nature of the data, and, second, it had limited number of observation. Therefore, the author has been inspired to investigate these gaps and propose a model which could address the earlier concerns.

Apart from the work zone congestion, the former researchers endeavoured to construct statistical models to identify the factors that can significantly enhance the safety of the roads. Accident frequency models, as well as, accident rate and injury severity models were developed to tackle this issue. Factors related to the road geometry, pavement condition, and traffic characteristic were found to be the most crucial factors in this area; one of the major concerns that was barely addressed in this field was related to the unobserved heterogeneity issue. The second motivation of the author was initiated by considering this gap. Also, in random parameter modelling, the coefficients of the random factor alters across the observations. Traditionally, researchers were not considering the luxury of having multiple coefficients for single factor by using the average of it. Therefore, author proposed a methodology which can take advantage of having multiple coefficients for various scenarios. It is believed that applying the recommended methodology improves the prediction perfromance of the models.

Chapter 3

Methodological Framework

3.1 Statistical and Econometric Modelling Framework

This section presents an overview of the methodologies used for the statistical modelling of the work zone throughput and injury severity analysis. Multiple Linear regression, and Negative Binomial Regression will be used to model the work zone throughput; Ordered probability models (Logit, Probit, and Arctangent) are used to model injury severity of the collisions in the work zones. After the modelling stage, the developed models will be used as the basis of performance evaluation of work zones. These models will offer a comprehensive insight about the possible factors which can improve the safety of the work zones and also can lead to concise suggestions for future construction zones. Figure 3.1 illustrates the breakdown of the recommended methodology for this study.

3.1.1 Multiple Linear Regression (MLR)

Linear Regression is one of the most common statistical and econometric techniques. Two primary reasons made this tool popular,

1. Linear regression is appropriate for recognizing the relationship between a wide range of variables.
2. The assumptions of this tool, which will be discussed later, are satisfying many practical applications and make this method very easy to use and understand.

Linear regression generally serves as an excellent starting point for demonstrating statistical model estimation procedures. Although this method is easy to understand and interpret, using this method while there are other appropriate models is not recommended. Multiple linear regression could be defined as a model which includes at least two independent variables. These models are standard for continuous dependent variables [101].

3.1.1.1 Linear Regression Model Assumptions

Multiple linear regression could present the linear relationship between several independent variables and a continuous dependent variable. Most regressions try to detect a list of explanatory variables that are significantly affecting the dependent variable.

In general, explanatory models are based on data achieved from well-controlled experimentation in the laboratory; analytical models are based on data obtained from observational studies such as data collected from work zones. There are also some other models such as quality control models which are out of the scope of this study. It should be noted that the dependent variable could change because of some unknown factors other than the already existing explanatory variables in the model; therefore, there is always a need for uncovering these hidden factors. Some assumptions should be considered to observe the direct effect of the independent variables [101]. When any of the requirements is not satisfied, corrective actions should be considered, and in some cases, alternative modelling approaches should be tried. The followings are the assumptions of the linear regression model,

1. Continuous Dependent Variable (Y)
2. Linear in Parameters (Relationship between Y and X)
3. Observations Independently and Randomly Sampled
4. Uncertain Relationship between Variables
5. Disturbance Term Independent of X and Expected Value Zero
6. Disturbance Terms not Auto-correlated
7. Regressors and Disturbance Uncorrelated
8. Disturbances Approximately Normally Distributed

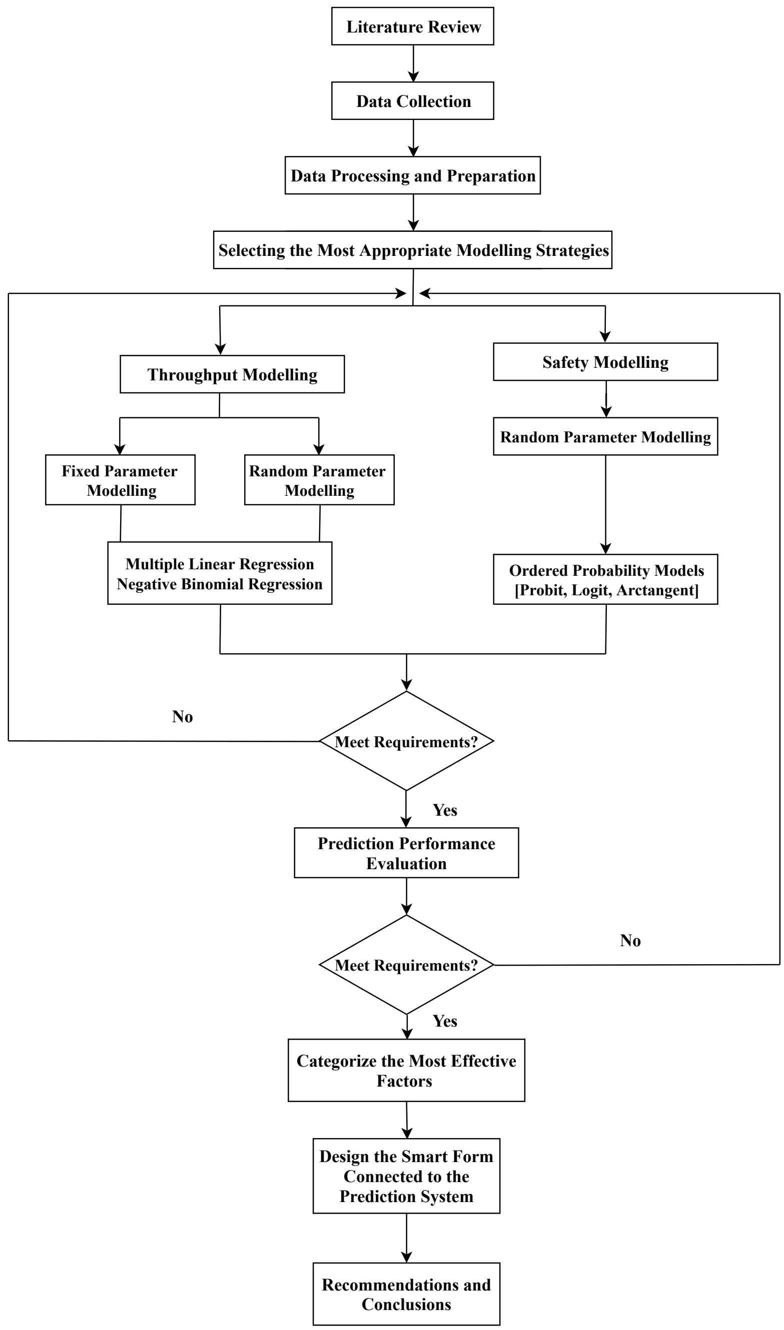


Figure (3.1) Research Methodology

3.1.1.2 Regression Fundamentals

The main goal of the regression is to state the relationship between dependent variable Y with at least one independent variable. Generally, these regression models aim to present some practical information and properties of the population by only reviewing the properties of the sample.

The matrix expression of the linear regression model can be presented as follows,

$$Y = X\beta + \varepsilon \tag{3.1}$$

$$\varepsilon = Y - \widehat{Y} = Y - X\hat{\beta} \tag{3.2}$$

where X is properties matrix, β is the factors coefficient matrix, Y is the dependent variable, $\widehat{Y}$ is the prediction matrix, and ε is the difference between actual and predicted values.

One of the most common ways to solve the above problem is the Least Squares Method. This method is also known as "Ordinary least squares" or "OLS". In this method, the goal is to estimate the model parameters by using the sample data. To that end, the following formula could be used to estimate these parameters,

$$\beta = [(X^T X)]^{-1} X^T Y \tag{3.3}$$

Equation 3.3 presents the estimation of model parameters. The next step is to check whether the estimated factors are statistically significant or not. There are three common methods for this purpose:

1. Confidence Interval
2. Hypothesis Testing
3. P-value Estimation

The decisions based on all these methods should be in-line with each other. In this thesis, the confidence interval and p-value approaches were mostly used for final decisions.

3.1.1.3 Goodness of Fit for Linear Regression

The aim of calculation of goodness-of-fit (GOF) statistics is to have a better insight in comparison procedure of various studies. R-squared and adjusted R-squared are generally reported in many of the previous studies. The following notions should be calculated to estimate the R-squared statistics [102],

1. Sum of Square Errors (SSE), which shows the variation of the fitted regression line around the observations.

$$SSE = \sum_{i=1}^{n} (Y_i - \widehat{Y_i})^2 \tag{3.4}$$

2. Regression Sum of Squares (SSR), which presents the variation of the fitted regression around the $\bar{Y}$.

$$SSR = \sum_{i=1}^{n} (\widehat{Y_i} - \bar{Y_i})^2 \tag{3.5}$$

3. Total Sum of Squares (SST), which represents the total variation of each observation around the $\bar{Y}$.

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y_i})^2 \tag{3.6}$$

Based on the above information, R-squared could be defined as,

$$R^2 = \frac{SST - SSE}{SST} = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{3.7}$$

R-squared could vary from 0 to 1. R-squared zero shows that there is not any relationship between X and Y; however, r-squared one means that the estimated model explains all the variance. Adjusted R-squared accounts for the degree of freedom changes as a result of various numbers of factors in the model. This statistic can be calculated as equation 3.8,

$$R^2_{Adjusted} = 1 - \frac{n-1}{n-p}\frac{SSE}{SST} \tag{3.8}$$

where, n is the number of observations and p is the number of factors or model parameters.

3.1.2 Count Data Modelling- Poisson and Negative Binomial Regression

Count data usually refers to the non-negative integer values, which is frequently observed in transportation-related phenomena. Some of the most common count data variables in transportation can be listed as the number of vehicles waiting in the queue, the number of accidents per road segment, etc. This method meets the minimum requirement of having integers as the outcomes of the modelling process [101].

Several methods, such as Poisson Regression and Negative Binomial Regression, can be used to model the count data. The most common method among all methods is the Poisson regression model. One of the requirements of using this model is that the ratio of mean to variance should be approximately 1 (i.e. they should not be statistically different from each other). If the mean and variance are significantly different from each other, it could be concluded that over-dispersion exists in the data and developing a Negative Binomial regression model can highly improve the model results and performance.

3.1.2.1 Maximum Likelihood Estimation (MLE)

Maximum Likelihood Estimation (MLE) is another popular and long-established statistical estimation method. This estimation is based on Equation 3.9 [101],

$$f(x_1, x_2, x_3, \cdots, x_n, \beta) = \prod_{i=1}^{n} f(x_i, \beta) = L = (\beta|X) \tag{3.9}$$

This equation could be reformed based on n independent and normally distributed disturbances as Equation 3.10,

$$L = (2\pi\sigma^2)^{-\frac{n}{2}} EXP[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - X_i^T\beta)^2] \tag{3.10}$$

The logarithm of Equation 3.10 is known as log-likelihood function. Generally, solving log-likelihood function is considered more straightforward in comparison with likelihood function itself. This function is shown in Equation 3.11,

$$LN(L) = LL = -\frac{n}{2}LN(2\pi) - \frac{n}{2}LN(\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta) \tag{3.11}$$

Maximization of the log-likelihood function based on β and σ^2 leads to similar betas from OLS coefficient estimation approach. The most notable advantage of MLE estimation is its consistency and accuracy in large sample sizes. Other than that, this method is more suitable to use when the dependent variable is not normally distributed.

3.1.2.2 Poisson Regression Model

One of the goals of this study is to determine the number of vehicles passing the work zone during the queuing time. Equation 3.12 presents the Poisson Regression model, where the probability of work zone i having y_i throughput per hour is:

$$P(y_i) = \frac{exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \tag{3.12}$$

where P(y_i) is the probability of work zone i having y_i throughput per hour and λ_i is the Poisson parameter of work zone i, which is equal to work zone i's expected number of throughput per hour, E[y_i]. The relationship between independent (explanatory) variables with the Poisson parameter can be presented as Equation 3.13.

$$\lambda_i = exp(\beta X_i) \quad or, equivalently \quad Ln(\lambda_i) = \beta X_i \tag{3.13}$$

X_is are the vector of the statistically significant explanatory variables directly derived from the collected data set; β is the vector of the estimated parameters.

3.1.2.3 Negative Binomial Regression Model

Sometimes the data do not meet the requirement of equality of mean and variance to allow the use of Poisson Regression. Negative Binomial Regression is one of the most popular models that can account for this inequality. The Negative Binomial model 3.14 has a similar formulation to the Poisson model, and it only counts for error term (ϵ):

$$\lambda_i = exp(\beta X_i + \epsilon_i) \quad or, equivalently \quad Ln(\lambda_i) = \beta X_i + \epsilon_i \tag{3.14}$$

where exp(ϵ_i) is Gamma-distributed disturbance term with the mean 1 and variance α. One of the methods to estimate the accuracy of the Poisson and Negative Binomial models is McFadden ρ^2. This statistic is equivalent to R^2 in linear regression and can be calculated as Equation 3.15.

$$\rho^2 = 1 - \frac{LL(\beta)}{LL(0)} \tag{3.15}$$

where LL(β) is the log-likelihood at convergence with parameter vector β, and LL(0) is the log-likelihood at zero. A perfect model has ρ^2 equal to 1; a model with ρ^2 equal to zero is considered flawed [101].

Similarly to adjusted R-squared, adjusted McFadden ρ^2 [103] could be calculated by accounting for the number of factors that have been used in the modelling procedure. To that end, adjusted McFadden ρ^2 could be achieved by using Equation 3.16,

$$\rho^2 = 1 - \frac{LL(\beta) - k}{LL(0)} \tag{3.16}$$

where k is the number of statistically significant factors in the model, and the rest of the factors are presented in McFadden ρ^2's description.

3.1.3 Ordered Probability Models

There are many cases in the transportation safety field, where researchers have used discrete ordered data as a dependent variable. Results of questionnaires with qualitative scales, hierarchical opinions (such as agree, neutral, disagree), or categorical frequency data (such as KABCO classifications) are some examples for these types of data. In the mid-70s, ordered probability models were introduced in the transportation field; the main idea of this methodology depends on a variable referred to as z, which estimates the ordinal ranking of the data. This unobserved variable (z) could be formulated as Equation 3.17:

$$z = \beta X + \epsilon \tag{3.17}$$

In this equation X represents the vector of significant independent variables, β is the set of estimated coefficients for each independent variable, and ϵ is the random disturbance [101].

In Equation 3.18, the dependent variable (y) is categorized hierarchically as 1,2, 3, ..., k; then, predictions could be made by comparing the estimated thresholds with z,

$$
\begin{aligned}
&y = 1 \quad if \quad z \le \mu_0 \\
&y = 2 \quad if \quad \mu_0 < z \le \mu_1 \\
&... \\
&y = k \quad if \quad z \ge \mu_{k-1}
\end{aligned}
\tag{3.18}
$$

The probability of each category could be calculated as Equation 3.19,

$$
\begin{aligned}
P(y = 1) &= F(-\beta X) \\
P(y = 2) &= F(\mu_1 - \beta X) - F(-\beta X) \\
P(y = 3) &= F(\mu_2 - \beta X) - F(\mu_1 - \beta X) \\
&... \\
P(y = k) &= 1 - F(\mu_{k-1} - \beta X)
\end{aligned}
\tag{3.19}
$$

F(z) is the cumulative distribution function. Any changes to the type of the cumulative distribution function., e.g. Probit, Logit, and Arctangent, could alter the estimation procedure and affect the outcomes. Note that the first threshold in these types of estimations is generally considered as zero. Consequently, Equation 3.20 is the general equation for calculating the probabilities:

$$
P(y = i) = F(\mu_i - \beta X) - F(\mu_{i+1} - \beta X) \tag{3.20}
$$

Also, based on the underlying mathematical structure, the likelihood and log-likelihood functions could be formed as Equations 3.21 and 3.22, respectively,

$$
L(y \mid \beta_1, \cdots, \beta_s, \mu_1, \cdots, \mu_{k-1}) = \prod_{n=1}^{N} \prod_{i=1}^{k} [F(\mu_i - \beta X_n) - F(\mu_{i+1} - \beta X_n)]^{\delta_{in}} \tag{3.21}
$$

$$
LL(y \mid \beta_1, \cdots, \beta_s, \mu_1, \cdots, \mu_{k-1}) = \sum_{n=1}^{N} \sum_{i=1}^{k} \delta_{in} Ln[F(\mu_i - \beta X_n) - F(\mu_{i+1} - \beta X_n)] \tag{3.22}
$$

N is the number of observations and δ_{in} is equal to 1 for observation n if the observed outcome is i, and it is equal to 0 otherwise. The most crucial assumption in this estimation procedure is $\mu 0 \leq \mu_1 \leq \cdots \leq \mu_{k-1}$. One of the concerns associated with ordered probability models is that βs cannot give a full picture of their effect on each category. The positive sign of the βs shows that it increases the probability of the very high category and reduces the probability of the first category, but it does not represent details about the changes in intermediate categories. It is recommended to calculate the marginal effect, which could be estimated as Equation 3.23, to have a better insight into the effect of each β on the interior categories [101],

$$\frac{\partial(P(y=i))}{\partial x} = [f(\mu_i - \beta X) f(\mu_{i+1} - \beta X)]\beta' \tag{3.23}$$

where f(z) represents the standard density function.

As previously mentioned, several F(z) cumulative functions could be adapted for probability estimation. Three of the cumulative density functions that are used in this study are Probit, Logit, and Arctangent; their respective f(z)s and F(z)s are presented in Table 3.1. In the ordered Logit model, it is assumed that ϵ has a standard logistic distribution instead of a normal distribution, which is the assumption of the Probit model. The ordered Arctangent model is generally used to capture any asymmetric distributions [104].

Table (3.1) List of PDFs and CDFs *

Probit	**Logit**	**Arctangent**
$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} = \phi(z)$	$f(z) = \Lambda(z)[1-\Lambda(z)]$	$f(z) = \frac{2}{\pi}\frac{1}{1+z^2} = g(z)$
$F(z) = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx = \Phi(z)$	$F(z) = \frac{e^z}{1+e^z} = \Lambda(z)$	$F(z) = \frac{2}{\pi} arctan(z) = G(z)$

* f(z) and F(z) are representing Probability Density Function and Cumulative Density Function, respectively

3.1.4 Random Parameter Modelling

The modelling approaches that were presented earlier in this chapter do not offer flexibility to coefficients to vary across the observations. In other words, the coefficient would be constant for all the scenarios. However, this fixed-parameter assumption might not be correct in all cases. As an example, the effect of the paved and wide shoulders on the number of collisions could be divided into two categories. Some researchers believe that the paved shoulders are more convenient; however, sometimes continuity of the paved shoulders for long-distance might confuse the motorists in a way that they assume the paved shoulder as an extra lane (especially at night time). This might lead motorists to make sudden lane changing manoeuvres when shoulder ends or becomes narrower, which potentially can cause conflicts or collisions. However, wide and paved shoulders are also found to be very critical and beneficial in many other cases. This explains that there might be some other underlying factors which are relevant to the width and type of shoulder. These hidden factors might not be observed, or data related to those were not collected; this phenomenon is known as unobserved heterogeneity. Random parameter models account for this issue and allow coefficients to change across the observations. This is a significant consideration as if unobserved heterogeneity is not addressed, the parameter estimates could be incorrect, the inferences unreliable, and the predictions inaccurate [55] [62] [101]. General formulation of random parameter modelling is shown in Equation 3.24,

$$g(y_{it} \mid x_{it}, z_i) = f(y_{it}, x_{it}, z_i, a_i) \tag{3.24}$$

where f is the density function for the i^{th} individual observed dependent variable at time t, y_{it} is the i^{th} individual dependent variable at time t, x_{it} and z_i are measured covariates, and a_i is a factor specific parameter vector which alters randomly across each observations, with mean α and covariance matrix Ω.

Random parameter structure could be revised based on the model structure. For ordered probability models, random parameter formulation is,

$$Prob[y_{it} = j - [x_{it}, \beta_i] = F(y_{it} = j, \mu, \beta_i^{'} x_{it} + a_i), i = 1, 2, , N, t = 1, 2, , T_i \tag{3.25}$$

where F could be any of the mentioned distributions (Logit, Probit, Arctangent). The assumption is that parameters are randomly distributed with heterogeneous parameters generated by Equation 3.26,

$$E[\beta_i \mid z_i] = \beta + \Delta z \tag{3.26}$$

Finally, the model could generate the coefficients using Equation 3.27,

$$\beta_i = \beta + \Delta z_i + \Gamma v_i \tag{3.27}$$

where β is the fixed constant terms in the means of the distributions for the random parameters, Δ is the coefficient matrix, z_i is a set of observed variables which are not altered by time and the means of the random parameters, Γ is a lower triangular matrix which produces the covariance matrix of the random parameters, and finally v_i is the unobservable latent random term in the i^{th} observation in β_i.

A simulation-based maximum likelihood estimation method was used to simplify the additional complexity due to the presence of random parameters in the calculation of outcome probabilities. To achieve simulated probabilities, numerous iterations of the same procedure should be conducted [101].

One of the major concerns in random parameter models is the sampling technique of β values from the selected distribution. In order to solve this issue, the technique of Halton draws [105] is chosen. This method can provide an appropriate context to efficiently approximate probabilities by limiting the number of draws to the smallest possible extent. Moreover, the non-random selection of draws provides more robust estimates, as compared to the use of merely random draws [106] [107] for further details on the simulation approach. Based on the literature, selecting 200 Halton draws can yield a stable and accurate estimation of parameters [63] [106] [108]. For the functional form of the parameter density function several distributions, such as Normal, Weibull, Log-Normal, Uniform, and Triangular distributions, will be considered and results will be compared in order to choose the best candidate in terms of statistical fit. A parameter will be considered as random when the mean and standard deviation of the parameter density function is statistically significant, at the 95% confidence level.

3.2 Data Collection

This section presents the data collection procedure and details about sites that data were collected from.

3.2.1 Data Collection for Work Zone Throughput

The main idea of estimating the throughput in the work zones was started in 2007. After the first phase of data collection, more data was collected in 2011 to have a more robust database. Finally, in 2016, [4] [5], more data was collected from various 400-series highways to provide an appropriate context for statistical modelling. The data collection methodology was consistent over the three data collection time frames. The site criteria to trigger data collection were the same for all phases of the study. For data collection purposes, it was decided to start recording the queuing data in the construction zone when the traffic speed was 10-15 km/hr or less. Two criteria need to be satisfied for a site to be qualified for data collection,

- Reduction of at least one lane in the highway [but not complete highway closure]
- Observable queuing beyond the taper line. Traffic needed to be totally stopped or very slow with the maximum observed speed of 10-15 km/hr

The first step of the data collection process involved coordination with the Contract Administrator (CA) of the respective projects to determine whether the two characteristics were satisfied or not. Once a site was deemed appropriate for data collection, a site visit was scheduled. Safety was one of the major concerns at the time of field visit. In some cases, the contractor did not allow the research team to be on the site due to lack of a safe vantage point for the researchers. In those cases, the research team found a public location from where a clear view of the site was available. The data collection process comprises two parts, site observations and traffic data collection.

3.2.1.1 Site Observations

The information from site observations could be set as parameters for the throughput modelling procedure. A site characteristics form, which was designed and used in early phases, was used in the 2016 data collection process as well. Factors that were documented on the form included:

- Date, time, and location of site visit
- Weather
- Facility type
- Speed limit (posted and temporary in km/h)
- Roadway geometry (i.e. proximity to curve, estimated lane widths, roadway grade, etc.)
- Length of work zone
- Duration of closure
- Proximity to intersections/ramps

- Lane configuration (i.e. number of travel lanes, number of closed lanes, etc.)
- Direction of traffic
- Presence of Ontario Provincial Police (OPP)
- Type of traffic control (i.e. traffic barrels, concrete barriers, signage, etc.)
- Pavement conditions

3.2.1.2 Traffic Data Collection

Since queuing happens gradually, it was crucial to have a good judgment to find the optimum time to start the data collection procedure. Improper timing for data collection could result in biased and incorrect results. There were several site visits where no queuing was observed. Lower than expected traffic for that specific night was one of the major causes of not observing any congestion at those site. Several reasons such as long delays in closing the lanes, and poor weather conditions prevented the contractors to set up the work zones in some of the other nights that the data was not collected.

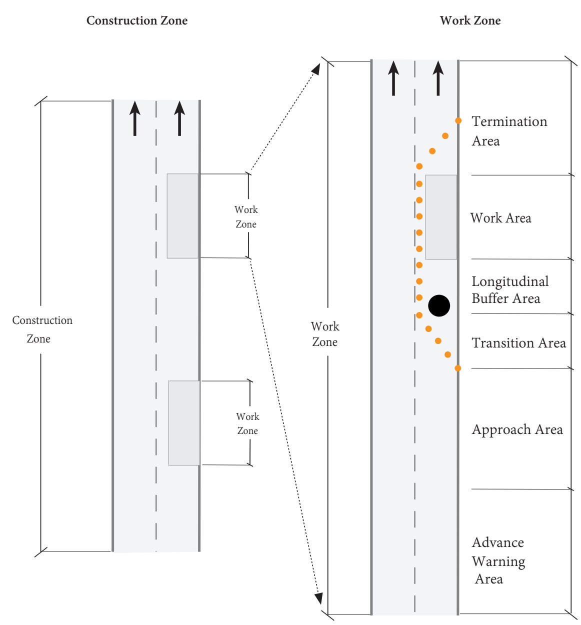


Figure (3.2) Work Zone Configuration and Data Collection Point (OTC 2014)

During data collection, vehicle counts were split into three categories: passenger vehicles, mid-size heavy vehicles, and heavy vehicles. On fifteen-minute intervals, the number of vehicles passing through the work zones was recorded. In addition to the throughput records being collected, videos and still images were recorded to validate any anomalies and allow further research investigation by reviewing the videos. As an illustration, Figure 3.2 shows a general overview of a construction work zone when a two-lane highway is converted into one lane with the use of barrels. The black circle represents the tapper point, where the research team was located for data collection (in most instances). From this point, the dispersion of the queue could be seen, triggering the end of data collection.

3.2.1.3 Identification of New Sites - 2016 Data Collection

The MTO "Central Region Rehabilitation Program" provided the schedule and characteristics of the potential work zone projects on 400-series highways, as presented in Table Appendix A. The provided list includes various information such as the location, length, and type of project. This list also presents the possible operational duration of each work zone. In addition, this detailed program helped the team to identify the potential data collection sites. Some of the projects on 401 highway were more than 35 km long.

3.2.1.4 Visited Sites

Although the Ministry identified thirty-six construction sites as potential sites for data collection, only 9 of the sites were visited. The rest of the sites could not be visited due to:

- No queuing reported by the Construction Administrators;
- No lane closures; or
- Cancellation of construction work due to weather conditions.

In four of the visited sites, data was not collected due to the free-flow of traffic at the work zone at the time of the visit. The detailed list of the visited sites is presented in Table 3.2. This table presents information such as contract number, location, comments related to the queuing condition and location of investigators at the time of data collection procedure.

Table (3.2) List of Sites Visited for Data Collection

Contract No.	Location	Comments	Standing Location
2014-2038	Hwy. 427/Hwy. 401 Interchange - Various Bridge Rehabs.	Queued	Inside of WZ
2014-2026	Hwy 410 - 401 to Queen Street	Queued	Inside of WZ
2014-2044	Hwy. 401 Collectors Warden to Bayview, Four Bridge Rehab. And Collector Resurfacing	Queued	Inside of WZ
2015-2035	Hwy 401 WB Collectors from East of Bayview to East of Allen Road	Queued	Inside of WZ
2015-2021	Resurfacing and ATMS, Hwy. 400 (Hwy. 401 to RR #7)	No queuing	Inside of WZ
2016-2015	Hwy 404 - Noise Barrier Wall and Mount Albert Rehabilitation	No queuing	Inside of WZ
2015-2018	Hwy 401- Widening from East of McLaughlin to East of Credit River	Queued	Outside of WZ
2013-2036	Precast Concrete Pavement Slab repair on Hwy 400 N	No queuing	Inside of WZ
2013-2014	Installing Concrete barrier on 401 E from Allen Rd to Avenue Rd	No queuing	Inside of WZ

Table (3.3) Traffic Throughput Mean and Standard Deviation

Site ID	Location	Data Recorded (hrs)	Throughput (vphpl)	
			Mean	St. Dev
2014-2038	Hwy. 427/Hwy. 401 Interchange - Various Bridge Rehabs.	0.5	1372	28
		1.5	1144	38
		0.5	988	122
2014-2026	Hwy 410 - 401 to Queen Street	0.5	1240	68
		0.5	1008	23
		1.0	796	240
2014-2044	Hwy. 401 Collectors Warden to Bayview, Four Bridge Rehab and Collector Resurfacing	0.25	1076	N/A
		0.5	1652	238
2014-2035	Hwy 401 WB Collectors from East of Bayview to East of Allen Road	0.75	1296	60
		1.25	760	30
		0.75	1004	557
		0.75	1200	123
2015-2018	Hwy 401- Widening from East of McLaughlin to East of Credit River	0.5	1120	119

The hourly throughput frequency was calculated for the sites that queuing was observed, based on the collected data. Table 3.3 summarizes the average throughput volumes in terms of the vehicle per hour per lane (vphpl) and the associated standard deviation for each visit. The results show that there are some variations in throughput in each site. The throughput range varies from 760 vphpl to 1372 vphpl.

3.2.1.5 Descriptive Statistics

Figure 3.3 presents a summary of the descriptive statistics of the work zone throughput data set. Based on the available data set, it was captured that police officials were not present in the work zones in 91% of the closures. Also, 83% of the data set was collected during weekdays. Most of the data (around 88%) were collected at night time, which was expected. In 94% of the work zones, barrels were used, and in only 6% of the construction areas, concrete barriers were used for road closures. Right and left lanes were closed almost equally in the visited work zones. In addition, in 73% of the construction sites, 2 or more lanes were closed to conduct the required maintenance, rehabilitation, and reconstruction activities. Among all visited construction sites, 401 and Queen Elizabeth Way (QEW) were visited the most.

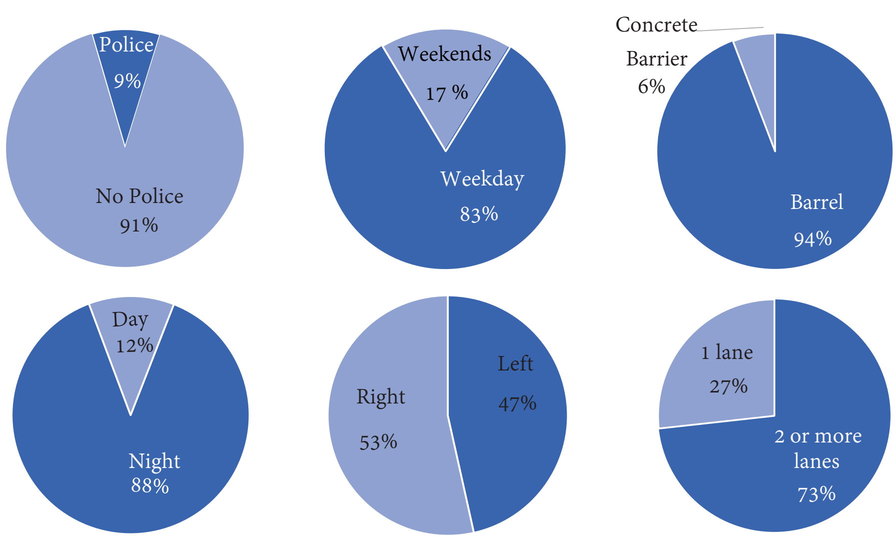


Figure (3.3) Summary of Descriptive Statistics of Throughput Data set

3.2.2 Data Collection for Safety Evaluation of Work Zones

In addition to the data collected for throughput analysis, this research needed additional data to evaluate the safety of the work zones. Ideally, access to the following list of data could ensure the robustness of the models, which will be discussed later in the results chapter. For safety evaluation of the work zones, the following factors were commonly used in various studies that were mentioned before,

- Specific origin of work zone contract (location)
- Human related factors
  - Age of driver
  - Gender of the driver
- Type of activity in the work zone
  - Resurfacing
  - Pavement Rehabilitation
  - Bridge Rehabilitation
  - Deck Rehabilitation
- Work zone characteristics
  - Duration of the Work zone
  - Length of the Work zone
  - Urban/Rural
  - Work zone Cost and Time Characteristics
  - Final Cost
  - Time Delays
- Geometric characteristics of the work zone area
  - Presence of Horizontal and Vertical Curve
  - Number of Lanes
  - Barrier Type (Guardrail, Concrete, Barrel, etc.)
  - Presence of Median
  - Median Width
  - Presence of Inside/Outside Shoulder

  - Shoulder Width

- Traffic characteristics of the work zone area
  - Average Annual Daily Traffic (AADT)
  - Equivalent Single Axle Load (ESAL)
  - Number of Throughput
  - Truck Percentage

- Pavement condition of the work zone area
  - Rutting Depth
  - International Roughness Index (IRI)
  - Pavement Condition Rating (PCR)

Gathering data was one of the most challenging, time-consuming, and complicated steps of this study. To that end, several methodologies have been adapted, and each of those will be discussed individually.

3.2.2.1 Transport Canada and Ministry of Transportation of Ontario Databases

Transport Canada needs to assure the safety of the highways, and they have access to several detailed collision databases in Canada. Annually, they publish a summary of all incidents and events occurred in the highways under a report called "Canadian Motor Vehicle Traffic Collision Statistics". This summary provides some descriptive statistics, and this information could be useful for checking the intuitiveness of the future results and also begetting insight into the current status of the roads.

Figure 3.4 summarizes the number of fatal and personal injury collisions in Canadian highways between 1998 and 2017 [109]. The general trend in the last two decades shows a gradual reduction (except some inconsistencies between 1999 and 2002) in the frequency of both fatal and personal injury collisions by 35% and 23%, respectively.

Similarly, Figure 3.5 depicts the number of fatalities, serious injuries, and total injuries in Canadian highways between 1998 and 2017. 36% reduction in the number of fatalities, as well as 39% reduction in the frequency of severe injuries demonstrate significant improvements in the safety of the roads; however, it is still mandatory to identify what factors caused these improvements.

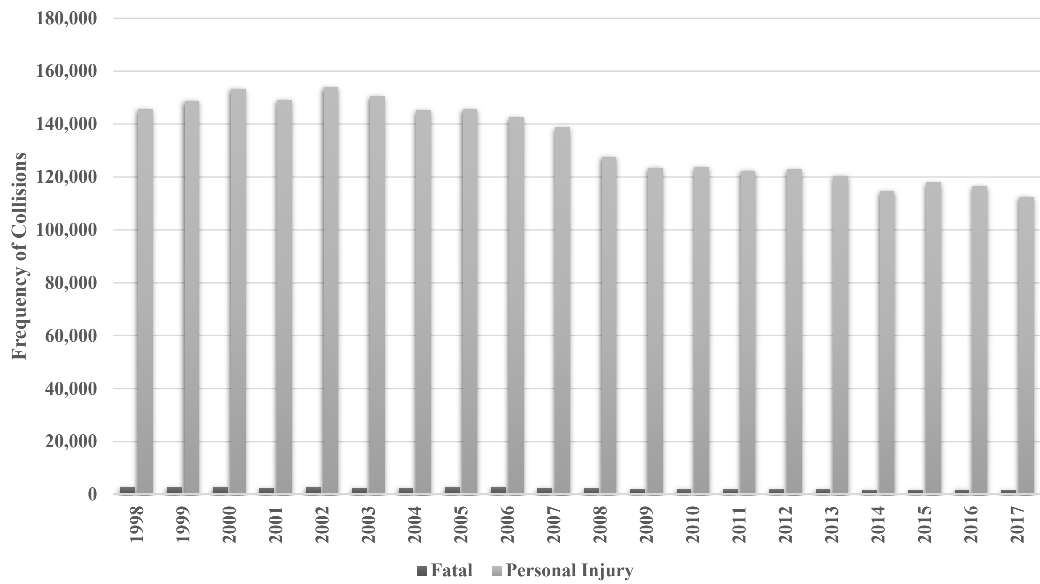


Figure (3.4) Transport Canada Annual Fatal and Personal Injury Collisions Summary

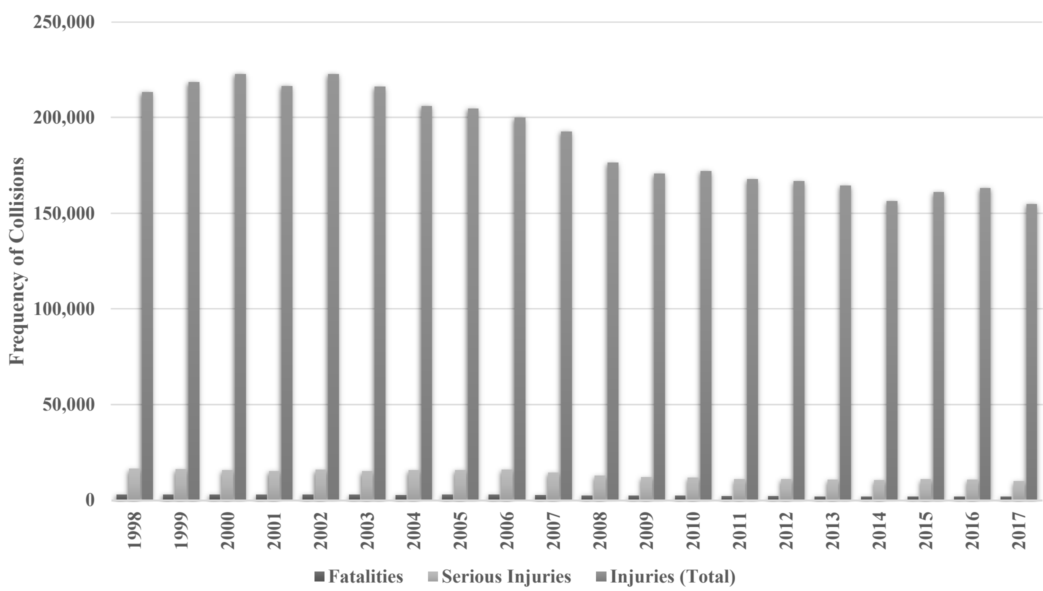


Figure (3.5) Transport Canada Annual Injury Severity Summary of Victims

The database from Transport Canada includes various information as follows,

- Collision date
  - Year
  - Month
  - Day of week
  - Collision hour
- Collision details
  - Collision configuration (first event)
  - Road alignment
  - Road surface
  - Roadway configuration
  - Collision severity
  - Traffic control
  - Weather
- Vehicle details
  - Number of people in the vehicle
  - Vehicle type
  - Vehicle year
- Person details
  - Injury severity
  - Person age
  - Person position
  - Person sex
  - Road user class
  - Safety devices

Ministry of Transportation of Ontario (MTO) also has a detailed database for all collisions within the province. This database's structure is fundamentally similar to Transport Canada's database. MTO annually publishes a summary of descriptive statistics of collisions under the "Ontario Road Safety Annual Report". Between 2012 and 2016, overall of 2352 fatal collisions were recorded; unfortunately, this report claims that in some of these fatal collisions more than one victim (overall 2589 fatalities) was involved. Table 3.4 demonstrates the details related to the victims in these collisions.

Table (3.4) Fatality Distribution by Role of the Victims (2012-2016)

Description	**Year**				
	2012	**2013**	**2014**	**2015**	**2016**
Number of Fatal Collisions	505	470	484	454	439
Number of Fatalities	568	518	517	503	483
Drivers	319	317	323	302	292
Passengers	128	96	78	94	87
Pedestrians	113	100	110	100	96
Other	8	5	6	7	8

Due to harsh winters and restrictions for performing maintenance, preservation, rehabilitation, and reconstruction activities, generally companies schedule their activities for summertime; this period is usually referred to as 'Construction Season'. Ontario Road Safety Annual Report (2012-2016) verifies that accident frequency through the construction season is higher than other times of the year. Figure 3.6 (a: Fatalities; b: Personal Injury) depicts the validity of this claim and demonstrates that from May to September of each year there is a noticeable increase in both fatal and personal injury collisions' frequency.

The presented databases could be used in several statistical analyses for highway safety, but unfortunately not for this research. The primary concern is because of the incompatibility of the databases with the research objectives. In the presented databases, unfortunately there were no records that were discerning whether the collisions occurred inside the work zones or not. Lack of this critical factor led the author to investigate the availability of extra unpublished information by these organizations. Several attempts were made to get the permission for accessing the raw data, but because of confidentiality issues, the author's requests were declined. At this point, it was decided to adopt a methodology that allows using other data resources and perform an independent data collection procedure.

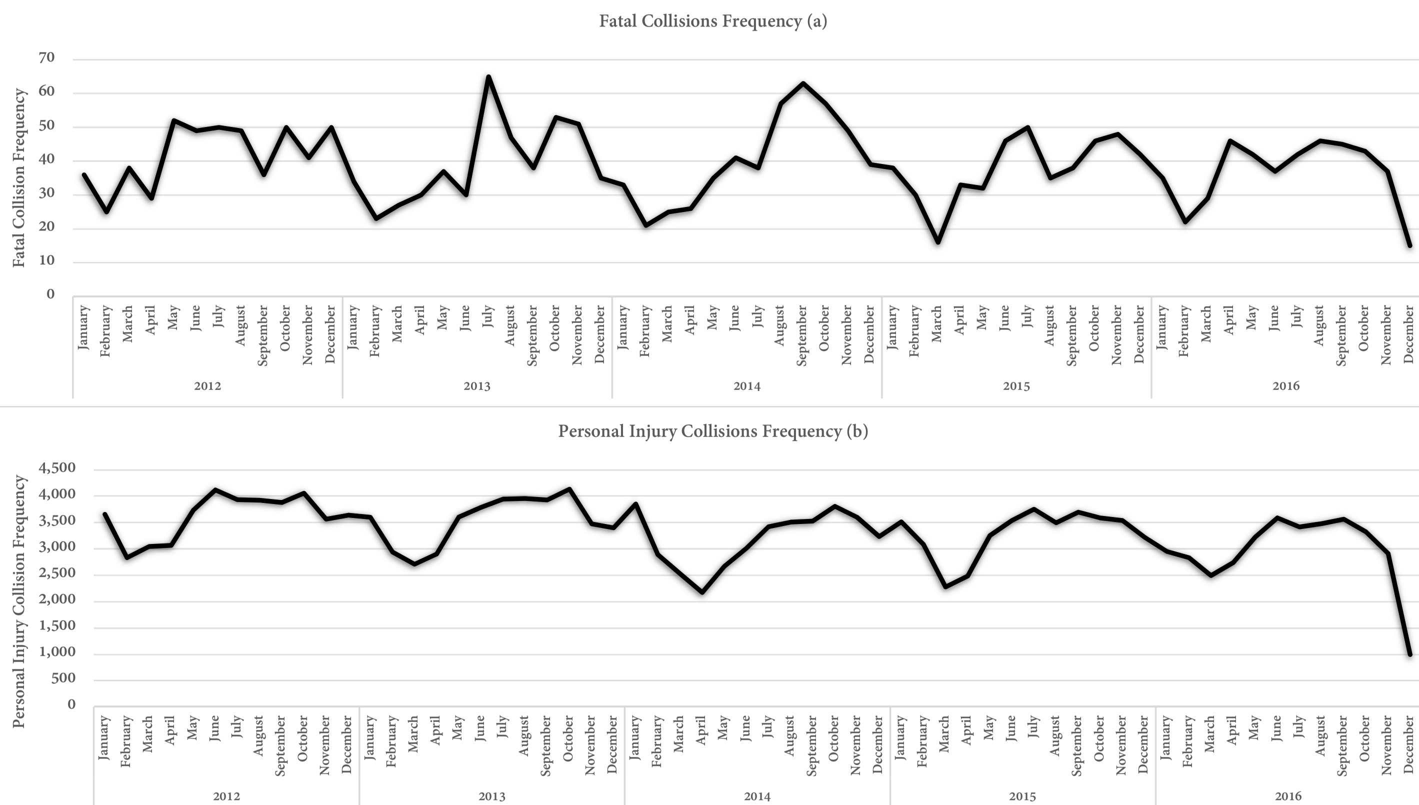


Figure (3.6) Fatal and Personal Injury Collisions Summary from 2012 to 2016

3.2.2.2 Work Zone Safety Survey

Surveys and questionnaires are the most well-known and well-established technique for data collection. For this research, a survey was designed to be sent to several engineers, contractors, managers, and organizations who were directly or indirectly involved in projects that required any type of closure in high-volume highways. The designed survey passed necessary clearance levels under the "Office of Research Ethics" of the University of Waterloo and was approved with "ORE-23031" identification code.

This survey includes five major sections, an overall of 45 questions, see Appendix B. The break down of the questions are as follows,

- Participant's information (2 short answer questions)
- Accident related questions (7 multiple choice questions and 2 short answer questions)
- Work zone and highway geometry questions (5 multiple choice questions and 6 short answer questions)
- Pavement related questions (5 multiple choice questions and 8 short answer questions)
- General information (5 multiple choice questions and 5 short answer questions)

This survey was sent out to more than 150 participants from all over Canada (but mostly concentrated in Ontario and Quebec provinces). Participants could be divided into the following categories,

- Participants from Federal Government
- Participants from Municipal (City) Governments
- Participants from Consultants
- Participants from Academia

Since this survey needed at least 45 minutes to finish and it should have been completed for each collision individually, many of the participants did not complete the survey; among all the participants, only two people fully completed the survey. Although the author is grateful for participants who fully submitted their surveys, this amount of responses was not statistically sufficient and reliable to conduct the study based on them.

3.2.2.3 National Highway Traffic Safety Administration (NHTSA)

National Highway Traffic Safety Administration in the US publishes a detailed data set that summarizes all of the occurred collisions within each year. This database presents the recorded information from collision reports, and it contains four major categories which are listed as follows,

- Person level information
- Pre-crash level information
- Vehicle level information
- Crash level information

Each of these categories then divides into multiple subcategories, which will be discussed individually. Also, all the presented data in all of the following sections are recorded in the work zones. Since there were some obstacles and limitations in accessing the Canadian collision data, it was decided to conduct the study on some of the US states which are relatively similar to Ontario, Canada. After careful considerations of laws and regulations, weather condition, and demographic information, it was decided to select the data from the following states for this research,

- New York State
- Pennsylvania State
- Illinois State
- Michigan State

The raw data set has been carefully polished based on two main factors,

- Whether the collisions occurred in the work zone or not
- Whether the data points have enough information [no or low number of missing values for each specific data point]

Person Level Information

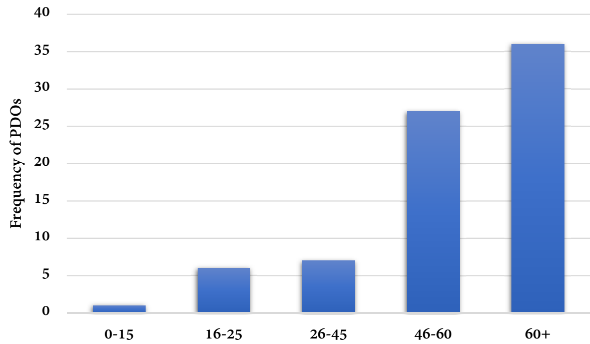
This category mostly focuses on general drivers' and occupants' related information such as,

- Age of driver and occupants
- Alcohol and drug test results
- Gender of driver
- Driver's previous convictions
- etc.

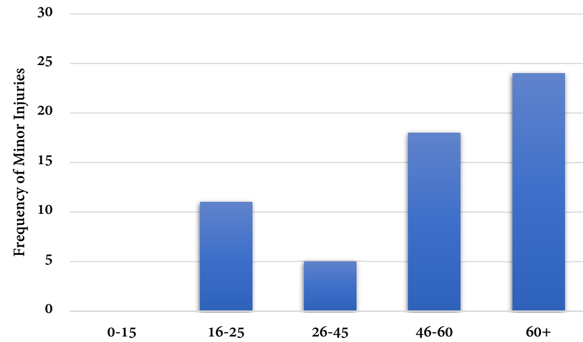
These pieces of information aim to investigate how demographic information impacts work zone collisions. Besides, this information could be valuable in terms of improving the driving quality by possibly reconsidering education techniques, laws, and regulations. One of the noteworthy facts in this data is that motorists older than 45 have the highest number of collisions in all of the injury severity categories (see Figure 3.7). Reconsidering the driving licence renewal policies for motorists older than 45 could potentially improve these statistics significantly. Middle age motorists who are between 25 and 45 years old have the lower fatality and major injury frequency in comparison with elder motorists. It was captured that young drivers experienced the lowest number of fatalities and PDOs. Figure 3.7b and 3.7c demonstrates that minor and major injury collisions occurred more frequently among young drivers.

Drug and alcohol involvement is one of the factors that was recorded and presented in this data set. Based on the available data set, alcohol and drugs consumers were responsible for almost 47% of the collisions occurred in the work zones between 2013 and 2016. Figure 3.8 presents the frequency of alcohol and drug-involved collisions in comparison with the total number of collisions in each injury severity category. Under the influence of alcohol and drug consumption, it is very likely to observe a delay in reaction of the motorists in the areas that road geometry changes, e.g. work zones; therefore, it was expected to see poor manoeuvres from drivers under the influence. Consequently, these delayed responses from drivers could end up on several unpleasant events such as fatalities and major injuries.

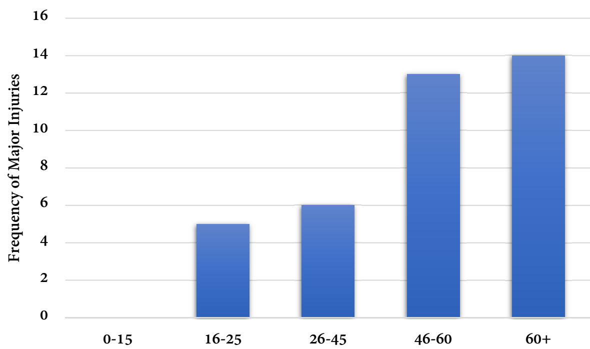
Gender of the driver was also reported after each collision. Out of 256 records, 182 of the victims where men and only 74 of them were women. Interestingly, 32% of both male and female drivers experienced PDO level collisions in the work zones. Similar behaviour



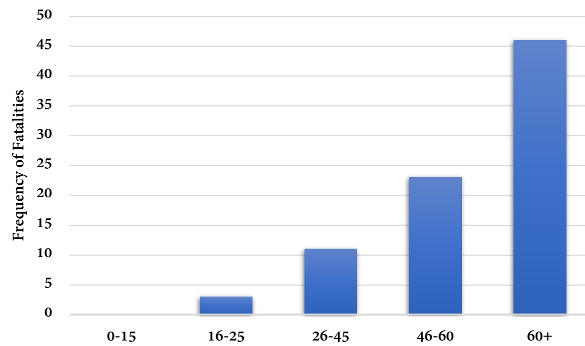
(a) Property Damage Only



(b) Minor Injury



(c) Major Injury



(d) Fatality

Figure (3.7) Frequency of Each Injury Severity Class for Each Age Group

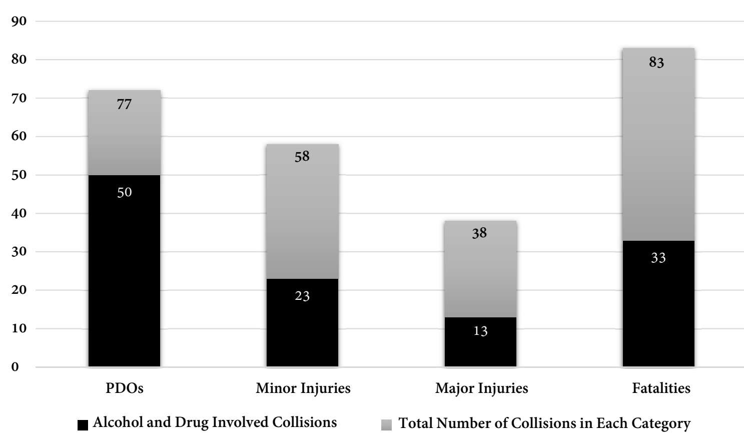


Figure (3.8) Alcohol and Drug Involvement in Work Zone Collisions was also captured for major injury class. However, in 34% of fatal collisions in the work zones drivers were male; this percentage is 14% lower for female drivers (approximately 20%). Figure 3.9 also demonstrates that female drivers have a higher ratio of minor injury collisions. Appendix C presents the related form which was used by NHTSA for this part of data collection.

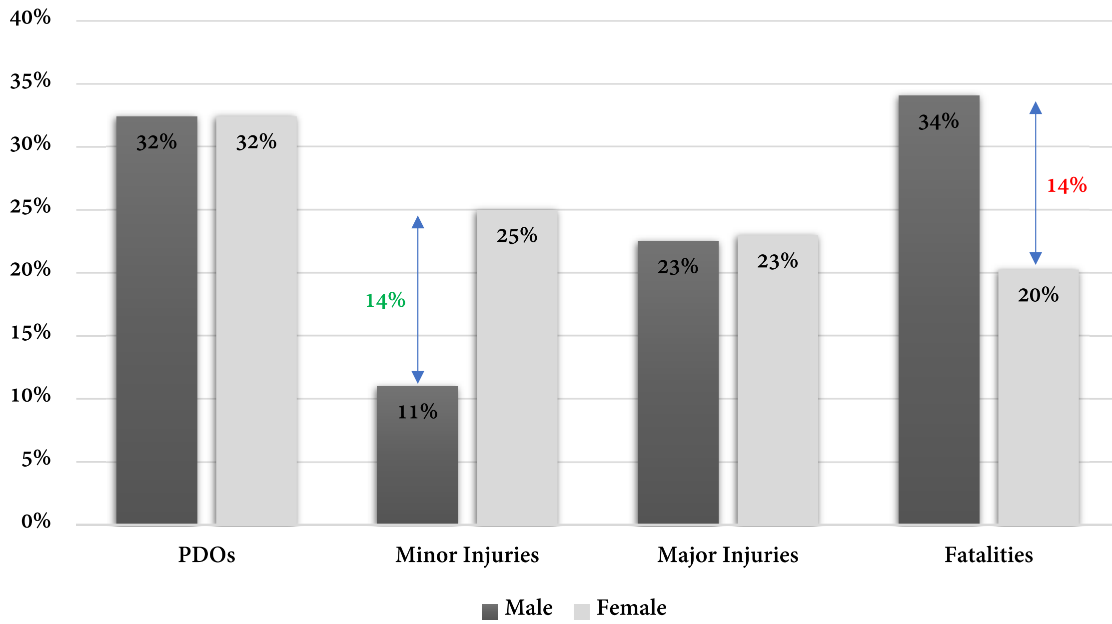


Figure (3.9) Gender of Victims Distribution in Work Zone Collisions

Pre-crash Level Information

Fifty-seven major critical pre-crash events are presented in the original data set, which 14 of those events were observed in the work zones. In this list of pre-crash events, several crash contributing factors such as loss of vehicle control, tailgating, and many others are filed. Some of the factors, e.g. 'Rollover' and 'Hit and Run', were eliminated from this list by the NHTSA decision.

Crash types also have been categorized in this part of data. This column of data could be useful to predict the possible crash type based on various scenarios before the work zones. One of the most common scenarios, as an example, was aroused by high-speed manoeuvres of motorists, which causes vehicle control loss and rear-ended collisions or immediate lane changes which could result in side-swiping collisions [110]. Generally, most important pre-crash factors could be listed as,

- Crash type
  - Drive off road
  - Control/traction loss
  - Parked vehicle.
- Critical event pre-crash; vehicle control loss due to:
  - Travelling vehicle
  - Other motor vehicles in lane
  - Other motor vehicles in lane encroaching into lane
  - Object or animal
- Driver distraction
  - Using or reaching to a device
  - Cellular phone
  - Inattention
- Roadway alignment
  - Straight
  - Curve right
  - Curve left
- Roadway grade
  - Level
  - Crest or Sag
  - Up hill or Down hill
- Roadway surface condition
  - Dry
  - Wet

- Snow
- Sand
- Slush

- Speed limit (35 mph to 70 mph)
- Traffic control devices
  - Flashing traffic control signal
  - Yield sign
  - Warning sign
  - Other regulatory sign and highway traffic signal

Although in this section of the data, several "Unknown" and "Not Reported" cells were noticed, two trends were not negligible in data set for the factors related to "Traffic Control Devices". The first trend as shown in the Figure 3.10 shows that in only 33% of the work zones with fatal collisions, warning signs were used; however, in almost 50% of the work zone with PDO, minor injury, and major injury collisions warning signs were used. This fact emphasizes the importance of warning signs before the work zones and its possible effect on reducing the severity of the collisions.

The other significant trend is related to the absence of traffic control devices (signs) in or around the work zones. Figure 3.11 demonstrates that surprisingly, in 55% of the fatal collisions, no traffic control devices were placed. Installation of traffic control devices for other injury severity classes was 10-15 percent more frequent. Based on this observation, it could be concluded that the presence of traffic control devices could play an important role in avoiding severe collisions and fatalities.

Vehicle Level Information

Vehicle-related information presents general information about the overall type of vehicles, areas of impact, hazardous material involvement, travel speed, many other factors. NHTSA decided to include two separate columns for "Rollover" and "Hit and Run" which were eliminated from critical pre-crash events.

Even though setting an appropriate speed limit for the work zone is mandatory, controlling the travel speed of the vehicles in the work zones should also be considered as a crucial task. Higher travel speed is the main reason of rollovers, loss of control, rear-ended collisions, and

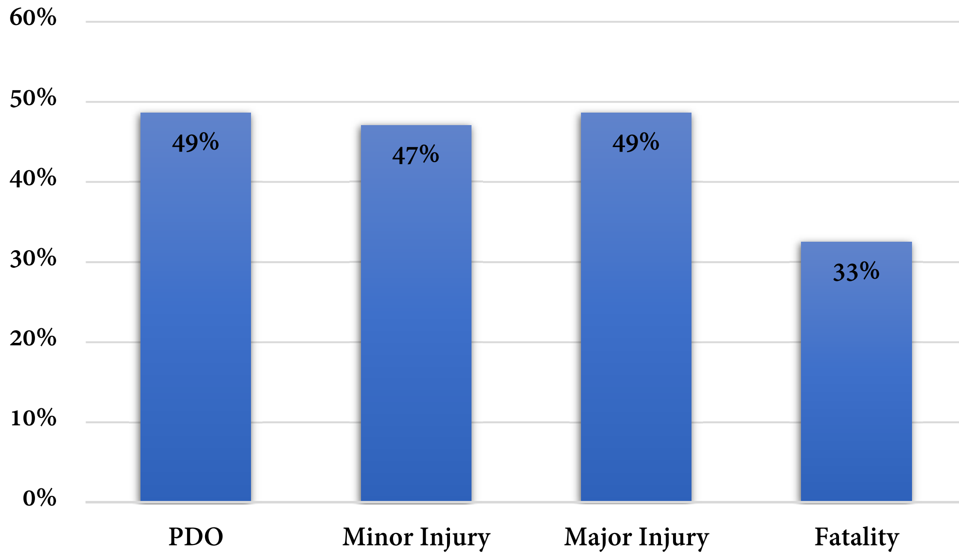


Figure (3.10) Warning Sign Effect on Occurrence of Each Injury Severity Class

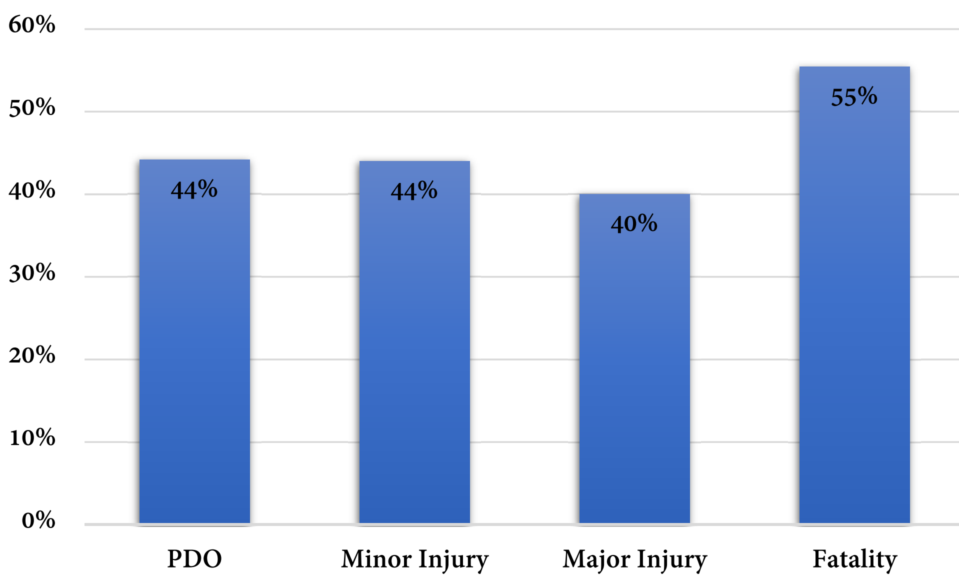


Figure (3.11) Effect of Absence of Traffic Control Devices on Each Injury Severity Class several other unpleasant events, and it is expected to observe more severe consequences of speeding prior and within the work zones. Figure 3.12 demonstrates that high travel speeds were recorded for about 60% of the fatal collisions that occurred in the work zones. Travel speed between 85 and 100 miles per hour has the highest share of the PDO, minor injury, and major injury collisions as well; this fact clarifies the importance of travel speed control in the work zones. Another speed-related factor concerns stationary vehicles; these vehicles are ranked second in causing collisions in the work zones, and they could be categorized into two major groups,

- Vehicles which are waiting in the queue to pass the work zones
- Parked trucks and construction vehicle around the work zones

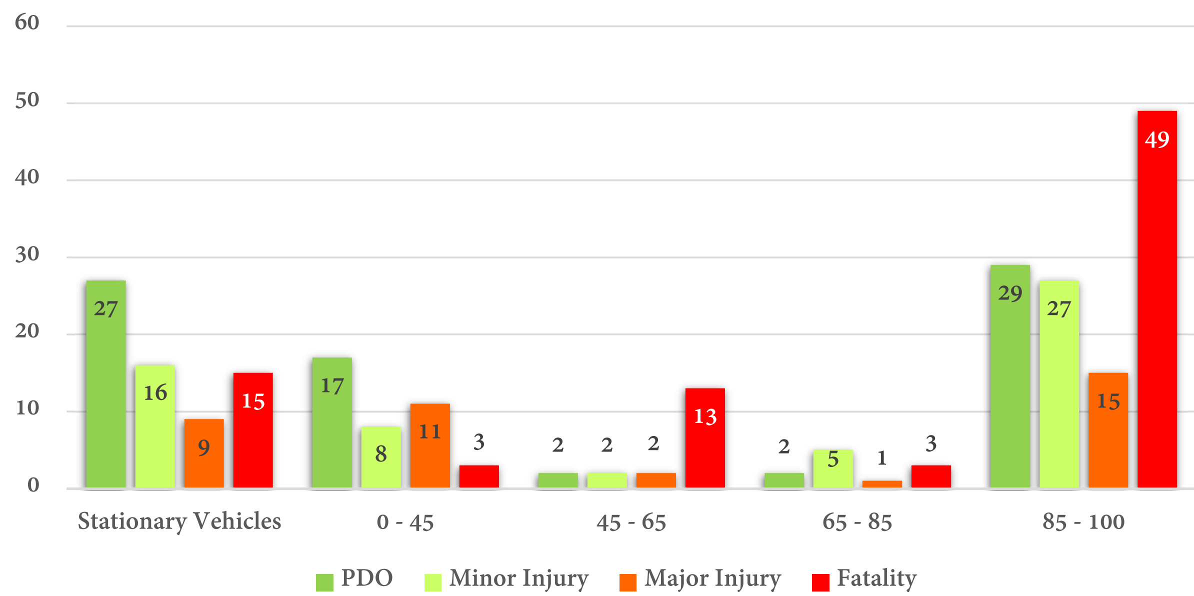


Figure (3.12) No. of Collisions in Each Injury Severity Class with Various Travel Speeds

Another important fact in this data set was related to rollovers. Rollover occurred in almost 28% of the collisions, which resulted in fatalities. Rollovers could be the consequence of fast lane-changing manoeuvres, double lane changes, and on some occasions hitting the concrete barriers with high speed prior or within the work zones. Figure 3.13 portrays the fact that only in 1% of the collisions with PDO severity, rollover occurred.

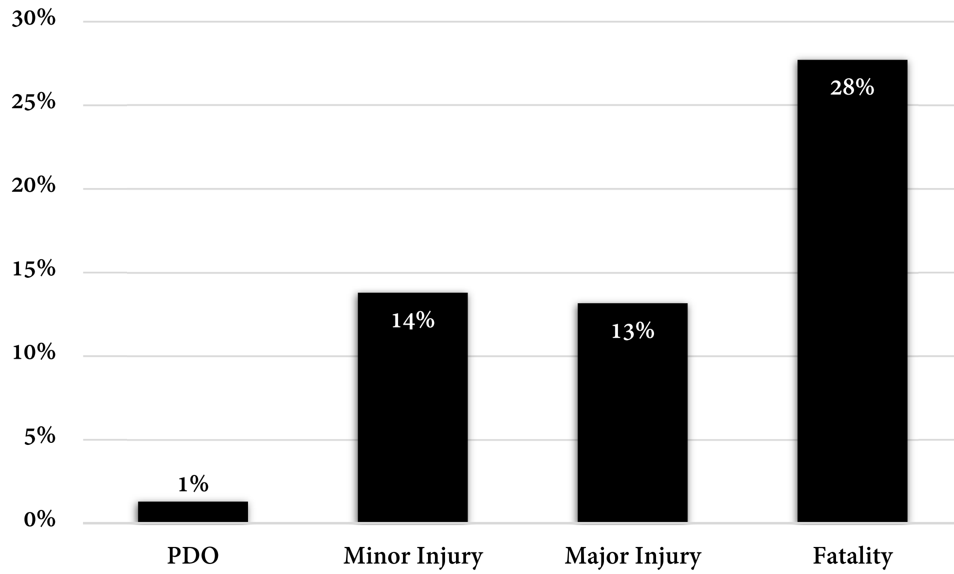


Figure (3.13) Observed Rollover Rate in Each Injury Severity Category

In the US and Canada, truck drivers go through some strict training programs and they must pass complicated exams to receive their authorized driving license for driving commercial vehicles. Nevertheless, it was quite noticeable that trucks had a large share of the befell collisions in the work zones. 166 out of 256 collisions for the study period involved at least one truck, approximately 65%. This statistic brightens the importance of truck restrictions in the work zone. Figure 3.14 represents the rate of truck-involved collisions in each injury severity category. Trucks were responsible for 74% of the PDOs, 60% of minor injuries, 71% of major injuries, and 57% of fatalities.

Even though it was investigated that fatality has the lowest rate among other severity types in truck-involved collisions, it is worth mentioning that this category was ranked second in terms of the number of fatalities; this means that one truck-involved collision could harm several people in the process. Figure 3.15 illustrates the number of truck-involved collisions for various injury severity classes.

Appendix D shows a data collection form for vehicle-related data.

Crash level information

This part of the data is mostly focused on each collision itself. Information such as day of the collision, month of the collision, location, weather conditions, lighting conditions, manner of the collision, relation to the junction, and relation to traffic way, is presented in

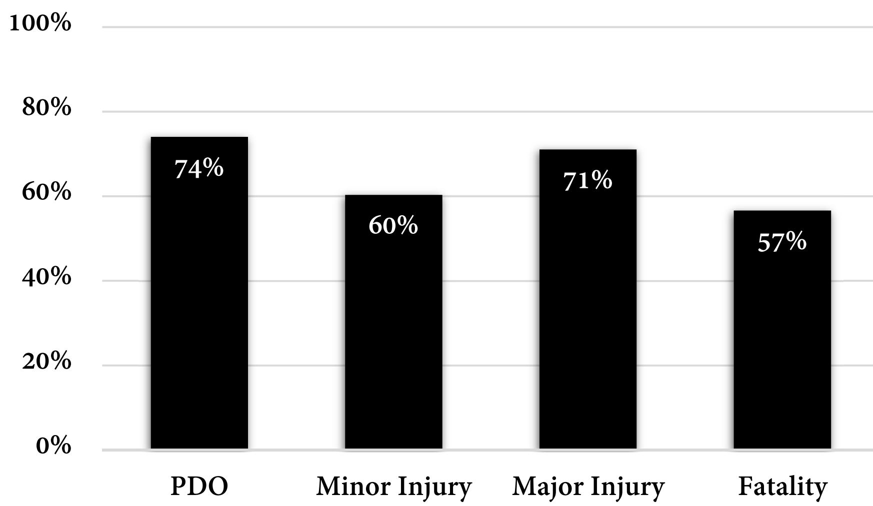


Figure (3.14) Truck involvement Rate in Each Injury Severity Category

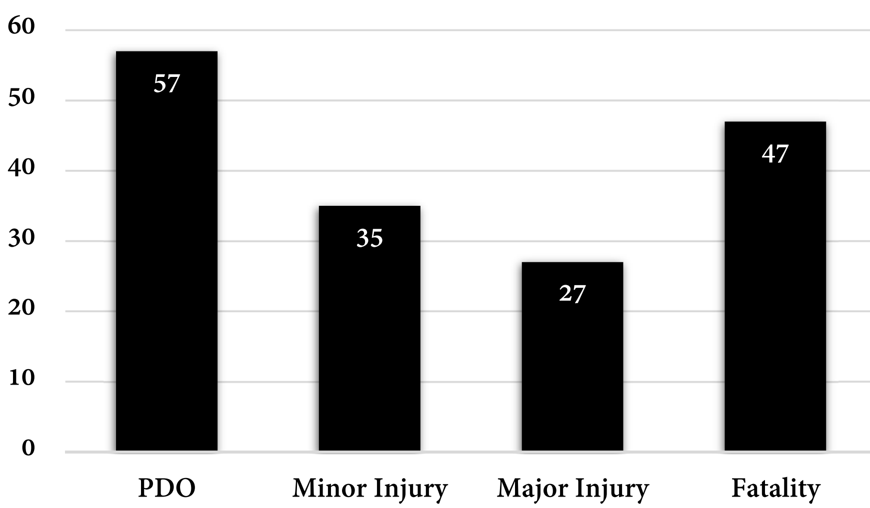


Figure (3.15) Number of Truck Involved Collisions in Each Injury Severity Category the corresponding forms. Many of the factors presented in this part of data could be potentially found as the most fundamental ones which can ensure the safety and convenience of the motorists based on the literature.

As mentioned in the literature review chapter, many studies have been conducted to investigate the best timing for setting up the work zones [3] [78] [79]. Deciding on the best timing for closures depends on several factors such as type of activity, AADT, and local rush hours. Figure 3.16 presents the hourly plot of the occurred collisions throughout the study period. As it is shown in Figure 3.16, most of the collisions in any injury severity class happened between 10 AM and 5 PM. It is also worth mentioning that between 2 and 3 PM, all the injury severity categories, except minor injuries, are at their highest level.

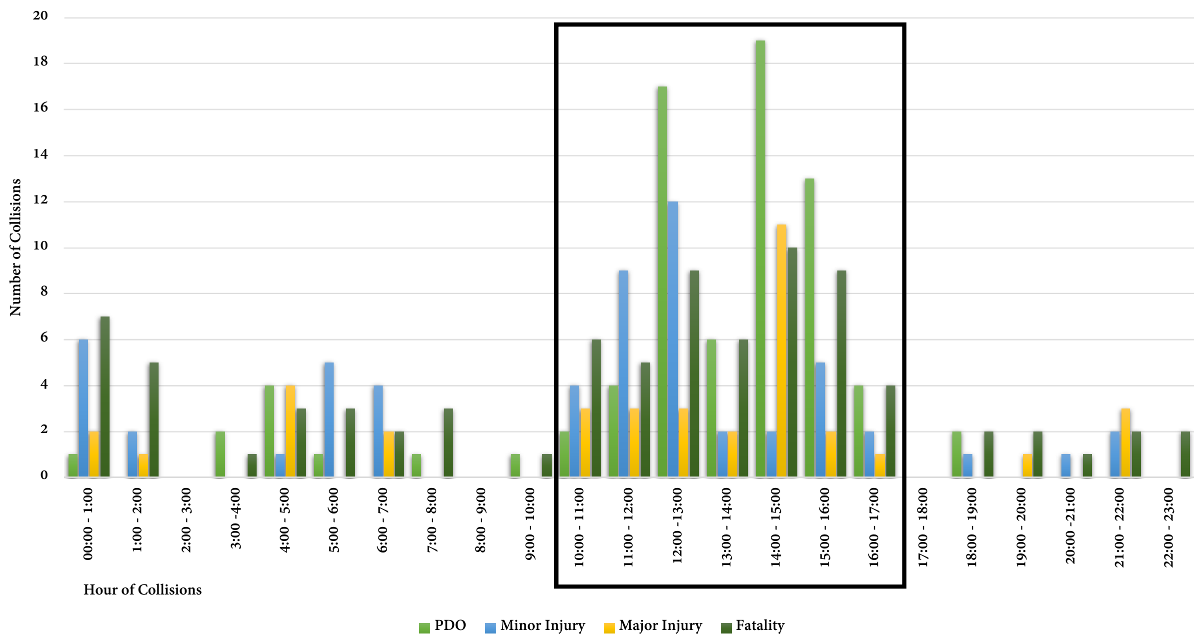


Figure (3.16) Hourly Collision Frequency for Each Injury Severity Category

It is also observed that the number of collisions between 9 PM and midnight were lower in comparison to earlier hour of the day. However, between midnight and 2 AM, the number of fatalities increased, and this could be due to interruptions caused by work zone setups. Another possible reason for observing such a boost could be "Drink and Drive". Overall, Figure 3.16's findings are consistent with Shepard & Cottrell [3] results which concluded that nighttime closures are safer and more convenient.

Several engaging facts have been found after a detailed analysis of the data set, which are summarized as follows,

1. 79% of the collisions occurred during the weekdays, and only 21% of them occurred during the weekends or holidays.

2. Tuesdays and Saturday have the lowest number of recorded collisions in the available data set. However, more than 60% of the collisions reported in only three days of the week (Mondays, Wednesday, and Thursdays).

3. 80% of the collisions in the work zones have happened in broad daylight. Rest of the collisions took place at dark, dawn, or dusk.

4. Males between the ages of 40 to 60 were found to experience more collisions in the work zones. This group is followed by male drivers aged between 25 and 39. Figure 3.17 depicts the collision rate by age groups and driver gender.

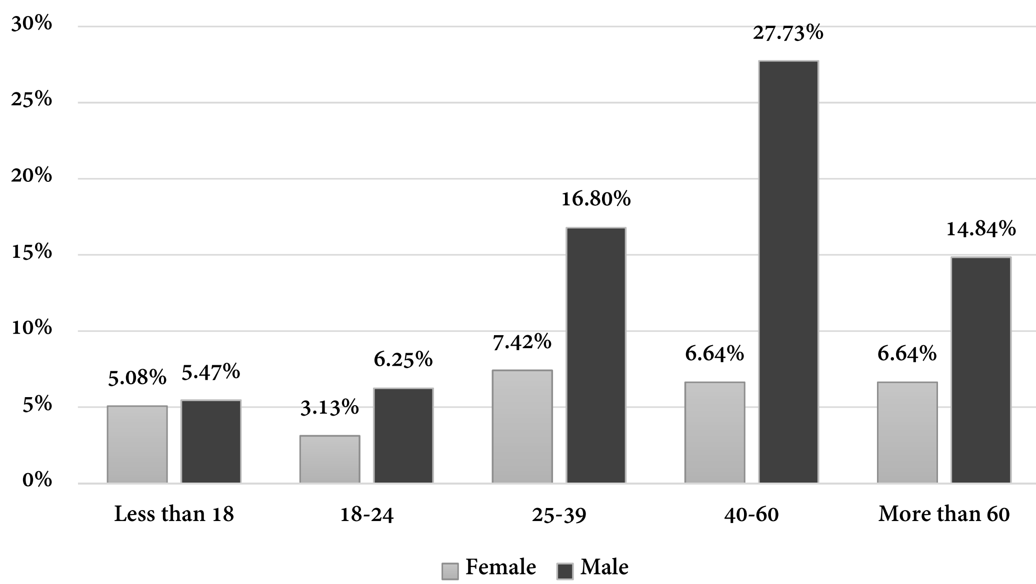


Figure (3.17) Collision Rate by Age Groups and Driver Gender

5. July has been found to have the highest number of collisions overall. May, June and August are following this month with 42, 33, and 32 collisions. This also emphasizes the fact that throughout the "construction season" the odds of observing the high frequency of collisions are higher. Figure 3.18 presents more details about collision frequency by each month of the year during the study period.

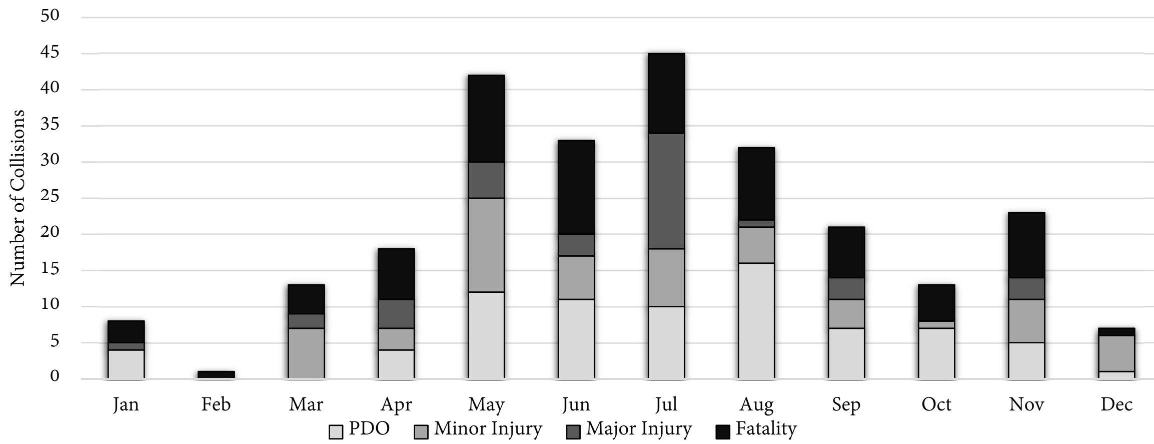


Figure (3.18) Collision Frequency by Month of Occurrence (2013-2016)

A sample of crash data collection form is presented in Appendix E.

3.3 Summary

This chapter summarized the fundamentals of statistical modelling techniques that have been adopted for this research. Also, the structure of this research has been discussed. The importance of gathering valid and useful databases as one of the most challenging and vital tasks for scientific research has been discussed. It further explained the author's several attempts in collecting required data for further analyses. In short, four sources of data have been evaluated, and the National Highway Traffic Safety Association (NHTSA) source has been selected for this research. Finally, this chapter provided the context for better understanding and knowledge about the available data set by introducing and discussing descriptive statistics of the most decisive factors of the data. Results of the developed models will be discussed in more detail in the following chapters.

Chapter 4

Work Zone Throughput Analysis

As mentioned earlier in Chapter 3, the initial phase of this study was to investigate the possible factors that can significantly reduce traffic congestion in the work zones. Distinguishing these factors could help decision-makers such as MTO to design more efficient work zones with less congestion. s could utterly help decision-makers such as MTO to design more efficient work zones with less congestion. The main purpose of developing these statistical models is not to simulate the work zones, but to find additional contributing factors, other than traditional factors such as upstream traffic state, downstream traffic states, in development of traffic congestion in work zones located in high-volume highways.

4.1 Multiple Linear Regression's (MLR) Results

4.1.1 Fixed Parameter MLR Results

Fixed parameter multiple linear regression model is considered as the base model used for this study, and it provided some insight into the possible influential factors on throughput analysis. This model as the starting point was developed for comparison purposes with its other competitors. Results of this model are also in line with previously developed models [4] [5]. The most significant advantage of the new model in comparison with antecedently developed models is the use of continuous variables. Having continuous variables instead of just dummy variables gives the luxury of having a more flexible model. Developed model investigated that several factors had a compelling effect on the throughput of the work zones as follows,

1. Presence of Police
2. Work zone configurations
   - One lane closure during the day-time
   - Two or more lane closures in the work zones with length of 3 km or more
3. Truck percentage
4. Work zone location

Table 4.1 demonstrates the estimated coefficients for statistically significant factors as well as their standard deviation. Also, t-values and p-values for each factor have been presented to confirm the significance of them.

Table (4.1) Fixed Parameter Multiple Linear Regression Model

Factors	**Coef.**	**St. Dev.**	**t-value**	**p-value**
Constant	1397.450	49.602	28.170	0.000
Presence of Police (1 if present, 0 otherwise)	-229.571	58.206	3.940	0.000
WZ Configuration (1 if it's day and one lane closed 0 otherwise)	499.296	68.074	7.330	0.000
WZ Configuration (1 if two or more lanes closed and the length of the work zone is greater than 3 km)	-158.116	41.850	-3.780	0.000
WZ Location (1 if 401 Highway, 0 otherwise)	-76.099	36.244	-2.100	0.039
Natural Logarithm of Truck Percentage	-70.050	20.614	-3.400	0.001

Presence of law enforcement officers has been found to have an adverse effect on the frequency of vehicles passing through the work zones. This parameter is believed to have a noticeable effect on the safety of the work zones, which will be discussed in later sections. Generally, motorists react rapidly to flashing lights of the police vehicles in or around the work zones. These reactions consist of immediate speed reduction and lane changes to provide safe clearance area for law enforcement officers, and other emergency vehicles; these actions are mandatory, and in case of not obeying them, the motorists could be penalized under the "Move over Law" (See Figure 4.1) [111]. The factor's coefficient is confirming this behaviour of the drivers. In addition, speed reduction prior to the work zones as well as lane changes slows down the stream entering the work zone, and as a result of it, it is expected to have more congestion in these locations.

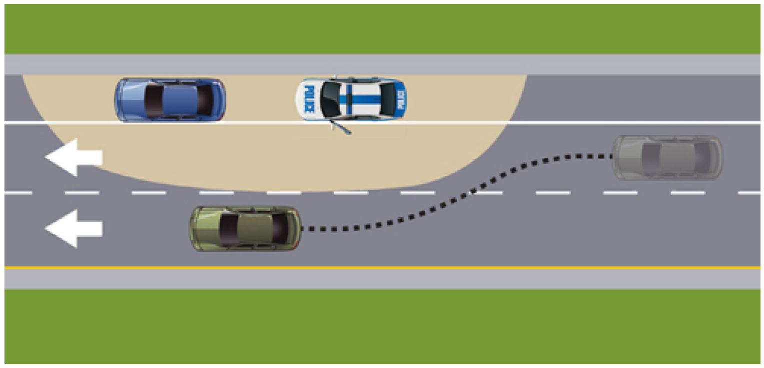


Figure (4.1) Move over Law Demonstration

Two other factors that have effectively altered the throughput frequency are related to work zone configurations and setups. The first factor in this category shows that although the motorists would experience a notable speed reduction during the daytime single lane closures, the traffic could still move in an acceptable rate (a positive sign of the coefficient). On the other hand, this is not the same for the work zone with 2 or more lanes closed and lengthier than 3 km. 2 or more lane closures itself can significantly interrupt the traffic flow of high-volume highways. Also, the speed limit in work zones is lower than highways' regular speed limit. This means in more extended work zones, motorists should drive more distance with lower speed. Knowing the fact that speed limit prior to the work zone is higher than work zone itself, it would be logical to claim that the charging rate of the section is higher than the discharging rate of it; therefore, queuing would be very probable. Combination of 2 or more lane closure and long work zone (longer than 3 km) was found as a factor which increases the queuing in the work zones.

The United States has the most extended (6.58 million kilometres in total length) and biggest road network in the world, however, it does not necessarily mean that the busiest highway also belongs to this road network. Road traffic- technology website[1] claimed that the most hectic highway in North America belongs to Canada. As stated in this report in 2014, the annual average daily traffic (AADT) count of highway 401 in Ontario was as high as 431,900 between Weston Road and Highway 400 exits. However, in some specific days of the year, this count outpaced 500,000 vehicles. Figure 4.2 presents the summary of AADT, Summer Average Daily Traffic (SADT)[2], Summer Average Weekday Traffic (SAWDT)[3],

[1] www.roadtraffic-technology.com

[2] Summer Average Daily Traffic; defined as the average twenty-four hours, two-way traffic for the period July 1st to August 31st including weekends.

[3] Summer Average Weekday Traffic; defined as the average twenty-four hours, two-way traffic for the

Winter Average Daily Traffic (WADT)[4] between 1988 and 2016. This figure emphasizes the intensity of traffic on 401 highway and possible congestion issues by setting up the work zones. These facts justify the significance of the 401 related factor in the MLR model. The negative sign of this dummy variable indicates that it is very probable to see long queues and delays in or around 401 work zones.

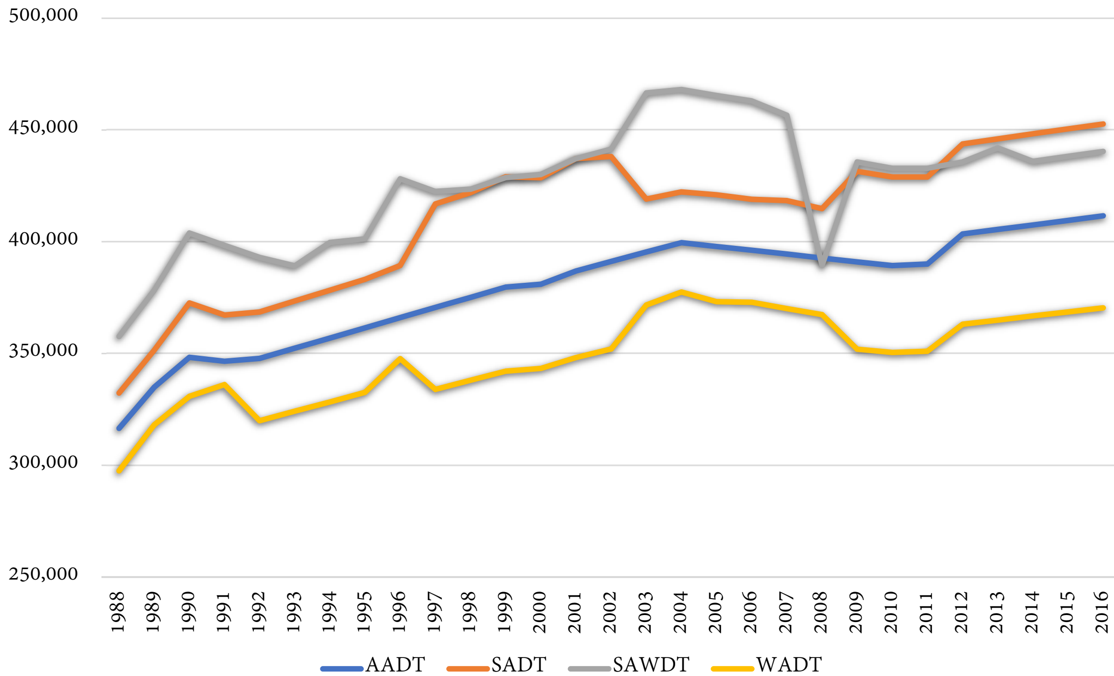


Figure (4.2) Highway 401's Traffic Summary 1988-2016

Throughput frequency in the work zones also found to be related to the natural logarithm of the trucks percentage. The developed model indicates that throughput decreases by 70 vehicles per hour for each 1 unit increase in truck percentage, which is quite notable. Several studies in the past have confirmed the possible relationship between truck percentage and traffic congestion [112].

After checking the intuitiveness of the significant factors, the ability of the model in prediction has been assessed. To that end, it has been checked how distant the predictions are in comparison with actual observed values. This model was able to predict the events

period July 1st to August 31st, excluding weekends.

[4]Winter Average Daily Traffic; defined as the average twenty-four hours, two-way traffic for the period January 1st to March 31st, plus December 1st to December 31st, including weekends.

as shown in Table 4.2,

Table (4.2) Prediction Assessment of Fixed Parameter MLR Model

Number of correct predictions			Total # of Observations	RMSE
within 100 veh	within 200 veh	within 300 veh		
45 54%	65 79%	80 98%	82	146.946

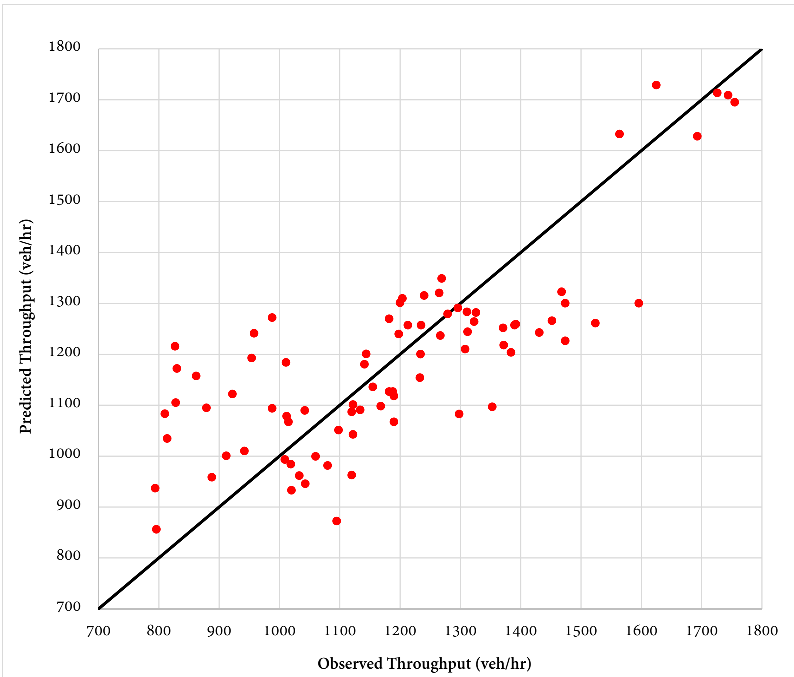


Figure (4.3) Predicted vs. Observed Throughput plot for MLR Model

Figure 4.3 presents the observed vs. predicted graph. Although all of the points are approximately around the 45-degree line, some of them are slightly off of this line. Also, this figure shows that predictions for work zones with lower throughput is slightly worse than mid-range and high throughput work zones. Predicting slightly higher throughput for a specific work zone is not necessarily a disadvantage, but could be considered as a conservative prediction, which could encourage the contractors to design for more challenging situations.

The R-squared and adjusted R-squared for this model are 0.71 and 0.68, respectively. Author believes that by using more detailed data sets and adapting more advanced statistical models, these results could be improved. This provided the context of developing a Negative Binomial model, which is a more established technique for count data modelling. Also, random parameter approach will be used to account for unobserved heterogeneity.

4.1.2 Random Parameter MLR Results

Random parameter multiple linear regression model is the next model, which is developed based on the fixed model that has been established in the previous section. This model also found very similar significant factors, but was not only limited to those. These additional factors demonstrate the ability of the random parameter modelling approach to allow more factors to be significant by allowing them to change across the observations. Random parameter multiple linear regression model found the following additional significant parameters,

- Number of closed lanes
- Time of day; Night or Day
- Type of barrier (as random parameter); Barrel or Concrete

In the fixed parameter MLR, it was shown that in the work zones longer than 3 km and 2 or more lanes closed, discharging rate of the traffic occurs in a slower pace and as a result of it, queuing could be initiated. This factor was also found to be significant in random parameter MLR. However, this model also captured that no matter how long is the work zone, closure of 2 or more lanes can reduce the number of vehicles passing through the work zones significantly. Based on the magnitude of these factors, it could be concluded that if 2 or more lanes are closed, almost 118 fewer vehicles per hour per lane could pass the work zone. This value could be even greater when the length of the work zone is more than 3 km, i.e. approximately 232 fewer vehicles per hour per lane.

This model also emphasized the importance of nighttime closures. The negative sign of the factor means that nighttime closures increase the queuing. However, delays caused by these queuing at night cost less than daytime closures. Since 88% of the data was collected at night, this factor was highly expected to be statistically significant. The leading cause of delays and congestion in the work zones during the nighttime closures could be a lack of visibility. Moreover, work zones generally interrupt the regular flow of the highways, and no matter when the closure occurs, it is expected to see congestion in or around these zones. Therefore, although nighttime closures might potentially cause delays, it is more efficient than daytime closures. Furthermore, in the previously discussed studies in the literature review chapter, it was determined that nighttime closures are preferred over daytime closures. One of the actions to improve the throughput flow in the work zones is to upgrade the lighting system prior to and within the work zones at nights. According to the American Traffic Safety Services Association (ATSSA) report [113], it is mandatory to have a proper lighting plan for the work zones. This plan should be developed by considering the following details:

1. Determine the activities in the work zone and required lighting levels
2. Assurance of work zone illumination
3. Select the type of lighting system and its corresponding source
4. Decide on fixture location of the lighting system
5. Final verification of the designed plan

Some of the states and provinces in the USA and Canada have established some preliminary guidelines toward providing appropriate lighting conditions in the work zones. Vecellio [114] summarized some of these studies, as shown in Table 4.3. Adopting the proper lighting system plans based on the existing studies and guidelines could result in less congested nighttime work zones. The implementation of the efficient and proper plans could potentially offer the following advantages [115],

- Having safer and more practical construction zones by designing a decent lighting level;
- Offering a solid traffic control plan as well as better supervision on motorists who are passing through the nighttime work zones;
- Enhancing the safety of construction workers and motorists.

The other statistically significant factor is related to the type of the barrier which was used for the closures. This factor as a random parameter with Normal distribution shows that in almost 97% of the time throughput is decreased by installation of concrete barriers, and the mean effect of this factor on throughput is -279 veh/hr. (See Figure 4.4). This congestion could be due to drivers' perception; drivers' seemed to decelerate faster when they recognize concrete barriers. The regular barrels are made of low-density polyethene, which may cause

Table (4.3) Summary of Illumination Practices in some of the DOTs in the US

Agency	**Activity or Task**	**Specifications**
Caltrans	All Nighttime Operations	10 fc *
Florida DOT	Proper workmanship and inspections	5 fc
Georgia DOT	All Nighttime Operations Average maintained horizontal illuminance	20 fc over the work area for tower lights
		Minimum 50,000 lumens for a tower light
Illinois DOT	All nighttime operations	Provide a minimum of 5 fc throughout the work area
Missouri DOT	Construction equipment and labor are active	5 fc
Missouri DOT	Flaggers and other specified locations in lighting plan	0.6 fc
New Jersey DOT	Tasks on and around Equipment	100 lux
New Jersey DOT	Specific tasks such as crack filling, saw-cutting, and joint sealing	200 lux
North Carolina DOT		50,000 460,000 Lumens
North Carolina DOT		22,000 50,000 lumens output to provide 10 fc
Nova Scotia DOT and Public Works	All Areas Where Workers and Inspection Staff Work	60 lux average; 30 lux point
Rhode Island DOT		Use 250 watt Metal Halide type lights

* A foot candle (fc) is defined as one of the units for illumination measurement that is equal to one lumen per square foot or 10.764 lux.

Table (4.4) Random Parameter Multiple Linear Regression Model

Factors	**Coef.**	**St. Dev.**	**t-value**	**p-value**
Constant	1853.730	37.391	49.580	0.000
WZ Configuration (1 if two or more lanes closed and the length of the work zone is greater than 3 km)	-114.172	20.088	-5.680	0.000
Presence of Police (1 if present, 0 otherwise)	-195.473	27.416	-7.130	0.000
WZ Configuration (1 if two or more lanes closed, 0 otherwise)	-118.377	19.022	-6.220	0.000
WZ Location (1 if 401 Highway, 0 otherwise)	-64.968	17.668	-3.680	0.000
WZ Condition (1 if night, 0 otherwise)	-120.459	28.511	-4.230	0.000
Natural Logarithm of Truck Percentage	-66.029	9.439	-7.000	0.000
Type of Barrier (0: Barrel, 1: concrete)	-279.830	35.445	-7.890	0.000
Standard deviation of parameter distribution	152.883	8.235	18.570	0.000

no damage or minor damage when the vehicle collides with them. However, this would not be the same with concrete barriers. The damage caused by concrete barriers would be more significant in comparison with barrels; this fact intuitively encourages motorists to reduce their speed at a faster rate to avoid potential significant damages on their vehicles. Although this speed reduction could provide a safer environment for the motorists, it increases the queuing.

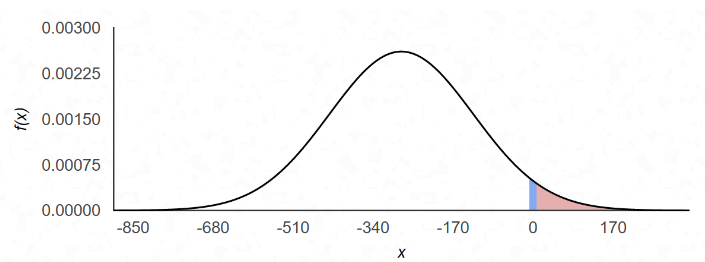


Figure (4.4) Concrete Barrier Effect on Throughput

Next step in evaluating the random parameter MLR model is to check this model's prediction power. Since this model is much more accurate than the fixed parameter competitor, the defined limits are stricter. Table 4.5 summarizes the results of the predictions and the corresponding Root Mean Square Error (RMSE) value for this model. It worth to mention that the R-squared and Adjusted R-squared are 0.98 and 0.97, respectively. These values show that the random parameter model has improved the predictability and accuracy.

Table (4.5) Prediction Assessment of Random Parameter MLR Model

Number of correct predictions				Total # of Observations	RMSE
Within 25	Within 50	Within 75	Within 100		
48 59%	72 88%	80 98%	82 100%	82	33.271

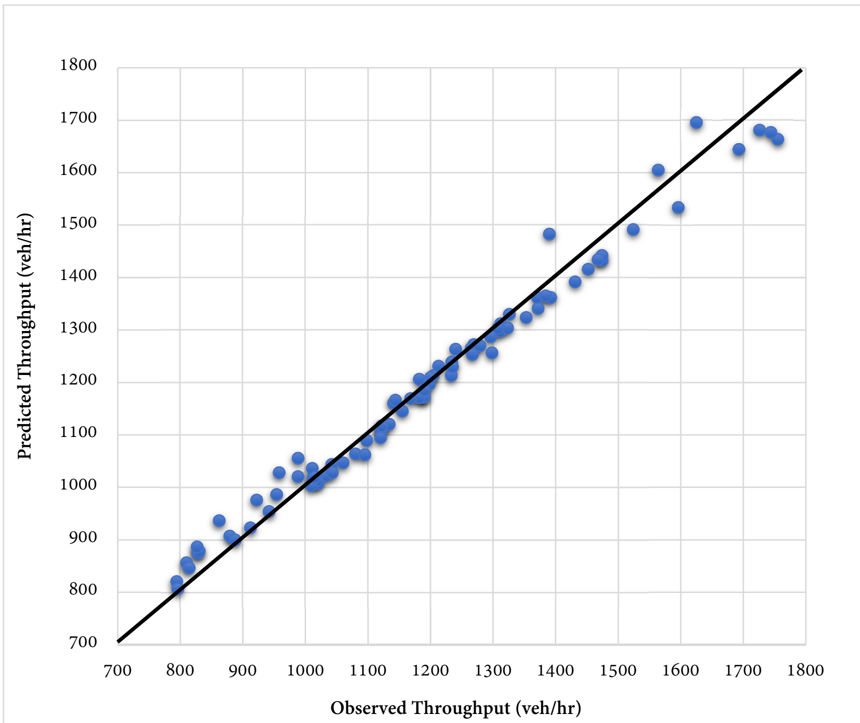


Figure (4.5) Predicted vs. Observed Throughput plot for Random Parameter MLR Model

The RMSE value has been improved undoubtedly, as well as both R-squared and Adjusted R-squared. The results of the prediction show that the random parameter MLR model could predict within 75 vehicle per hour per lane range almost as good as how the fixed parameter MLR performed in a much higher range of 300 vehicle per hour per lane. Figure 4.5 demonstrates the "Predicted vs. Observed" graph for the random parameter MLR model and its distribution around the 45-degree lane. Finally all of the normal probability plots and residual plot were analyzed for both fixed and random parameter models and none of them were following any specific or observable trend.

4.2 Negative Binomial (NB) Results

4.2.1 Fixed Parameter NB Results

Since the dependent variable could be categorized as discrete data (e.g. physically it is not possible to have 1954.6 veh/hr/lane passing the work zones), Negative Binomial regression is believed to yield more accurate and reliable results in comparison with linear regression. Statistically, adopting the best modelling approach based on the nature of the data is essential [101]. The factors which were found to have a reasonable influence on the throughput of the work zones are summarized in Table 4.6. Apart from similar factors which were found to be influential in the literature and the models presented in the previous section, some additional factors have been introduced as critical components of work zone congestion by this model. The factors which are similar to fixed and random parameter MLR models are as follows,

- Closures of 2 or more lanes could increase the congestion in the work zones.
- Closures of 2 or more lanes in the work zones longer than 3 km have greater congestion level.
- Work zones in 401 highway generally have more congestion than other 400-series highways.

Aside from these factors, some new findings were also explored. Usually, the speed limit of the King's highways (400-series) is 100 km/h. However, in individual sections of these highways, it was decided to post a lower speed limit of 80 km/hr or 90 km/hr. This lower speed limit is due to several factors such as [5],

- Improving the safety around the approaches to the "International Bridges"
- Satisfying the horizontal curve standards
- Satisfying the vertical curve standards

[5] http://www.thekingshighway.ca/intro.html

Table (4.6) Fixed Parameter Negative Binomial Model

Factors	Coef.	St. Dev.	t-value	p-value
Constant	7.138	0.037	193.160	0.000
WZ Configuration (1 if two or more lanes closed and the length of the work zone is greater than 3 km)	-0.144	0.040	-3.600	0.000
WZ Location (1 if 401 Highways, 0 otherwise)	-0.053	0.032	-1.670	0.095
WZ Condition (1 if speed limit is 80 km/h and it's night, 0 otherwise)	-0.226	0.060	-3.780	0.000
WZ Configuration (1 if it's weekend night and only one lane closed, 0 otherwise)	0.284	0.079	3.590	0.000
Truck Percentage (1 if it's higher than 20%, 0 otherwise)	-0.077	0.041	-1.870	0.061
Truck Percentage (1 if it's less than 6%, 0 otherwise)	0.088	0.042	2.070	0.038
WZ Condition (1 if 2 or more lanes closed, 0 otherwise)	-0.323	0.077	-4.210	0.000
Dispersion Parameter (α)	0.014	0.003	4.890	0.000

Table (4.7) Marginal Effect for Fixed Parameter NB Model Factors

Factors	Effect	St. Dev.	t-value	p-value
WZ Configuration (1 if two or more lanes closed and the length of the work zone is greater than 3 km)	-163.863	43.562	-3.76	0.000
WZ Location (1 if 401 Highways, 0 otherwise)	-245.514	59.598	-4.12	0.000
WZ Condition (1 if speed limit is 80 km/h and it's night, 0 otherwise)	-412.929	108.461	-3.81	0.000
WZ Configuration (1 if it's weekend night and only one lane closed, 0 otherwise)	313.032	83.369	3.75	0.000
Truck Percentage (1 if it's higher than 20%, 0 otherwise)	-89.7396	46.951	-1.91	0.056
Truck Percentage (1 if it's less than 6%, 0 otherwise)	106.204	52.316	2.03	0.042
WZ Condition (1 if 2 or more lanes closed, 0 otherwise)	-63.0734	37.489	-1.68	0.052

Based on the aforementioned analysis, setting up the work zones in these locations could be more challenging, and it is expected that these would also exerience more congestion. In the fixed parameter NB model, it was observed that the number of vehicles passing the work zones significantly reduces during nighttime closures at segments with the speed limit of 80 km/hr, which could be considered as intuitive based on the existing conditions.

The other interesting factor in this model could undoubtedly help decision-makers to have better planning strategies prior to the work zones set-ups. The model shows that having one-lane closure at weekend nights could significantly improve the queuing condition. The marginal effect of this factor is 313 (see Table 4.7), which means that 313 more vehicles per hour per lane were passing the work zones under this configuration in comparison with other queued work zones.

The truck percentage was also playing a major role in built-up of the congestion in the work zones. The model was able to distinguish the effect of the trucks in 2 independent factors. The first factor emphasizes the work zones that they have more than 20% of truck traffic. As was expected, high truck traffic volume resulted in more congestion and eventually queuing. Under this circumstance, the discharging rate would be reduced by 90 vehicles/lane. However, when there is less than 6% truck traffic in the work zone, the number of throughput increases by 106 vehicles per lane. Details of each factor in this model are presented in Table 4.6 and Table 4.7.

In the Negative Binomial estimations instead of R-squared and Adjusted R-squared, ρ^2 and corrected ρ^2 should be used. The purpose of ρ^2 is very similar to R-squared, but its use is common among estimations with log-likelihood function optimization. Equation 3.15 and 3.16 show the corresponding equation for each of the mentioned goodness-of-fit measures.

The LL(β) for fixed parameter Negative Binomial model is -524.092, the LL(0) is -2238.997, and finally, the corresponding value for k is 8. Based on these values, ρ^2 and corrected ρ^2 would be equal to 0.77 and 0.76, respectively. The Root Mean Square Error (RMSE) value for this model is lower than the fixed parameter MLR model, which shows this model is performing slightly better than the fixed parameter MLR model.

In terms of prediction accuracy, this model predicted 3 and 5 additional events correctly within 100 vehicles and 200 vehicles range, respectively (Table 4.8). Fixed parameter Negative binomial model does perform better than its other fixed parameter competitor, but not random parameter MLR. This fact would be the primary motivation for developing the random parameter Negative Binomial model to assess whether the predictability and accuracy of it could provide an even better understanding regarding the work zone throughput.

Finally, Figure 4.6 demonstrates the distribution of observed vs predicted points around the 45-degree line. Similar to the fixed parameter model, this model also tends to predict more conservative in lower throughput. Being excessively conservative during the planning stage is not desirable for the contractors due to its higher cost. However, it could potentially reduce both congestion levels and user delay cost.

Table (4.8) Prediction Assessment of Fixed Parameter NB Model

Number of correct predictions			Total # of Observations	RMSE
within 100 veh	within 200 veh	within 300 veh		
48 58%	70 82%	80 98%	82	141.795397

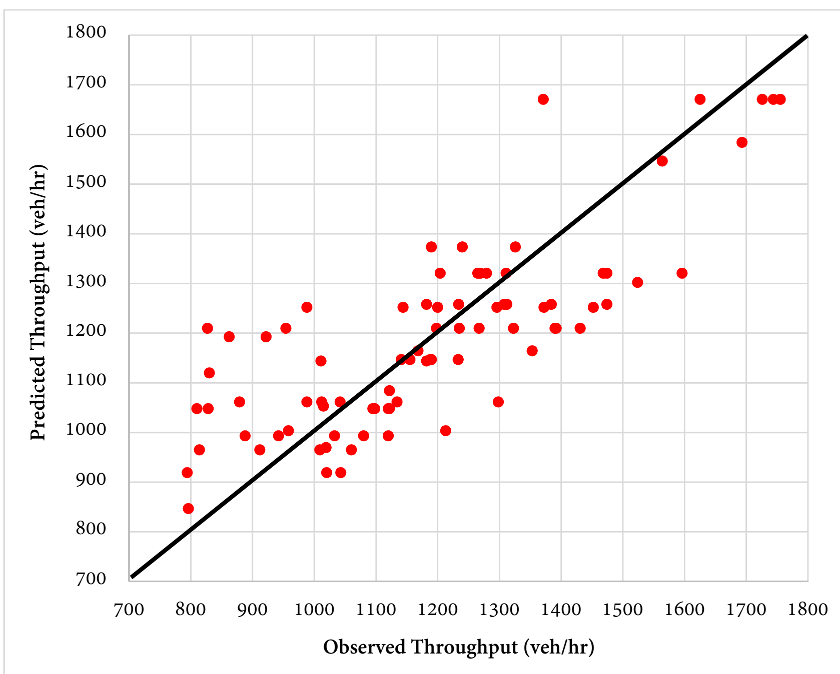


Figure (4.6) Predicted vs. Observed Throughput Plot for Fixed Parameter NB Model

4.2.2 Random Parameter NB Results

Random parameter Negative Binomial model, similar to the other developed models, accentuated the importance of work zone configurations and conditions. Seemingly, these factors are quite similar to the ones that were found to be significant in previous models, but not with the same effect. The effects of each factor could vary based on the estimation process. In this model, unobserved heterogeneity issue was addressed due to use of random parameter modelling, and it is believed that the results that will be discussed later in this section are a most realistic representation of the real-life events in the work zones. The effective factors based on this estimation could be listed as follows,

- Two or more lane closures in the work zones causes a reduction in the number of vehicles passing through them by 536 per hour oer lane. This value is quite high, and it shows the importance of this factor in controlling the traffic flow.

- If the work zone is 3 km or more and it has 2 or more lanes closed, 177 fewer vehicles would pass in comparison with shorter work zones with a similar condition. This factor emphasizes the importance of allowing motorists to adjust their speeds and driving behaviour throughout multiple shorter work zones, and consequently, this would allow less congestion and lower user delay cost.

- Nighttime closures found to be significant as a random parameter. This means that they might not behave similarly in all situations. Based on the results that are presented in Table 4.9, Table 4.10, and Figure 4.7, it could be concluded that in 78% of the occasions in average 120 more vehicles were passing through the work zones, while in the rest of events nighttime closures had negative effect on frequency of passing vehicles. This dualistic behaviour of this factor encouraged the author to investigate the specific situations which cause this inconsistency. Next two factors are discussing some of the scenarios which can explain this duality.

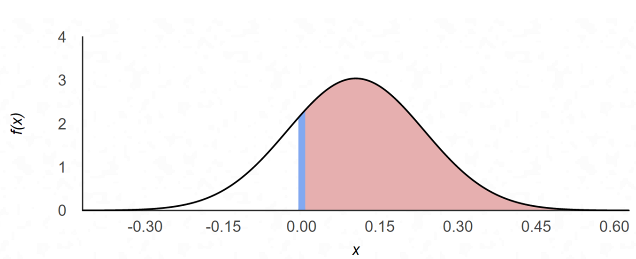


Figure (4.7) Nighttime Closure Effect on Throughput as Random Parameter

- As it has been explained in fixed parameter Negative Binomial model, in some particular segments of King's highways (400-series) the speed limit is as low as 80 km/hr due to satisfying geometric design specification and safety considerations. By setting up work zones in these sections during the night, almost 243 fewer vehicles could pass the work zones and congestion could be initiated (See Table 4.10).

- The next factor is associated with the specific configuration of the work zones. This factor shows that 493 more vehicles are successfully passing the work zones with one-lane closure during the weekend nights in comparison with other set up timings and configurations. Considering this dummy variable in the work zone planning and design could potentially reduce the level of congestion in these critical locations.

- The final two factors quantify the effect of the trucks on the work zone congestion. In the work zones with more than 20% of trucks, it is expected to see 102 fewer vehicles passing through them in comparison with the opposite condition. Thus, in segments with this issue (e.g. highways close to international borders), it would be more reasonable to redirect trucks to detours or temporarily prohibit trucks to pass from these specific segments. However, having less than 6% of the trucks in the work zones could slightly increase the number of passing vehicles by 70.

The final component of the Negative Binomial model is α, which is the over-dispersion parameter. The significance of this factor verifies the use of Negative Binomial model over the Poisson model.

After verifying the significance of the factors both statistically and intuitively, there is a need to verify the performance of the model by presenting the goodness-of-fit and prediction power. According to Equations 3.15 and 3.16, the values of LL(0), LL(β), and k are required. LL(0) is the same as the one for fixed parameter Negative Binomial model, -2238.997. LL(β) values have been improved significantly, and its value got to -356.742. Based on these value, the ρ^2 and corrected ρ^2 are calculated as 0.84 and 0.83. These ρ^2 and corrected ρ^2 values could be considered as very high for these type of models (equivalent to R^2 higher than 0.9) [103]; this measure will be discussed later in more details. Figure 4.8 shows observed vs predicted values for this model, and it shows its correctness of prediction by having them very close to the 45-degree line. The results from Table 4.11 also authenticate this claim.

Table (4.9) Random Parameter Negative Binomial Model

Factors	**Coeff.**	**St. Dev.**	**t-value**	**p-value**
Constant	7.012	0.026	271.980	0.000
WZ Configuration (1 if two or more lanes closed and the length of the work zone is greater than 3 km)	-0.152	0.016	-9.410	0.000
WZ Condition (1 if the speed limit is 80 km/h and it's night, 0 otherwise)	-0.209	0.022	-9.650	0.000
WZ Condition (1 if two or more lanes closed, 0 otherwise)	-0.461	0.032	-14.190	0.000
WZ Configuration (1 if it's weekend night and only one lane closed, 0 otherwise)	0.424	0.034	12.410	0.000
Truck Percentage (1 if it's higher than 20%, 0 otherwise)	-0.088	0.017	-5.120	0.000
Truck Percentage (1 if it's less than 6%, 0 otherwise)	0.060	0.015	4.130	0.000
WZ Condition (1 if night, 0 otherwise)	0.104	0.021	4.820	0.000
Standard deviation of parameter distribution	0.131	0.007	20.080	0.000
Dispersion Parameter (α)	524.384	121.389	4.320	0.000

Table (4.10) Marginal Effect for Random Parameter NB Model Factors

Factors	**Effect**	**St. Dev.**	**t-value**	**p-value**
WZ Configuration (1 if two or more lanes closed and the length of the work zone is greater than 3 km)	-176.675	-0.033	-9.010	0.000
WZ Condition (1 if the speed limit is 80 km/h and it's night, 0 otherwise)	-242.676	-0.020	-9.350	0.000
WZ Condition (1 if two or more lanes closed, 0 otherwise)	-535.948	-0.332	-12.910	0.000
WZ Configuration (1 if it's weekend night and only one lane closed, 0 otherwise)	493.430	0.341	11.290	0.000
Truck Percentage (1 if it's higher than 20%, 0 otherwise)	-102.546	-0.016	-5.060	0.000
Truck Percentage (1 if it's less than 6%, 0 otherwise)	70.160	0.016	4.060	0.000
WZ Condition (1 if night, 0 otherwise)	120.394	0.091	4.420	0.000

Table (4.11) Prediction Assessment of Random Parameter NB Model

Number of correct predictions				Total # of Observations	RMSE
Within 25	Within 50	Within 75	Within 100		
55	75	81	82	82	30.88
67%	91%	99%	100%		

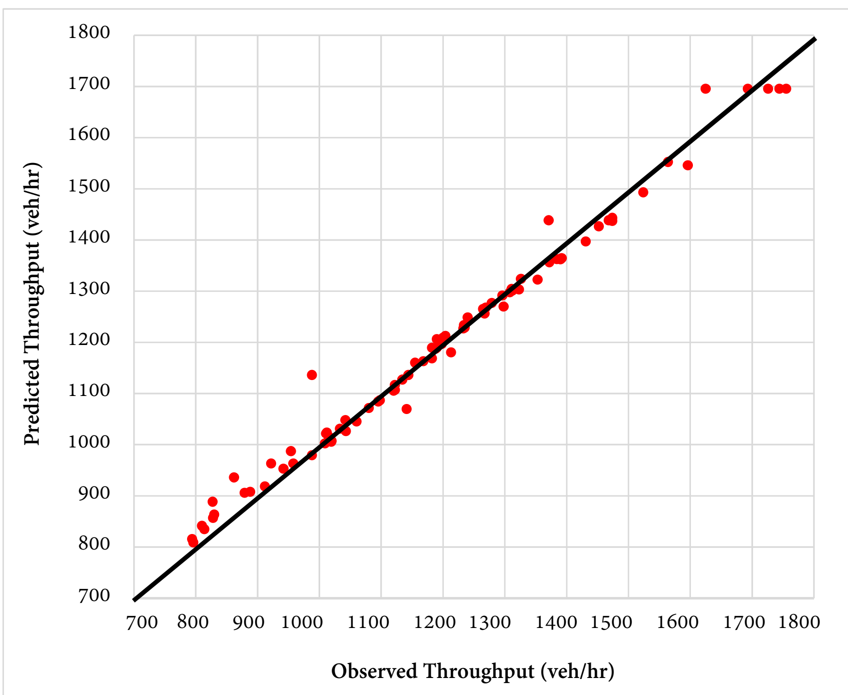


Figure (4.8) Prediction vs. Observed Throughput Plot for Random Parameter NB Model

4.3 Summary

This phase of this research focused more on identifying the influential factors on work zone throughput. To achieve this goal, four different statistical models have been developed as follows,

- Fixed parameter multiple linear regression model,
- Random parameter multiple linear regression model,
- Fixed parameter Negative Binomial model, and
- Random parameter Negative Binomial model

Among these models, random parameter estimation methods provided the luxury of accounting for unobserved heterogeneity and let the coefficients change across the observations. This led the author to have more accurate and reliable models. The summary of these two models is presented in Table 4.12.

Apart from the accuracy of the models, it is also vital to choose the most appropriate modelling approach based on the nature of the dependent variable. Previously, the Ministry of Transportation of Ontario adopted the fixed-parameter multiple linear regression model, which might not be the most suitable model for estimating the discrete nature of work zone throughput. Thus, this research aimed to present some innovative and up-to-date concepts in statistical modelling that can offer more accurate results by accounting for both unobserved heterogeneity and the nature of the dependent variable. These results can significantly avoid congestion in the work zones located in highways of Ontario. The final methodological conclusions of this phase could be summarized as,

- Random parameter models are performing better than fixed parameter models.
- Since Negative Binomial models account for the discrete nature of the dependent variable; it is more logical and reasonable to use them instead of multiple linear regression.
- Although multiple linear regression is not the most suitable estimation method for work zone throughput, it could provide a context to have an understanding of potential significant factors.

Table (4.12) Summary of Random Parameter Models

Random Parameter Model	**Multiple Linear Regression**		**Negative Binomial Regression**	
Factors	**Coefficient**	**t-value**	**Coefficient**	**t-value**
Constant	1853.730	49.580	7.012	271.980
Presence of Police (1 if present, 0 otherwise)	-195.473	-7.130		
WZ Configuration (1 if two or more lanes closed and the length of the work zone is greater than 3 km)	-114.172	-5.680	-0.152	-9.410
WZ Configuration (1 if two or more lanes closed, 0 otherwise)	-118.377	-6.220	-0.461	-14.190
WZ Configuration (1 if it's weekend night and only one lane closed, 0 otherwise)			0.424	12.410
WZ Condition (1 if night, 0 otherwise)	-120.459	-4.230		
WZ Condition (1 if speed limit is 80 km/h and it's night, 0 otherwise)			-0.209	-9.650
WZ Condition (1 if night, 0 otherwise)			0.104	4.820
Standard deviation of parameter distribution			*0.131*	*20.080*
WZ Location (1 if 401 Highway, 0 otherwise)	-64.968	-3.680		
Natural Logarithm of Truck Percentage	-66.029	-7.000		
Truck Percentage (1 if it's higher than 20%, 0 otherwise)			-0.088	-5.120
Truck Percentage (1 if it's less than 6%, 0 otherwise)			0.060	4.130
Type of Barrier (0: Barrel, 1: Concrete)	-279.830	-7.890		
Standard deviation of parameter distribution	*152.883*	*18.570*		
Dispersion Parameter			524.384	4.320
R-Squared or Rho-Squared	0.98		0.98	
Adjusted R-Squared or Rho-Squared	0.97		0.97	

- The best model for this phase of the study is ‘Random Parameter Negative Binomial Model’.

Also, according to the results, some recommendations and suggestions could be provided as follows,

- It is recommended to not only have police vehicles within work zones but prior to them. This could lead motorists to reduce their speed before they get to work zones and adapt to existing environment of work zones. This could highly improve road safety and avoid congestion in work zones.

- Although sometimes this is not possible due to busy planning schedule and workload to have multiple short distance closures, it is suggested to avoid more than 3 km work zones when 2 or more lanes are planned to be closed.

- Closures during the weekend nights with only one lane obstructed are found to be more efficient. Thus, MTO, as well as a contractor, should attempt to plan and design their work zones under these circumstances.

- Nighttime work zones tend to have higher traffic congestion; this could be due to data collection process and it is recommended for future studies to collect data equally from daytime an nighttime closures to avoid biased results. Despite this fact, it is expected to see some congestion when the segments speed limit is reduced to 80 km/hr due to geometric design limitation and safety issues. To potentially avoid congestion in these locations, work zones illumination system should be improved. Also, proper active signage system as well as informative intelligent transportation system technology ahead of the work zones can prevent or lessen the queuing.

- Due to the significance of truck percentage in the models, it is suggested to limit the truck access to work zones that are subjected to high truck traffic volume. This action could be performed by temporarily designating detours for the trucks.

- Work zones with concrete barrier tend to have more congestion in comparison with work zones that are redirected/closed with barrels. Although this conclusion is made based on the throughput model, and it justifies the use of barrels over concrete barriers, the safety advantages of the concrete barriers are not negligible. The actual effect of each barrier’s usage should be specifically investigated in future by accounting factors such as work zone’s location, configuration, and illumination condition.

Chapter 5

Work Zone Safety Assessment

Assessment of the work zone safety is the next step of this study. In order to have more concise results, some of the methodological conclusion from previous chapter will be adopted. This means that in this phase of the research, only random parameter models will be developed due to their accuracy of prediction and robustness. Random parameter ordered probability (Probit, Logit, Arctangent) models will be presented, and the results of each will be discussed individually. Apart from that, some new strategies to find the best coefficients for having even more accurate predictions will be introduced in this chapter. Note that these models are developed based on the US data due to confidentiality and limitations in accessing the detailed Canadian collision database.

5.1 Random Parameter Ordered Probability Models

In this section, the results from models that are developed from the full data set are presented, and the effect of each will be discussed independently. As a reminder from Chapter 3, the ordered probability models have been chosen due to their ability to capture the hierarchical nature of the dependent variable (PDO[1], Minor Injury, Major Injury, Fatality). In order to be consistent throughout the decision-making process, and provide the context for a fair comparison, it was aimed to investigate factors which are statistically significant in all three ordered probability models (Probit, Logit, Arctangent) at the same time. This was achieved by conducting several trials with various variables, and recording

[1]Property Damage Only

the most effective factors in every single trial to guarantee the involvement of the essential elements. The summary of the results for these models is presented in Table 5.1.

Table (5.1) Summary of Full Random Parameter Ordered Probability Models

List of statistically significant parameters	Probit Model		Logit Model		Arctangent Model	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Constant	5.561	7.140	12.311	7.280	10.458	7.070
Collision Hour (1 if collision occurred between 8PM and 6 AM, 0 otherwise)	1.348	4.320	3.551	5.660	3.089	5.650
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	1.172	4.710	2.518	5.170	2.148	5.110
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	0.997	3.830	2.112	4.220	1.831	4.250
Speed at the time of the collision (1 if it is less than 50 mph, 0 otherwise)	-1.510	-4.850	-2.870	-4.810	-2.451	-4.770
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	5.865	5.200	15.559	5.570	13.274	5.530
Driver related factor (1 if Failing to keep in the proper lane or run off the road or following improperly, 0 otherwise)	1.522	3.920	4.506	5.420	3.802	5.320
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	1.473	5.110	3.407	5.760	2.924	5.700
Traffic control devices (1 if used, 0 otherwise)	-2.331	-4.360	-4.631	-4.460	-4.003	-4.490
Manner of Collision (1 if rear-end, 0 otherwise)	-1.449	-3.600	-1.842	-2.490	-1.534	-2.430
Work zone signage (1 if a warning sign is available, 0 otherwise)	-1.660	-3.160	-4.296	-4.150	-3.762	-4.230
Heavy truck involvement (1 if it is involved, 0 otherwise)	1.903	5.140	3.838	5.370	3.337	5.370
(St. dev. of parameter distribution) Normally distributed	2.053	9.070	4.672	8.520	4.001	8.210
Roll over (1 if roll-over didn't occur, 0 otherwise)	-2.930	-6.760	-6.735	-7.070	-5.721	-6.870
(St. dev. of parameter distribution) Normally distributed	2.927	10.340	5.164	8.890	4.408	8.530
Gender of the driver (0 if male, 1 if female)	-1.197	-4.410	-2.087	-4.180	-1.762	-4.110
(St. dev. of parameter distribution) Normally distributed	1.976	9.010	6.493	8.940	5.515	8.570
Mu(01)	2.793	8.110	6.053	7.650	5.164	7.440
Mu(02)	4.431	10.080	9.643	8.990	8.222	8.630
Log-likelihood at convergence	-265.712		-254.662		-260.385	
Likelihood at zero	-372.337		-372.337		-372.337	

According to the developed models, nighttime collisions, between 8 PM and 6 AM, generally increase the probability of fatalities. This factor initiates a challenge for decision-makers. Chapter 4 results demonstrated that nighttime closures are performing better than daytime closures, and they can help to avoid congestion in the work zones. However, the developed models for this chapter are bringing up some issues regarding the nighttime closures. More impaired drivers, less visibility, drowsy drivers, and more heavy truck percentage at nights in comparison with days are possible reasons that can intensify the probability of more severe collisions.

Appendix F presents a summary of the partial effect of each significant factor in the models. In order to have a better understanding regarding the severity of the collisions, the effect of each factor on injury severity level has been calculated. Tables F.1, F.2, and F.3 demonstrate that actually PDO and fatal collisions are slightly affected by closure time. Although closures between 8 PM and 6 AM reduce the probability of minor injury occurrence, they increase the probability of major injury collisions. According to the ordered Logit and Arctangent models, the probability of minor injury collisions are reduced by approximately 70%, and from the other side, the probability of major injuries and fatalities increased by roughly 60% and 10%, respectively. The ordered Probit model results show less extreme variation toward these factors, and indicating a 35% reduction in minor injuries, and a 25% enhancement in major injury probabilities.

These results depict the reality that there is a trade-off between having safer highways and less congested highways. However, it is believed that stricter law enforcement and regulations at nighttime and especially prior to the work zones, as well as, encouraging the contractors to add one or more "Construction Zone" signs prior to the work zones, could improve the safety level of the work zones. Book 7 [116] recommends installing these signs 2 km prior to the work zones. Under these circumstances, a driver who approaches the work zone with 120 km/hr at nighttime could travel this distance in only one minute. Corfitsen [117] conducted a study which evaluated the reaction time of the young drivers under various conditions such as tired, very tired and rested. He claimed that the reaction time of the drivers could significantly increase at nighttime in almost all of the cases, but the rested drivers. This shows that offering more time for drivers to make a proper decision and adapting themselves to the existing condition could potentially avoid or at least lessen the severity of the collisions. Based on this discussion, by merely installing one extra sign, 3 km prior to the work zone, two important safety considerations will be addressed,

1. The travel time to work zone could be increased by 50% (1.5 minutes)
2. Motorists will be exposed to more frequent signs (It reduces the chance of driver to not recognizing any signs prior to the work zone at night)

Driver's age is also found to be statistically significant as a factor which can increase the probability of fatal collisions. This factor emphasizes on two major age categories,

- Drivers with age less than 25
- Drivers with age more than 50

According to the partial effects tables that are presented in Appendix F, this factor enhances the probability of major injury collisions by 18% in the Probit model, and 41% in both Logit and Arctangent models. Since both young and elderly drivers are susceptible to more severe collisions, there is an immediate need to resolve this issue by adopting more comprehensive approaches. Although this factor gathers young and elderly drivers in the same group, the strategies for overcoming this issue for each of those would be different.

The Organization for Economic Co-operation and Development (OECD) published a report on October 2006 [118], which emphasized the possible solutions for the enhancement of road safety by focusing on young drivers aged between 15 to 24. This report claimed that 8500 young car drivers died in OECD [2] countries. This statistic affirms that young drivers are not only putting themselves in danger but other motorists as well. This undeniable fact could be one of the reasons that insurance companies offer higher rates for young drivers (less than 25). Since the seriousness of this factor is clarified, some actions should be taken to lessen the safety concerns regarding the presence of young drivers in roads, and mainly work zones. Some of the recommended solutions are as follows,

- Enhancement of public awareness toward the existing problem; this factor could be implemented by encouraging young drivers to be more responsible drivers through social campaigns.
- Improvement in law enforcement strategies; this solution focuses mostly on presence of police in roads that are exposed to more population of young drivers. It is also recommended to have stricter rules and higher penalties for drivers who are not obeying the rules that are concerning seat belts, and drugs and alcohol consumption.
- Applying proactive restriction in the first year of solo driving period; this solution also leads to have less young and impaired drivers in the roads. In OECD report it was mentioned that the first year of solo driving is considered as the most dangerous time frame for the young drivers and they should be strictly supervised and controlled.

[2] List of OECD countries 2019

- Improvement in education system; this could possibly be one of the most efficient ways to enhance road safety. Having more mandatory practice sessions, more detailed pre-licensing educations, stricter driving examinations, and explanation of the consequences of careless driving in larger extends could effectively result in offering a safer driving environment for the drivers.

Apart from young drivers, senior drivers with age more than 50 years are also increasing the probability of having more severe collisions in the work zones located in the high-volume highways. It is worth mentioning that the raised concern does not question the seniors' driving experience and skills, but their physical limitations. Unfortunately, elderly drivers might be affected by some age-related impairments such as loss of sensory, cognitive, and motor skills [119] [120] [121] that make them more susceptible to collisions. Keskinen [121] assessed the effect of education for older drivers, and he concluded that educating older drivers is not enough by itself. Teachers should help senior drivers to have a better understanding of their abilities, skills, and challenges that they might face throughout driving. Moreover, some provinces of Canada, such as Ontario have a specific procedure for license renewal of senior drivers, but only for drivers more than 80 years old. Based on Statistics Canada report at 2009[3] only 6% of the senior drivers are more than 85 years old, and there are almost 3.5 million senior drivers on Canadian roads.

One of the ways to assess the health and skills of the elderly drivers is to conduct proper tests to assure whether they are eligible to drive on roads or not. These tests need to examine issues such as lack of vision, eye cataract, detection of moving objects, recognition of judging distance and several other age-related impairments. These tests are currently biennially in Ontario for drivers older than 80 years old. Other than conducting tests, it is also highly recommended to reduce the testing age to 50; this conservative action causes more frequent and detailed assurance of senior drivers' overall condition and reduces the severity of collisions.

As discussed previously, drivers usually have some issues regarding the visibility of the highways during the night. Due to the popularity of nighttime closures to avoid long queues and congestion, it is quite indispensable to make sure drivers can clearly distinguish work zones from a distance; this could be achieved by using an appropriate illumination system and signage. However, the government should be aware that although engineers are responsible for assuring the safety of the roads to some extent, this goal could not be achieved unless there is a constructive and fruitful collaboration between government and engineers. The government can play its role by making sure all the young and elderly

[3] Statistics Canada 2009 Report

drivers are ready to drive on the highways and engineers should satisfy the safety of the highways by effectively applying the findings of the researchers in this field, as well as, implementing guidelines and standards properly.

Next factor could be one of the biggest challenges to address and control. Consumption of alcohol or drugs is taking many lives annually. Ministry of Transportation of Ontario (MTO) stated in the "2016 Ontario Road Safety Annual Report" that 20% of fatalities are caused by drinking and driving. It is expected to see a significant increase in this percentage by involving fatalities caused by driving under the influence of drugs. Also, in 2014 report by the same organization, the alcohol and drug consumers were responsible for almost 30% of the fatal collisions. The developed models also confirm this trend. Table 5.1 shows that this factor has a positive sign, which indicates this factor decreases the probability of the PDO and increases the probability of fatalities, and it is in line with the provided statistics in Canada. Based on partial effect tables in Appendix F, more details could be concluded; the probability of PDOs and minor injuries decrease in collisions that alcohol and drug were involved and the probability of major injuries and fatalities increase significantly. All three models confirm this behaviour; however, the magnitude of this factor is higher in Logit and Arctangent models by capturing the growth in major injuries and fatalities by 37% and 6%, respectively.

Perreault published a report which emphasized on the statistics related to impaired driving in Canada [122]. In 2015, almost 75,000 alcohol and drug-impaired driving collisions were reported by Police. Perreault also claimed that young drivers aged between 20 and 24 are responsible for most of the drinking and driving collisions (480 incidents per 100,000 drivers) at 2015. Although this number is quite high, it still dropped by 36% in comparison with the recorded collision rate for same age group at 2009 (751 collisions per 100,000 drivers).

According to the significance of alcohol and drug consumption, several strategies could be adapted to reduce or preferably eliminate the effect of this factor. The first action requires the collaboration of law enforcement to have stricter rules and punishments for motorists who are not respecting the corresponding laws and regulations related to this issue. This action could take place by having more frequent sobriety checkpoints close to the interchanges when vehicles are still not merged to the highways [123]. In addition, improvement in education, as well as increasing public awareness, could also reduce the effect of this factor in future. As an example, charitable organizations such as "Mothers Against Drunk Driving"[4] in Canada are actively running campaigns and holding educational seminars to enhance the public awareness. There is a hope to have roads with no impaired drivers in

[4] Official Website of MADD

Ontario since the number of impaired drivers is in its lowest level in last 30 years [122]; this could be achieved by offering some incentives to such organizations and revising the current laws and regulations.

Developed models also demonstrated that collisions with speeds less than 80 km/h tend to cause less severe injuries. Effects of this factor could be summarized as follows,

1. Random Parameter Ordered Probit Model; this factor increases the probability of having PDO and Minor injury collisions by 13% and 18%, respectively; however, it reduces the probability of major injuries by 23% and fatal collisions by 8%.

2. Random Parameter Ordered Logit Model; this factor captured to have less effect on PDO collisions, but more on minor injury collisions. Based on the results, collisions with the recorded speed of 50 mph (80km/hr) or less increase the probability of having PDO collisions by only 3%, which this amount for minor injuries is 44.5%. On the other hand, the probability of occurrence of major injuries and fatalities reduced by 44.5% and 3%, respectively. This shows that by setting a proper speed limit, the likelihood of major injury and fatal collisions could be reduced by 47.5%.

3. Random Parameter Arctangent Model; the findings of this model were very similar to the Logit model.

One of the critical factors that is directly related to the speed is stopping distance. Stopping distance involves two major components: Reaction distance and Braking distance. Reaction distance could be defined as a distance that a vehicle travels between the moment driver detects the hazard and action of braking, and braking distance is the distance that vehicle travels from the moment driver pushes the brake pedal and the moment that vehicles get to full stop condition. Both of these distances are highly correlated with travel speed. American Association of State Highway and Transportation Officials (AASHTO) provided a table, as shown in Table 5.2, which shows the required distance for the drivers to stop their vehicles with various speeds completely. Figure 5.1 shows the trend of each component of stopping sight distance presented in Table 5.2 [124].

Table 5.2 demonstrates that the stopping sight distance increases by 85 meters when travel speed changes from 80 km/h to 100 km/h. It is worth mentioning that these distances will be significantly greater than the presented values when the pavement is wet or tires are not in acceptable condition (due to lower friction). According to this fact and the results of the developed models, the speed limits of the work zones should be modified based on the roadway condition to avoid or at least lessen the major injuries and fatalities. In addition to this action, it is highly recommended to make sure drivers are driving through the work zones with safe speed (as posted prior to the work zones), especially at nighttime. This could be implemented by proper signage and the presence of law enforcement before the work zones.

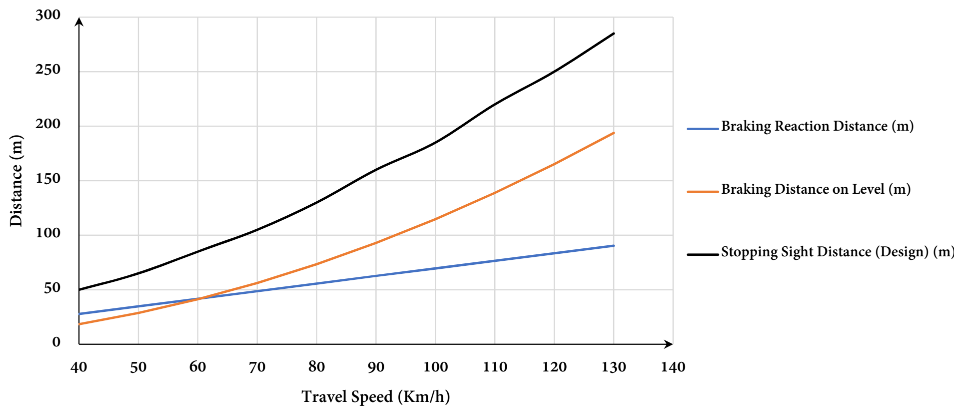


Figure (5.1) Stopping Distance for Various Travel Speeds

Table (5.2) Stopping Distance of Vehicles by Various Travel Speeds

Design Speed (km/h)	Braking Reaction Distance (m)	Braking Distance on Level (m)	Stopping Sight Distance	
			Calculated (m)	Design (m)
40	27.8	18.4	46.2	50
50	34.8	28.7	63.5	65
60	41.7	41.3	83.0	85
70	48.7	56.2	104.9	105
80	55.6	73.4	129.0	130
90	62.8	92.9	155.5	160
100	69.5	114.7	184.2	185
110	76.5	138.8	215.3	220
120	83.4	165.2	248.6	250
130	90.4	193.8	284.2	285

Based on the data and the estimated models, when the work zone is located in a section of a highway which has 4 or more lanes, the probability of fatal and major injury collisions increases. Figure 5.2 depicts, as an example from the scenario that there are more than four lanes, why the probability of experiencing conflicts due to improper lane changes in hghways with more than four lane is higher than the cases that there are fewer lanes.

Figure 5.2 shows a 5-lane highway with two lanes blocked by the work zone. In this case, five traffic conflict scenarios could be defined as listed below,

1. The conflict between grey car and green car;
2. The conflict between grey car and yellow car;
3. The conflict between green car and yellow car;
4. The conflict between yellow car and pink car; and
5. The conflict between pink car and blue car.

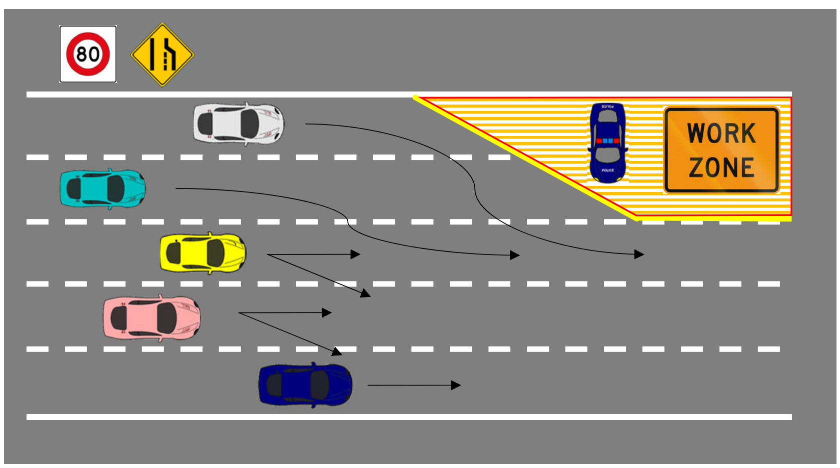


Figure (5.2) The Traffic Interruption due to Work Zones in a 5-lane Highway

However, in a very similar situation in 3-lane highways the number of traffic conflict scenarios would decrease to 3; these scenarios as shown in Figure 5.3 could be listed as follows,

1. The conflict between grey car and green car;
2. The conflict between grey car and yellow car; and
3. The conflict between green car and yellow car.

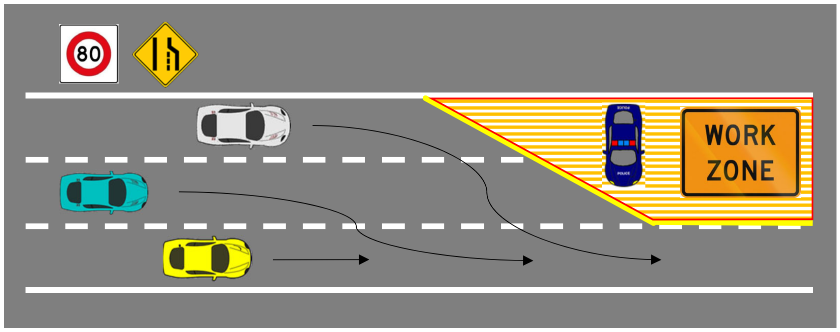


Figure (5.3) The Traffic Interruption due to Work Zones in a 3-lane Highway

In scenarios such as the one presented in Figure 5.2, any mistake by any of the drivers could lead to chain reaction collisions, which is not desirable due to safety hazards and congestion. This finding does not necessarily mean that having highways with 4 or more lanes is problematic, but it emphasizes that some additional safety consideration should be applied in work zone setups to assure the safety of the drivers and workers. It is believed that daytime closures are potentially more sensitive to this factor due to higher traffic.

The other factor could be categorized as driver-related factors. This factor shows the effect of 3 different scenarios that can statistically increase the probability of more severe collisions. These behaviours could be listed as follows,

1. Failing to keep in the proper lane; or
2. Run off the road; or
3. Tailgating (following improperly)

All these actions were found to be influential as one factor, and in the case of occurrence of any of them the likelihood of more hazardous outcomes would increase. According to the calculated effects for all three developed models, this factor had the highest impact on fatal collisions. This factor enhances the probability of having fatal collisions by 33% based on Logit and Arctangent ordered probability models. The inevitable effect of this factor led the author to investigate how the existing conditions in the work zone should be altered to minimize the effect of these scenarios. After careful evaluation, it was concluded that these scenarios are mostly the consequences of several careless actions by the driver such as improper speed, lack of attention, failing to have appropriate following distance, and impaired driving. This emphasizes that designing the work zones by only considering various safety countermeasures and engineering factors do not guarantee the safety of the highways. Educating drivers in order to make them understand the aftereffect of their actions is mandatory.

The month of the collision also has some impact on the severity of it. It is well known that in North America, especially in Canada and Northern U.S. states, usually the end of spring and summer are referred as "Construction Season". In other words, due to extreme winter seasons and limitations for setting up the work zone during wintertime, most of the maintenance, rehabilitation, and construction activities are limited to the short period between April and September.

The developed models illustrated that collisions which happened in the work zone during the first three months of construction season have the higher probability to be more severe. Interestingly, Transport Canada [125] reported that the majority of the heavy truck collisions that caused fatalities have occurred at the daytime, dry pavement, and clear weather conditions. It was also emphasized that lousy weather condition was responsible for around 20% of highway collisions. Although this percentage is quite significant, it raises a question about the remaining 80% of collisions; it needs to be studied that what was the role of the clear weather conditions in the occurrence of these collisions. This issue could be tackled by focusing on the drivers' perception of road conditions. One of the reasons which could lead to collisions under bright weather conditions is lack of attention and excessive confidence of the drivers; especially, new drivers who did not experience any conflicts or collisions during the bad weather condition. Short and concise announcements from radio stations that are emphasizing the importance of road safety could be one of the effective solutions for this issue. This method was used to increase societal awareness for other matters such as "Public Health Facts", and "Drinking and Driving".

Two of the other findings of this research discuss the effect of "Traffic Control Devices" in general, and a specific one "Warning Signs" as the independent variable. Having more frequent and well-designed traffic control devices and positioning them in locations which are noticeable by drivers can improve the safety of the work zones and highways. Based on the random parameter ordered Probit model, the probability of major injuries and fatalities reduces by 37% and 15%, respectively in case of proper use of traffic control devices. The influence of this factor on injury severity levels was found to be even higher in random parameter ordered Logit and Arctangent models. In these two models, the probability of occurrence of major injury collisions drops by 66% by proper installation of traffic control devices. Besides, by properly applying this finding, a 10% reduction in the probability of fatal incidents would be expected.

Book 7 [116], as the manual for temporary conditions of highways in Ontario, which presents the various traffic control devices used in this province. These signs are categorized in 3 major groups as listed,

- Devices for channelization, guidance, and information such as cones, construction makers, signs, and dynamic messaging signs and devices;
- Devices to regulate/control the flow of traffic such as traffic control persons, pilot vehicles, and paid duty police officers; and
- Positive protection devices such as glare screen, and temporary transverse rumble strips.

Step by step guide for installation of each of the mentioned traffic control devices is also provided in this manual. Proper implementation of this guideline should be assured by safety officers; moreover, companies who are not following the guidelines accordingly should be penalized. Although generally following the guidelines is not compulsory and sometimes following such details is expensive and time-consuming, companies should attempt to at least satisfy certain critical safety expectations to lessen the probability of hazards in the work zones. This statement becomes even more critical by knowing the fact (from chapter 3) that in almost 55% of the work zones that fatal collisions were reported, traffic control devices were not used.

Apart from any other traffic control devices, the probability of severe collisions was lower by approximately 60% in events that warning signs were installed prior to the work zones in comparison with events that warning signs were not available. This shows that drivers react more to this specific sign, and by using them more frequently prior and within the work zones, many lives could be saved annually.

Rear-end collisions, as one of the most recorded crash manners in divided highways, have negative impact on work zone safety. Based on the estimated coefficients for this factor in all three developed models, it was found that these types of collisions are generally decreasing the probability of fatalities. Nevertheless, they still cause some severe damages to the vehicles and also increase the likelihood of minor injuries.

One of the most common injury types after rear-end collisions is referred as "Cervical Ligament Injuries". When rear-end collisions occur, the neck of the drivers and other occupants of the vehicles will suffer from hyperextension and hyperflexion phenomena which might cause neck pain, headache, and other minor injuries. Nevertheless, the actual effects of these incidents on the injured people could vary based on the intensity of the

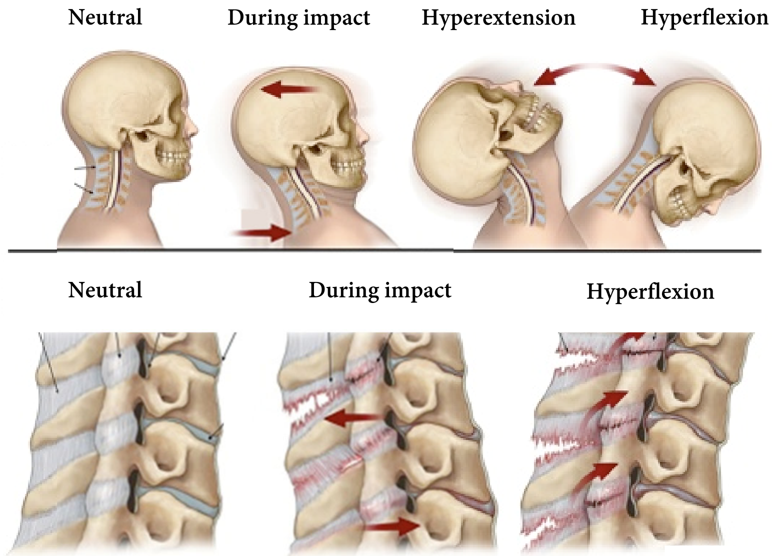


Figure (5.4) Cervical Ligament Injuries

collisions and safety features of the vehicles. Figure 5.4[5] shows how hyperextension and hyperflexion affect the ligaments of the motorists.

The interruption that work zones cause in the highways due to their design, increases the probability of rear-end collisions. Therefore, some additional safety considerations should be applied. One of the accessories that recently became popular is the soft car seat head/neck rest cushions. Using these accessories can potentially reduce the shock of the hyperextension and consequently reduce the hyperflexion reaction, which results in ligament injuries. Other solutions to prevent rear-end collisions in the work zones is to use dynamic traffic control devices which could update the drivers about the existing/special conditions of the highways and work zones. Radar enforced speed control, as well as encourage the drivers to avoid tailgating, especially close to vulnerable areas such as work zones are the other effective methods that can help avoiding rear-end collisions.

Heavy trucks are responsible for transporting goods, materials, and services from one location to another. On the average, more than 40,000 trucks travel through the busiest sections of highway 401. As mentioned in the throughput analysis, trucks usually increase the queuing in the work zones. This is not the only problem of having trucks in the work zones; random parameter ordered Logit, and Arctangent models captured that the probability of major injuries increased by 52% in collisions that trucks were involved.

[5] The source of Figure 5.4 click here

Random parameter ordered Probit model also captured 10% increment on fatalities in case of truck involvement. Federal Highway Administration (FHWA) categorized various commercial trucks by their weight in 3 major groups [126],

- Light trucks (less than 4,536 kg)
- Medium trucks (less than 11,793 kg)
- Heavy trucks (More than 11,793 kg)

According to the kinetic energy formula, the energy released in collisions could be defined as a relationship between speed (v) and weight (m) as shown in Equation 5.1,

$$Kinetic\ Energy = \frac{1}{2} m.v^2 \tag{5.1}$$

A heavy truck with 12,000 kg and speed of 100 km/h (approx. 28 m/s) releases approximately 4.63 MJ of energy; however, this energy for a regular sedan car with 1500 kg weight is as low as 0.6 MJ. Therefore, the damage caused by a heavy truck would be more severe.

Heavy truck involvement factor was also significant as a random parameter. Interestingly, heavy trucks increased the probability of more severe collisions in 83% of the collisions. In the remaining 17% of the collisions, this factor increased the probability of minor injuries and PDOs, and they could be due to scenarios such as passenger vehicles collisions with heavy trucks (rear-end). This could happen due to loss of control and poor speed management of the passenger vehicles' drivers who were tailgating the decelerating truck prior to the work zones (See Figure 5.5) [6].

Figure (5.5) Illustration of a Passenger Car Collision with a Heavy Truck

[6] Source of the Figure: www.shutterstock.com

Figure 5.6 shows the effect of this random parameter factor,

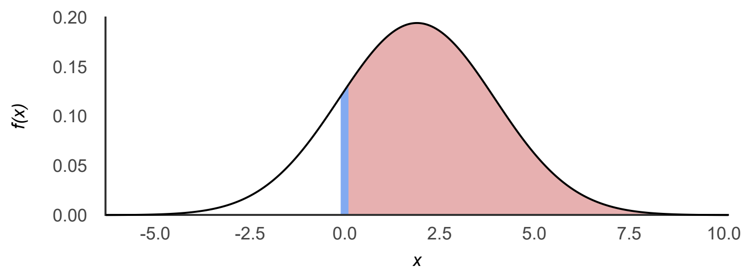


Figure (5.6) Truck Involvement Effect on Injury Severity Level as Random Parameter

Apart from rear-end collisions, rollover collisions were also found to be statistically significant as random parameters. This factor shows that in only 15% of the collisions that were not rollovers, the probability of the major injuries and fatalities increased. In other words, this emphasizes the level of danger that rollovers can cause. In 85% of non-rollover collisions, only the probability of minor collisions and PDOs increased. Rollovers could be due to loss of control, improper speed, and lack of attention of the drivers which could be limited and ideally eliminated by installing proper signs such as speed reduction warnings, work zone warnings, and lane reduction warnings prior to the work zones. Figure 5.7 demonstrates the effect of this random factor on the injury severity level,

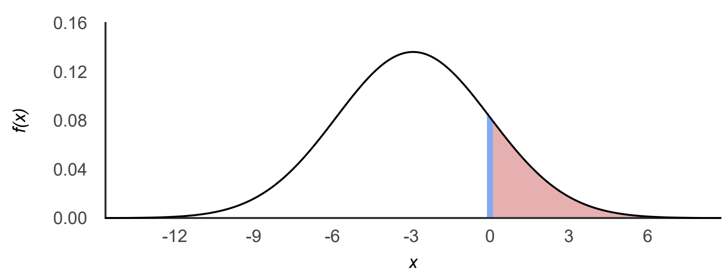


Figure (5.7) Roll Over Effect on Injury Severity Level as Random Parameter

Gender of the drivers is another statistically significant factor in the developed models for injury severity levels. According to the available data set and collected data, in 73% of the cases that the female drivers were controlling the vehicles, the probability of major injuries and fatalities decreased. This finding is in line with the descriptive statistic analysis, which was presented in the "Data Collection" Chapter. Simon and Corbett [127] conducted a study which targeted the decisions of the drivers based on their age, gender, and several other factors. One of the key findings of this research was that female drivers found to be more sensitive to road measured life stress. Considering this fact, they would be more cautious and responsive to the traffic condition caused by work zones. Marsh [128] conducted a comprehensive social and psychological study to investigate the relationship between the gender of the drivers and their driving behaviour. He claimed men are more probable to be convicted for offences such as speeding, drink-driving, and tailgating. This aggression and thrilling sensation by male drivers could be the reason of having more fatal collisions in comparison with female drivers.

Developed models for this research also showed that female drivers are tending to have less severe collisions in comparison with male drivers. For female drivers, the probability of having a fatal collision reduces by 9% under the random parameter ordered Probit model and 3% based on the other two models (Appendix F). It is also 41% less likely to have major injuries while the driver of the vehicle is female under the random parameter Logit and Arctangent models. Figure 5.8 depicts the effect of drivers' gender on the injury severity level,

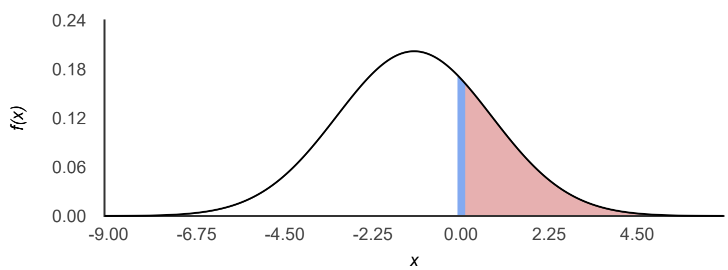


Figure (5.8) Drivers' Gender Effect on Injury Severity Level as Random Parameter

Conducting some psychological and social studies could potentially help researchers to have a better understanding of the fundamental differences between men and women drivers under various traffic conditions. This investigation would be extremely beneficial to tackle this issue and find a practical solution to avoid aggressive driving behaviour among male drivers.

5.2 Safety Models' Performance Assessment

After model development and getting some insight into the possible strategies that could be adapted to reduce the injury severity level of the work zone collisions, it is necessary to investigate the performance of these models and introduce a methodology to predict future events with random parameter models. Model performance assessment could be achieved by making sure that each of the models has the ability to predict the future events correctly and also by estimating goodness-of-fit measures; these measures could explain that equivalency of the predicted values and observed values.

Two of the most popular goodness-of-fit measures are pseudo-ρ^2 (McFadden ρ^2) and adjusted pseudo-ρ^2 (corrected [McFadden] ρ^2) as discussed in Chapter 3. Table 5.3 presents these values for each of the developed models.

Table (5.3) Model Comparison based on ρ^2 and Corrected [Adjusted] ρ^2

Goodness of Fit Measure	Random Parameter Ordered Probability Models		
	Probit	Logit	Arctangent
ρ^2	0.28	0.32	0.30
Corrected [Adjusted] ρ^2	0.24	0.28	0.25

According to calculated values, it is obvious that the random parameter ordered Logit model, which used Logistic distribution, for parameter estimation performs better than the other two models. It is worth mentioning, although ρ^2 and adjusted ρ^2 seem low, they could be still considered as rather satisfying results for ordered models. One of the ways to have a better understanding of these measures is to compare them with R^2 values. Figure 5.9 presents the empirical relationship between ρ^2 and R^2 [103]. Based on Figure 5.9, the random parameter ordered Logit model's ρ^2 and corrected ρ^2 are approximately equivalent to R^2 of 0.8, which could be considered as very good for ordered probability models [129].

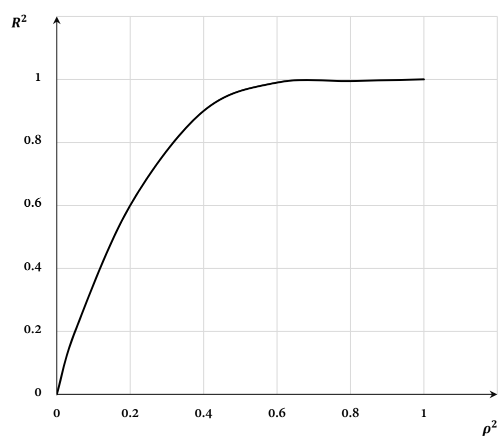


Figure (5.9) ρ^2 and R^2 Empirical Relationship

Another measure which could be taken into account is the Akaike Information Criterion (AIC). This estimator is able to perform a relative comparison between existing models and select the best model. The idea behind the development of AIC estimator is coming from information theory. Since statistical models are developed based on historical data, it is not possible to capture all the details of previous events by these models, and they lose some information while they attempt to describe the events. This estimator tries to quantify the amount of lost information. Therefore, the lower the AIC means a better model. This measure could be calculated by using Equation 5.2,

$$AIC = 2k - 2Ln(\hat{L}) \tag{5.2}$$

where k is the number of factors used in the model, and $Ln(\hat{L})$ is the log-likelihood at convergence.

Similar to AIC, Bayesian Information Criterion (BIC) is also developed to compare the several models based on their log-likelihood at convergence. This criterion introduces a penalty for the number of used factors in the models, and this penalty is more substantial in BIC in comparison with AIC. BIC could be calculated by using Equation 5.3,

$$BIC = Ln(n)k - 2Ln(\hat{L}) \tag{5.3}$$

where n is the number of observations, and the rest of the parameters were described earlier for AIC.

The lowest value for this measure is more preferable, but in order to conduct a more robust comparison, Kass and Raftery [130] suggested to calculate the ΔBIC, and higher differences mean stronger evidence for preferring one model over the other candidate. The decision-making boundaries have been presented in Table 5.4,

Table (5.4) Meaning of Differences Between BICs

ΔBIC	**Evidence against higher BIC**
0 to 2	Not worth more than a bare mention
2 to 6	Positive
6 to 10	Strong
>10	Very strong

Table 5.5 presents the calculated AIC and BIC values for the developed models. Also, according to calculated values, the lowest AIC and BIC values belong to the random parameter ordered Logit model.

Table (5.5) AIC and BIC values for Developed Models

Distribution	**Probit**	**Logit**	**Arctangent**
Number of Factors		19	
Number of Observations		256	
Log-Likelihood at Convergence	-265.712	-254.662	-260.385
AIC	569.424	547.324	558.77
BIC	636.782	614.682	626.128

Also ΔBICs were calculated as shown in Table 5.6. Since the differences of BICs are higher than 10 [130] (See Table 5.4) and random parameter ordered Logit model has the lowest BIC value; it could also be concluded that this model is the best one among all of the developed models.

Table (5.6) Differences in BIC values

ΔBIC	**Differences**
BIC_{Logit}-BIC_{Probit}	22.100
BIC_{Logit}-$BIC_{Arctangent}$	11.447

5.3 Cross Validation of the Selected Model

One of the statistical methods that are commonly used in machine learning to validate the model results is called cross-validation. The main idea behind performing cross-validation is to compare and choose the best model in terms of predictivity and avoid the over-fitting issue. This also helps to reduce the bias in the models.

One of the most simple cross-validation methods is referred to as "Holdout Method". In this method, the data set is usually divided into two different sets, training and testing sets. The training set will then be used as the input for the model development procedure.

When the model is finalized, the testing set could be used to assess the prediction ability of the model by considering the accumulated errors [131].

k-fold validation is the modified version of the holdout method. In this approach, the data set would be divided into k subsets, and the conventional holdout method should be repeated for k times. k-fold validation uses $k - 1$ of the subsets to train the model and uses the remaining subset as the testing data set. Finally, the average errors of each trial would be calculated to define the overall performance of the models. One of the notable advantages of this method is its independence on how the subsetting process occurs. In other words, each data point will be used $k - 1$ times in training procedure, and once as testing data. By increasing the number of folds (k), it is more probable to have lower variance and more consistent results. The only disadvantage of this approach is its computational cost, especially when k is equal to 10. Under this setup, ten independent models should be developed, which in comparison to the estimation of only one model is more computationally expensive; however, this would offer the luxury of reducing the concerns regarding overfitting. Figure 5.10 presents a schematic of this procedure [131].

Since ordered probability models are susceptible to distribution type used for parameter estimation and proportion of the data in each category, it should be assured that each of the folds have the same characteristics and proportion as the original data set. Dividing data in a way that each fold has the same characteristic as the original data set is known as stratified k-fold cross-validation. Details of this idea are presented in Figure 5.11.

According to the discussed methodologies, the data set was divided into ten independent folds, and each of those groups had the same proportion of PDOs, minor injuries, major injuries, and fatalities. Following steps has been followed to have unbiased folds,

1. Each of the injury severity classes was saved to an independent spreadsheet.
2. **RAND** function in Microsoft Excel was used to allocate a random number to each observation in each class and then sorted in ascending order.
3. Total number of data points in each class was divided by 10 to decide on the selection size.
4. By using steps 2 and 3, grouping was conducted.
5. **RAND** function in Microsoft Excel was used for one more time to assign a random number from 1 to 10 to each group in each injury severity class.
6. The groups with the same assigned number in each class was matched, and the fold was created.

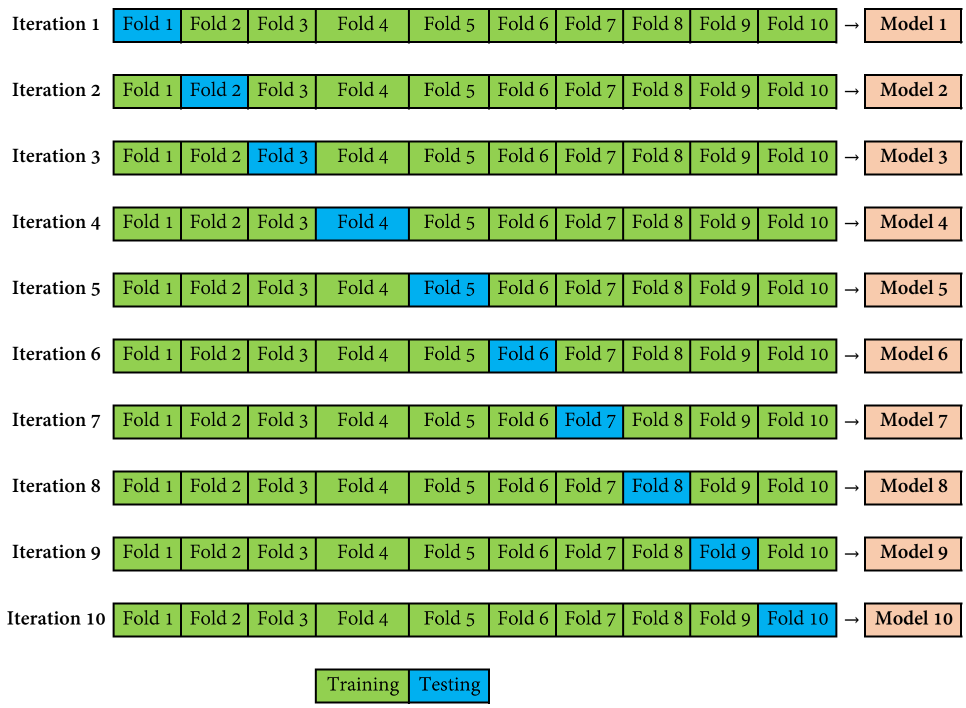


Figure (5.10) 10-Fold Validation Methodology

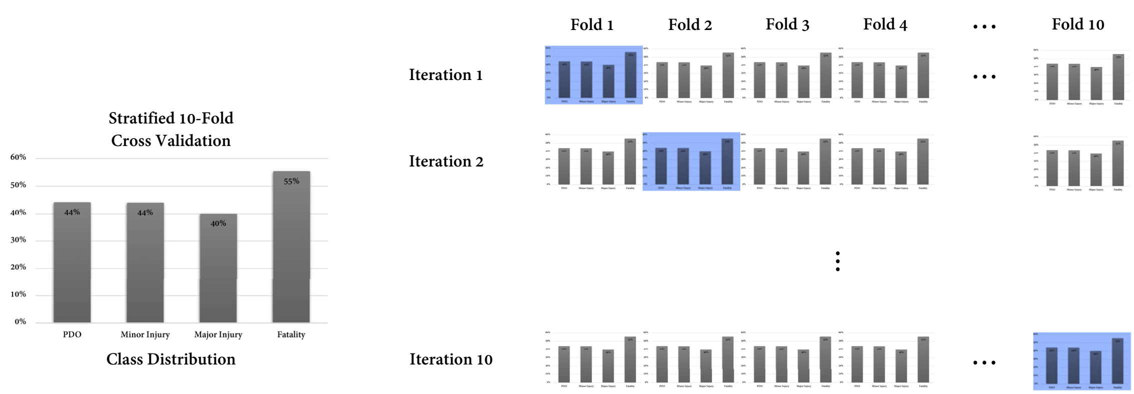


Figure (5.11) Stratified 10-Fold Validation Methodology

Using the framework as noted earlier, the steps for preparing the data for stratified k-fold cross-validation are presented in Figure 5.12.

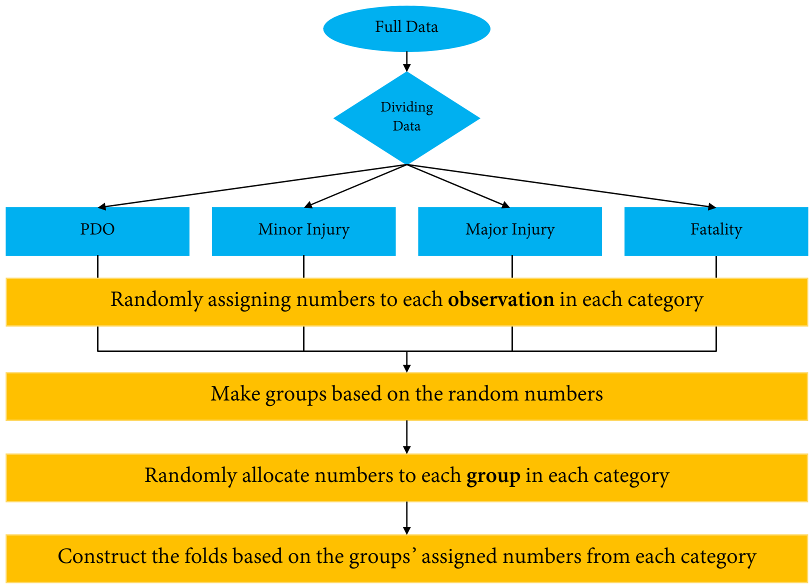


Figure (5.12) Data Preparation for Stratified k-fold Cross Validation

The next step after defining the folds is to develop independent models for each of those folds. Under the defined circumstances, ten random parameters ordered Logit models were developed. Appendix G contains all of the developed models; the detailed information about the used factors and their significance are presented for each fold in its corresponding table . Interestingly, when the data set was divided into smaller portions, some other factors which were not statistically significant in the full Random Parameter Ordered Logit model (See Table 5.1) turned out to be significant. These factors are listed as follows,

- In some of these ten models, it was found that if the collisions occur between December and March, the probability of having collisions with lower severity increases. This finding could be due to two major characteristics in the available data set,

    1. Lower number of collisions in the work zones in comparison with construction season; usually municipalities and construction companies do not schedule their work zones in the wintertime because of harsh winters. Consequently, lower accident frequency with lower severity could be expected.

2. Drivers are more cautious during the wintertime, and they drive with lower speed. Also, winter tires can help them to decelerate more rapidly.

- It was also found that work zones that are close to junctions (referring to both intersections and interchanges as is defined in the raw data) are more likely to have more fatal and major injury collisions. Figure 5.13 shows one of the scenarios which can cause a conflict and/or a collision. Junctions are usually connecting one freeway to another or the rest areas; therefore, the speed of the merging vehicles are yet not adjusted with the existing flow of the highway; this could potentially increase the probability of more severe collisions. In high-risk locations of the highways (e.g. around the junction) more conservative closure strategies and equipment such as concrete barriers and better illumination system are highly recommended.

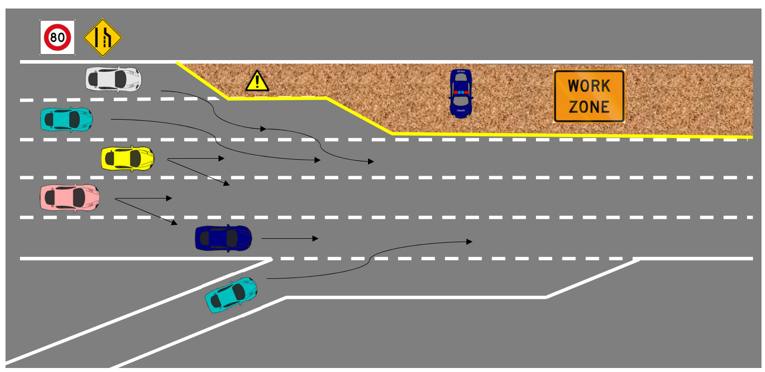


Figure (5.13) Potential Work Zone Collision Scenario in a Segment with a Junction

- In only one of the models (Fold 7 Model), the location of the collision was observed to have a significant impact on the severity of the collisions. This factor shows collisions, which occurred in the 'Median' or 'Shoulder' of the highways are increasing the likelihood of fatal and serious injury collisions. Workers usually work or rest around the work zone area and close to shoulders or medians (depend on the location of the work zone [right or left]). Under this circumstance, workers are highly exposed to get hit by passing vehicles. Lower speed limits, designing a proper resting area for workers, and stricter rule that prohibit workers to stand or rest close to passing vehicles, and using concrete barriers could be some of the solutions for this issue.

5.4 Cross Validation Models' Prediction Strategy and Performances

Prediction performance assessment of the random parameter ordered probability models is quite challenging, and finding a way to use the most efficient coefficients is vital. As mentioned in earlier chapters and sections, in random parameter modelling the number of estimated coefficients for the random parameter would be equal to the number of observations (coefficients change across the observations) and based on the chosen distribution and scenario, this coefficient might change. One critical question relates to which coefficient should be used for future events' prediction while the unobserved heterogeneity for various injury severity levels might vary in nature and magnitude. In earlier studies, researchers used the average value of the estimated coefficients for a single factor and used that for prediction, which logically is not inline with the whole ideology of the random parameter model. By using only the average value, estimating several coefficients for various scenarios seems not efficient and beneficial. Therefore, the author has decided on how to overcome this issue.

One of the significant contributions of this study is to tackle the mentioned concern by defining a factor referred as "Similarity Level". This factor compares the current scenario with any of the historical scenarios and selects the scenarios with more than 75% similarity; then, the corresponding coefficients for those scenarios will be used for that specific event (collision). Figure 5.14 demonstrates the algorithm of this process. Each table in Figure 5.14 shows a particular step of the prediction process. Each of these steps is summarized as follows,

1. Table a: In this step, the data from each significant factor and scenario was summarized in a table, and the random ones was marked to be distinguished for future steps (training data set for each fold).
2. Table b: This table summarizes the information about the current situation, environment, and facts about future work zone. Filling this table provided the context for conducting comparison (test data set for each fold).
3. Table c: This table counted the similar factors between each of the 'Table a' data and 'Table b'. In case of similarities, more than 75% in each specific category e.g. PDO, the coefficient for the random parameter was recorded in 'Table c'.

4. Table d: This table gathered all coefficients from scenarios that are sharing more than 75% similarity to the future event, and these coefficients were used for prediction purposes.

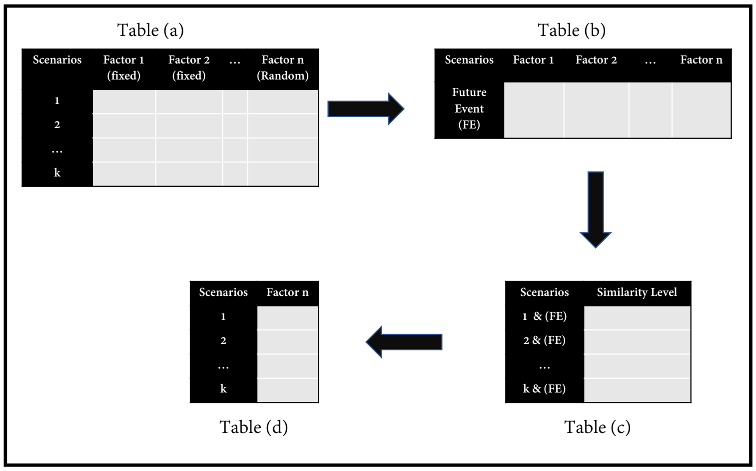


Figure (5.14) Prediction Algorithm for Random Parameter Models

Following this strategy offers the ability to select coefficients from the most similar cases for the sake of prediction and leads to more accurate results. The prediction for the training data set is more straightforward than the test data set due to having the coefficients of the random parameter for each case individually; therefore, more accurate results were expected for training data set. The results based on this approach has been shown in Table 5.7, which summarizes the prediction results (both Training and Test) from Random parameter Logit models [for each fold individually]. According to the presented results in Table 5.7, several conclusion could be drawn which are listed as follows,

- As it was expected, predictions for training data sets are significantly more accurate than the test data sets. This shows the importance of selecting the best coefficients for the sake of future prediction. It is worth mentioning that although the predictions are still not as accurate as with the training data set, they are more accurate than traditional random parameter prediction method.

- Approximately 70% of fatal collisions from test data sets has been predicted correctly by using the recommended methodology. Also, 20% of the fatal collisions were predicted as major injuries. The prediction accuracy of fatal collisions was more than 95% for the training data set.

- More conservative predictions were made by this algorithm for PDO collisions. These conservative predictions could generally be more beneficial to enhance the safety standards of work zones and highways.

- Minor and major injuries were predicted fairly accurate.

Table 5.8 shows the summary of 10-fold cross validation analysis and the accuracy of predictions in each category.

Among all of these classes, fatalities could be considered as the most critical category. This could be due to 2 primary reasons; the first reason is that fatalities are quite expensive for society and they induce lots of direct and indirect costs for the government. Capital Region Intersection Safety Partnership [132] conducted a study in 2017, and they have concluded that each fatal collision cost more than 225,000$ as a direct cost. This report also provided some statistics related to these costs. They estimated that the current cost is almost 27% higher than in 2010. The second reason is more related to the social impact of each fatality and the consequences of thess unpleasant events for families of the victims. Thus, minimizing the likelihood of fatal collision occurrence is vital, and it needs the collaboration of various organizations as well as the government.

Since the importance of predicting injury severity class is clarified, there is a need to investigate the significant reasons that are causing some of the mispredictions by the developed models. In order to have a deeper understanding of this issue, the author went through every single mispredicted data point and came to the following conclusions,

- Equation 3.17 is usually used to calculate the z for ordered probability models; then this z will be compared to the thresholds in order to decide on the injury severity level. In many of the mispredicted scenarios, the z value was slightly lower than the threshold, and consequently, it was predicted in the lower category. This could be considered as one of the limitations for these type of models.

- In some of the other scenarios, it was captured that an 'Unexpected Death' had occurred. For example, the model predicted the injury severity of a collision as PDO and the current information is also supporting this decision; however, the severity of this collision was recorded as fatality. In other words, based on the pieces of evidence, the collision was not intense or harsh enough to cause fatality; however, because of some unknown reason, it turned out to be a fatal collision. This could be due to 2 elemental reasons: first, humans are very vulnerable and sometimes even minor

Table (5.7) Prediction Performance of Each Fold [Fold 1 to Fold 3]

Training Dataset — Fold 1

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	69	0	0	0	69
Minor	0	52	0	0	52
Major	0	0	32	2	34
Fatalities	0	0	2	72	74
Total (Predicted)	69	52	34	74	229

Test Dataset — Fold 1

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	3	5	0	0	8
Minor	0	6	0	0	6
Major	0	1	3	0	4
Fatalities	0	1	1	7	9
Total (Predicted)	3	13	4	7	27

Training Dataset — Fold 2

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	69	0	0	0	69
Minor	0	52	1	0	53
Major	0	0	33	1	34
Fatalities	0	0	1	74	75
Total (Predicted)	69	52	35	75	231

Test Dataset — Fold 2

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	4	2	2	0	8
Minor	0	3	2	0	5
Major	0	1	2	1	4
Fatalities	0	3	0	5	8
Total (Predicted)	4	9	6	6	25

Training Dataset — Fold 3

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	69	0	0	0	69
Minor	0	52	0	0	52
Major	0	0	32	2	34
Fatalities	0	0	1	74	75
Total (Predicted)	69	52	33	76	230

Test Dataset — Fold 3

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	3	4	1	0	8
Minor	0	5	1	0	6
Major	0	0	4	0	4
Fatalities	1	0	2	5	8
Total (Predicted)	4	9	8	5	26

Table (5.7) (Cont'd) Prediction Performance of Each Fold [Fold 4 to Fold 6]

Training Dataset — Fold 4

	PDO	**Minor**	**Major**	**Fatalities**	**Total (Actual)**
PDO	68	1	0	0	69
Minor	0	52	1	0	53
Major	0	0	33	2	35
Fatalities	0	0	1	74	75
Total (Predicted)	68	53	35	76	232

Test Dataset — Fold 4

	PDO	**Minor**	**Major**	**Fatalities**	**Total (Actual)**
PDO	5	2	1	0	8
Minor	1	3	0	1	5
Major	0	1	2	0	3
Fatalities	0	0	1	7	8
Total (Predicted)	6	6	4	8	24

Training Dataset — Fold 5

	PDO	**Minor**	**Major**	**Fatalities**	**Total (Actual)**
PDO	69	1	0	0	70
Minor	0	51	1	0	52
Major	0	0	32	2	34
Fatalities	0	0	1	73	74
Total (Predicted)	69	52	34	75	230

Test Dataset — Fold 5

	PDO	**Minor**	**Major**	**Fatalities**	**Total (Actual)**
PDO	5	1	1	0	7
Minor	1	4	0	1	6
Major	0	1	2	1	4
Fatalities	0	0	2	7	9
Total (Predicted)	6	6	5	9	26

Training Dataset — Fold 6

	PDO	**Minor**	**Major**	**Fatalities**	**Total (Actual)**
PDO	69	0	0	0	69
Minor	0	47	5	0	52
Major	0	0	33	1	34
Fatalities	0	0	3	72	75
Total (Predicted)	69	47	41	73	230

Test Dataset — Fold 6

	PDO	**Minor**	**Major**	**Fatalities**	**Total (Actual)**
PDO	2	5	0	1	8
Minor	2	2	2	0	6
Major	0	1	2	1	4
Fatalities	0	0	2	6	8
Total (Predicted)	4	8	6	8	26

Table (5.7) (Cont'd) Prediction Performance of Each Fold [Fold 7 to Fold 9]

Training Dataset — Fold 7

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	68	1	0	0	69
Minor	0	52	0	0	52
Major	0	0	32	2	34
Fatalities	0	0	1	73	74
Total (Predicted)	68	53	33	75	229

Test Dataset — Fold 7

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	4	4	0	0	8
Minor	0	4	1	1	6
Major	0	1	2	2	5
Fatalities	0	1	2	6	9
Total (Predicted)	4	10	5	9	28

Training Dataset — Fold 8

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	69	1	0	0	70
Minor	0	52	0	0	52
Major	0	0	33	2	35
Fatalities	0	0	1	74	75
Total (Predicted)	69	53	34	76	232

Test Dataset — Fold 8

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	3	4	0	0	7
Minor	0	4	2	0	6
Major	0	0	2	1	3
Fatalities	0	1	2	5	8
Total (Predicted)	3	9	6	6	24

Training Dataset — Fold 9

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	69	0	0	0	69
Minor	0	51	1	0	52
Major	0	0	32	2	34
Fatalities	0	0	1	74	75
Total (Predicted)	69	51	34	76	230

Test Dataset — Fold 9

	PDO	Minor	Major	Fatalities	Total (Actual)
PDO	5	3	0	0	8
Minor	0	4	2	0	6
Major	0	1	2	1	4
Fatalities	0	2	1	5	8
Total (Predicted)	5	10	5	6	26

Table (5.7) (Cont'd) Prediction Performance of Each Fold [Fold 10]

Training Dataset						Test Dataset					
Fold 10						Fold 10					
	PDO	**Minor**	**Major**	**Fatalities**	**Total (Actual)**		**PDO**	**Minor**	**Major**	**Fatalities**	**Total (Actual)**
PDO	70	0	0	0	70	**PDO**	5	1	1	0	7
Minor	0	51	1	0	52	**Minor**	0	5	1	0	6
Major	0	0	32	2	34	**Major**	0	1	3	0	4
Fatalities	0	0	1	74	75	**Fatalities**	1	0	2	5	8
Total (Predicted)	70	51	34	76	231	**Total (Predicted)**	6	7	7	5	25

Table (5.8) Accuracy of Each Category Based on 10-fold Cross Validation

	PDO	**Minor**	**Major**	**Fatal**
Average percentage of correct predictions	51.1 %	72.7%	72.3%	69.8%

head injuries could result in fatalities; and the second reason could be lack of data. Information related to the pavement condition and geometric design of highways could be very useful in this study.

5.5 Smart Form Development

In the previous section, some of the concerns related to the prediction of the models were presented. One of the main reasons that these models could not be directly used is that these models only predict one outcome for a single data point; offering a definite prediction for the scenarios that human behaviour is involved is both challenging and somewhat not realistic. Therefore, the author recommends a new point of view for predicting future events by using all of the models that were developed for 10 fold validation. Figure 5.15 depicts the methodology that was adapted to implement the new approach. This methodology could be divided into six steps as follows,

Step 1- Development of a smart form: Complicated and hard to use statistical models sometimes make contractors and companies less eager to use the findings of academia. One of the major steps which should be taken by academia is to translate their findings into a

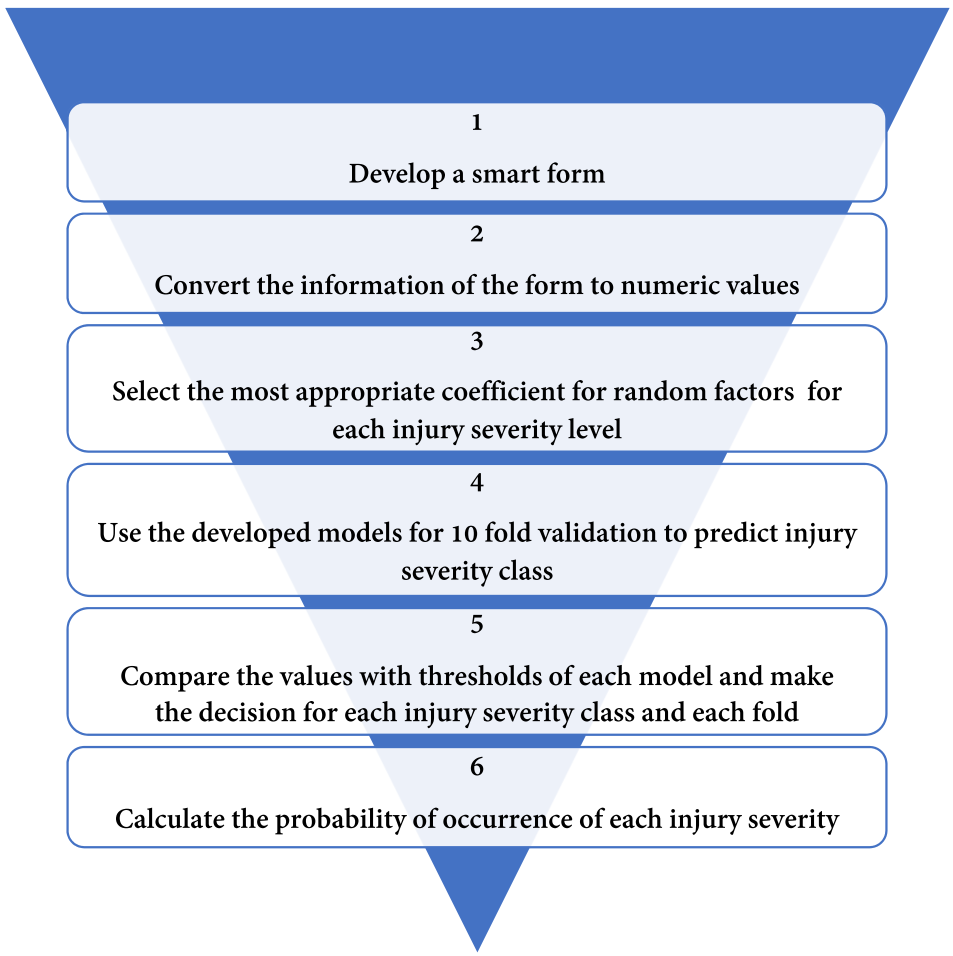


Figure (5.15) Injury Severity Class Prediction Strategy Using 10 Fold Validation Models user-friendly format. The prepared, smart form is connected to all of the developed models and contractors could easily use it in their time of need. This helps them to have a better understanding toward the possible scenarios in future work zones. Figure 5.16 shows a sample form with its inputs.

No.	Questions	Answers
1	When is the planned closure time?	Rest of the day
2	What the average age of the drivers in the region?	Between 25 and 50
3	Is you construction zone close to areas that might be affected by impaired drivers? (Downtown, City Centre, etc)	Yes
4	What is the precieved speed in or around the construction zone during the work period? (Not the posted speed)	Less than 50 MPH
5	Does the work zone section of the highway has 4 or more lanes?	No
6	Are there any reports regarding the slipery road condition, inappropriate curves, low visibility or any other issues in or around the work zone which might cause issues such as immediate lane changes, run-off-road, etc?	Yes
7	Are you planning to use traffic control devices during the closure?	Yes
8	Based on your previous experiences, what is the most probable crash type in similar closures?	Rear-end
9	which month are you planning to set up your work zone?	Between December and March
10	Are you going to use "Warning Sign" before and after your closures as described in the guidelines [Book 7]?	Yes
11	Is the closure closure to junctions?	Yes
12	What is the pavement surface condition at the time of the closure?	Dry
13	Is there any possibilities to have collisions to have collisions in the median or shoulder of the road? [e.g. Narrow lanes, no left shoulder, etc]	Yes
13	Is the area has ageomteric design deficiencies which might lead to roll-overs?	Yes
14	What is the dominant driver gender in the area?	Female
15	Does this section accomodates a high percentage of heavy trucks [more than 25%]?	Yes

Figure (5.16) An example for Step 1

Step 2- Conversion of information: When users fill the form, the information need to be translated into numeric values in order to be inserted into the developed statistical models. This step is responsible for this conversion. Figure 5.17 depicts a sample of this procedure.

Data Translation																	
Constant	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17
1	0	0	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1

Figure (5.17) An example for Step 2

Step 3- Selection of the most appropriate coefficients for random factors: One of the challenges for predicting future events is to find the most appropriate coefficients for the random factors. This selection procedure is discussed previously, and its procedure was shown in Figure 5.14. As mentioned earlier, in 10 fold validation, 90% of data was used for modelling, and 10% was used for testing purposes, and the prediction results from the models were comparable with the actual injury severity class (The actual injury severity is known for the test data set). However, this is almost not possible for the events that they have not occurred yet to check their prediction validity. Thus, it would be more logical to use the average of the test data sets' coefficients for each injury severity level. This offers the luxury of accounting for any possible injury severity class individually. Figure 5.18 shows a sample of this step for fold 2.

	Fold 2		
	Roll over	Gender	Heavy Truck
PDO	-2.585	-2.041	2.589
Minor Injury	-2.628	-2.355	2.602
Major Injury	-2.463	-2.065	2.597
Fatality	-2.535	-1.957	2.601

Figure (5.18) An example for Step 3

Step 4- Injury severity level prediction by developed models: After setting all coefficients for random factors in step 3, statistical models that are developed for each fold could be used to predict the potential future event. In this stage, an overall of 40 predictions will be made. Figure 5.19 shows a screenshot of a specific case for this step.

Coefficients from	Predictions									
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
PDO	-1.075	8.753	14.462	1.241	11.612	11.185	0.516	6.934	1.426	2.335
Minor Injury	-1.248	8.408	14.610	1.313	11.370	11.161	0.512	6.923	1.214	2.230
Major Injury	-1.043	8.859	14.459	1.249	11.400	10.637	0.369	6.592	1.275	2.615
Fatality	-0.954	8.899	14.786	1.281	11.731	11.189	0.653	6.953	1.173	2.369

Figure (5.19) An example for Step 4

Step 5- Deciding on injury severity level: In this step, results from each injury severity level model will be compared to the estimated thresholds of each fold's random parameter ordered Logit model. The outcome of this step could show how each model predicted a specific scenario. Figure 5.20 shows the sample of the outcome table of this step.

Step 6- Calculation of probability of each injury severity class: In this final step, the probability of occurrence of each injury severity level could be calculated by dividing

Coefficients from	Predictions									
	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
PDO	Minor Injury	Fatality	Fatality	Minor Injury	Fatality	Fatality	Minor Injury	Major Injury	Major Injury	Minor Injury
Minor Injury	Minor Injury	Fatality	Fatality	Minor Injury	Fatality	Fatality	Minor Injury	Major Injury	Major Injury	Minor Injury
Major Injury	Minor Injury	Fatality	Fatality	Minor Injury	Fatality	Fatality	Minor Injury	Major Injury	Major Injury	Minor Injury
Fatality	Minor Injury	Fatality	Fatality	Minor Injury	Fatality	Fatality	Minor Injury	Fatality	Minor Injury	Minor Injury

Figure (5.20) An example for Step 5

the number of each injury severity class by the total number of predictions. Figure 5.21 demonstrates an example of a sample outcome of this procedure.

Frequency of each class	Probability of having each injury severity level
PDO	0.1
Minor Injury	0.375
Major Injury	0.125
Fatal	0.4
Total	1

Figure (5.21) An example for Step 6

Ordered probability models usually have only one outcome with the highest probability among other options. Ten models from cross-validation process were used as the basis of the smart form. Among all developed models, some were conservative and strict, but some were more flexible; therefore, it is expected to have a fair distribution of predictions for each injury severity level. One of the advantages of the described procedure is to have more realistic predictions for future events; which was one of the core objectives of this research.

5.6 Summary

The main idea of this chapter was to develop 3 different random parameters ordered probability models which can investigate the most influential factors on work zone safety as well as injury severity of the collisions. According to these models, it has been identified that work zone configurations and conditions, socio-demographic information, and timing of the closures have an arithmetical impact on the injury severity level. Also, the intuitiveness of all factors has been discussed, and some possible solutions for enhancement of the work zones' and highways' safety for implementation have been recommended.

Table 5.1 summarized all the details of the estimated models, and effect of each factor for each model is presented separately in the Appendix F tables. The summary of the findings could be listed as follows,

- **Random Parameter Ordered Probit Model(RPOPM)**
  - Failing to keep in the proper lane, run off the road, and tailgating are increasing the probability of major injury collisions for 31%. This factor could be considered as the most critical factor to deal with; using proper signage prior to the work zones, increasing the drivers' awareness by education, and presence of law enforcement ahead of these vulnerable areas could significantly reduce the risk of major injuries in the work zones.
  - Presence of proper traffic control devices prior and within the work zones was also found to have a significant impact on the reduction of major injury collisions.
  - Highways with four or more lanes are more susceptible to the occurrence of fatal collisions. Higher probability of conflicts and collisions prior to the work zones due to the improper lane changes, lack of control, and excessive speeding could be the underlying reason for this factor's significance.
  - Rollover collisions are enhancing the probability of fatal collisions by 54%. Designing the work zones in a safer manner by following the available guidelines, and improving them with the academic findings could lessen the risks of rollover occurrence in the work zones.
- **Random Parameter Ordered Logit Model (RPOLM)**
  - The hour of the collision was found to be the most vital factor by increasing the probability of major injury collisions for 59%. More impaired drivers, low visibility, and higher driving speed could be identified as the main reasons for this factor's significance. Improving the illumination system, presence of Police prior to the work zones, and using dynamic traffic devices could reduce the effect of this factor in future.
  - Similar to the RPOPM, traffic control devices and more specifically "Warning Sign" could dramatically reduce the probability of major injury collisions.
  - Analogously to the RPOPM, highways with 4 or more lanes has the highest share in enhancing the fatal collisions probability.

- Any strategies and solutions which can avoid rollovers should be taken to the account. By eliminating this factor, the chance of having fatalities in the highways would reduce by 74%.

- **Random Parameter Ordered Arctangent Model (RPOAM)**
  - Findings from RPOAM model are relatively equivalent to the results of RPOLM, although the effects' magnitude is slightly different among these models.

The next step was to select the best model based on two criteria: Goodness-of-fit and prediction performance. After calculations, random parameter ordered Logit model has been selected as the most appropriate model for this study, with the highest rate of correct predictions, and (meaningful) lower AIC and BIC values. In order to have more accurate performance of the selected model, 10 fold validation technique was used to see whether the results of the initial random parameter ordered Logit model is reliable or not. Throughout this procedure, 'Similarity Level' of each scenario between the test data set and training data set was measured; and the most efficient coefficients for random factors were selected for future events predictions. Findings of these steps were used as a basis to construct a smart form which is connected to the developed models. This smart form gives the luxury of having a probabilistic view of future events, and it is believed that it could be a useful tool for designing a safer work zones in future.

Chapter 6

Conclusions, Contributions, Recommendations and Future Works

6.1 Conclusions and Recommendations

Canada has the eighth largest roadway network in the world consisting of more than 415,000 km of paved roads and more than 600,000 km of unpaved roads [1]. Roads exposed to a high-volume of heavy truck traffic, especially highways, experience significant pavement deterioration; often this deterioration is increased due to harsh Canadian winters. Therefore, the need for maintaining the serviceability of the roads in an efficient way would increase. In Ontario, the MTO is the leading authority who manages, schedules and sets up work zones to preserve and improve the overall condition of the roads.

Setting up work zones interrupts the regular flow of traffic, and it can cause two major issues: congestion and collisions. Throughout the years, researchers all around the world have tried to address these issues, and they have offered several solutions using historical data. Although using historical data is vital, but it should be remembered that future events are even more critical than the ones that they have already occurred. One method to overcome the limitations of data is the use of random parameter models. This innovative technique aims to eliminate the unobserved heterogeneity issue. However, predictions made using this approach could be very challenging as the original prediction method for random parameter models does not take advantage of the variability of coefficients across the observations. To that end, this research was initiated to recommend solutions for the

[1] The World Factbook by Central Intelligence Agency

mentioned gaps, and the conclusions of each chapter of this research will be discussed individually.

Chapter 1 gives a brief introduction about the background, research hypotheses, research objectives and scope. This chapter also discusses the main ideas behind this study and the expected results from it. **Chapter 2** as the literature review section of this thesis goes through several previous studies that have been conducted by other researchers. This chapter starts by introducing various road classifications, followed by maintenance, preservation, and rehabilitation strategies and concepts. Later in this chapter, work zone configurations and types are explained. Conventional work zone simulation techniques are discussed, and alternative work zone performance assessments, e.g. statistical modelling, are reviewed. **Chapter 3** proposes the methodologies that have been adapted to prove the hypotheses of this research. Various linear and non-linear regression approaches, techniques for model performance assessment, and several other details are discussed in this section. Also, it presents the data used for this research as well as its collection procedure and the sources that have been used for the modelling. This chapter describes how the throughput data was collected by the help of students and collaborators from the University of Waterloo, University of Toronto, CIMA+, and Ministry of Transportation Ontario. Also, it describes the practical challenges for gathering the work zone safety data. Analyses in **Chapter 4** and **Chapter 5**, as the main body of this thesis, discuss the significant factors in all of the developed models for both throughput and safety analysis. The goal of these chapters were, first, to demonstrate the influential factors which reduce congestion and improve the safety condition of work zones; and then to evaluate how each of these factors affected the outcomes of each model. These discussions are followed by multidisciplinary recommendations which could be potentially beneficial for work zone safety and traffic improvement. A new methodology for predicting with random parameter models was also recommended at Chapter 5. Finally, to have a more realistic perception of future, a smart form with probabilistic basis was developed. **Chapter 6** as the final chapter of this study aims to summarize all the findings of this research and present the contributions of the author.

Developed models for work zone throughput analysis have proven that work zone configuration, location, and time of closure have a significant effect on queuing. The random parameter negative binomial model results illustrated that multiple short work zones are more efficient than longer work zones where there are two or more lanes are closed. Reduced queuing times and providing space such that motorists can adapt their speed with this configuration allowed 177 more vehicles to pass safely. Among all work zone set up scenarios, nighttime closures during weekends had the least queuing frequency in comparison with other daytime and nighttime closures occurring during the weekdays by allowing extra 493 vehicles per hour per lane to pass. Highways such as 401 that have more than 20%

heavy trucks experienced more queuing; therefore, prohibiting truck traffic in work zones and designing appropriate detours could significantly improve the congestion level. Also, factors that were found to be statistically significant for determining the injury severity of the collisions demonstrated that not only engineers but other parties are also responsible for satisfying the safety of the work zones and highways. Government, through the enforcement of laws and regulations, and motorists by being more cautious and conscientious can enhance the safety of roads. As an example, motorists who consumed alcohol and drugs have an increased likelihood of fatal and major injury collisions by approximately 36%. Educating young drivers and establishing stricter laws and regulations, as two of the responsibilities of the government, could minimize the effect of this factor significantly. Also, aggressive driving behaviours by careless motorists increased the probability of major and fatal collisions by 78%. These examples, as part of the findings, emphasize on the importance of meaningful collaboration of government, motorists and engineers. From a statistical point of view, negative binomial models, due to their advantage in considering the discrete nature of traffic throughput, were found to be more efficient and accurate than multiple linear regression models. Moreover, it has been shown that using various distributions in the estimation of ordered probability models significantly changes the outcomes of the models. As it was expected, random parameter models presented more reliable and robust results in comparison with fixed parameter models as these models estimate multiple coefficients for random factors which vary across observations; this advantage accounts for unobserved heterogeneity. This study recommended a new methodology in selecting the most appropriate coefficient among all of the estimated coefficients for random parameters for the sake of future event prediction. Last but not least, all of the mentioned findings and models found for the safety of the work zones were summarized in a "smart form." This smart form is an easy-to-use platform that offers a practical starting point for contractors and engineers to determine the severity of potential collisions in work zones.

According to the findings of this research, the author offers several recommendations regarding the work zone setup, and way to improve their efficiency and safety. Some of the recommendations are summarized as follows,

- According to the **throughput analysis**, it is recommended
  - To encourage and instruct police to be present prior to the work zone set up. This act would lead to better speed adjustments and consequently, less queuing.
  - To set up work zones in a way that they do not interrupt more than 2 lanes and if they do, they should not exceed 3 km in length.
  - To have closures during weekend nights; these closures are less susceptible to congestion.

- To schedule closures for nighttime rather than daytime; however, the lower speed limit for nighttime closures and better illumination system is also suggested.
- To use barrels over concrete barriers. This factor could be negotiable due to the safety advantages of concrete barriers in specific cases.

- According to the **safety analysis**, it is recommended
  - To set up work zones between 8 PM and 6 AM, which is in line with the findings from throughput analysis as well.
  - To use more traffic control devices. This factor's effect could be even more noticeable when the 'Warning Sign' itself became statistically significant.
  - To be more conservative in work zones located in highways with 4 or more lanes, e.g. 401. Any of the other findings which reduces the probability of fatal collisions should be implemented more rigorously in these high-risk work zones.
  - To have a lower posted speed limit in work zones where rollover occurrence is more probable, such as horizontal curves, slippery roads, and so on.
  - To collaborate more actively with police to be present prior to the work zone set up as this avoids tailgating, improper lane changes, and several other factors that are found to increase the probability of fatal collisions.
  - To use an analysis tool like the developed smart form in order to have a better understanding of possible injury severity of future events.

- According to the **methodological point of view**, it is recommended
  - To conduct a detailed statistical analysis on data sets prior to the model development stage. Being aware of the nature of dependent variables as well as issues such as unobserved heterogeneity can significantly improve the outcome of the developed models.
  - To apply the suggested methodology (as discussed in Section 5.4) for random parameter predictions which aims to satisfy two major considerations. First, it considers the most advantageous characteristic of random parameter factors; this characteristic allows the coefficients of the random factors to change across the observations. Secondly, it selects the most similar events to the current scenario and uses the average of the selected coefficients to reach a more robust prediction.

6.2 Contributions

The Contributions of this study could be listed as follows,

- Estimation of random parameter and fixed parameter negative binomial models based on in-house collected data. These models were developed for the first time specifically for the work zones located on high-volume highways of Ontario by the knowledge of the author.

- Development of safety models for having a better perception about the possible injury severity of the future collisions in work zones.

- Recommendation of a whole new methodology for predicting future events with random parameter models through the introduction of the 'Similarity Level' factor.

- Performing stratified 10-fold cross validation for the 1st time to assess the prediction performance of random parameter ordered Logit model.

- Creation of a straight-forward and user-friendly smart form which summarizes all findings from safety assessment. This form's outcome provides a probabilistic point of view toward the injury severity level of the future work zone collisions.

6.3 Future Work

Investigation of all factors that can improve congestion and safety of work zones based on the historical data and statistical and econometric models provides a broad insight into the possible future events. Although in this thesis, it was aimed to address some of the literature gaps, there is still a need to carry on the research to achieve further improvements on this topic. The following suggestions could inspire future research,

- **Canadian road safety database**: One of the most challenging tasks of this research was to gain access to local Canadian road safety database. Using Canadian data could localize models and consequently lead to a better understanding of future events in a Canadian context. Also, throughput database could be improved, having more consistent and detailed data collection forms.

- **Random threshold models**: These models have the advantage of estimating the thresholds for each injury severity data point individually. This also could highly improve the predictability of the estimated models. The introduced 'Similarity Level' methodology could also be adapted to estimate the most appropriate and efficient thresholds for future predictions as well.

- **Sensitivity analysis**: Due to time limitations, only similarity level of '75% and higher' was considered in this study; however, it is recommended to conduct research which evaluates other similarity levels such as 65% and 85% to see whether it has a significant effect on the outcomes.

- **Cost analysis**: Conducting an analysis which compares all of the direct costs, e.g. illumination system improvement, and increasing number of signs prior to the work zones, as well as indirect costs, e.g. increasing the public awareness, and law enforcement presence. The implementation of the findings, as related to the amount saved by reduction in the severity of collisions, could widen the decision-makers' insight toward the necessity of them.

References

[1] Transport Canada, "Transportation in Canada - Overview Report," tech. rep., Transport Canada, Ottawa, Ontario, CA, 2018.

[2] E. Chen and A. P. Tarko, "Analysis of crash frequency in work zones with focus on police enforcement," *Transportation Research Record*, vol. 2280, no. 1, pp. 127–134, 2012.

[3] F. D. Shepard and B. H. Cottrell, "Benefits and Safety Impact of Night Work-zone Activities.," *Transportation Research Record*, pp. 31–36, oct 1986.

[4] M. A. Mushtaq, *Measuring Work Zone Throughput and User Delays.* PhD thesis, University of Waterloo, 2011.

[5] C. Hicks, B. McCabe, and S. Tighe, "Evaluation of work zone capacity models," in *Building a Sustainable Future - Proceedings of the 2009 Construction Research Congress*, pp. 101–110, 2009. cited By 1.

[6] Transport Canada, "Transportation in Canada 2017 - Overview Report," tech. rep., Transport Canada, Ottawa, Ontario, CA, 2017.

[7] Transportation Association of Canada (TAC), *Geometric Design Guide for Canadian Roads*. Ottawa, Ontario, CA: Transportation Association of Canada (TAC), 1986.

[8] Federal Highway Administration (FHWA), *Assessing Roadway Traffic Count Duration and Frequency Impacts on Annual Average Daily Traffic (AADT) Estimation Assessing Accuracy Issues with Current Known Methods in AADT Estimation from Continuous Traffic Monitoring Data*. Washington D.C., United States: U.D. Department of Transportation, 2014.

[9] SRF Consulting Group Inc., "Transpotation Systems Management and Operation: Work Zone Management Service Layer," tech. rep., IOWA Department of Transportation (IOWADOT), 2018.

[10] Transportation Association of Canada (TAC), *Pavement Asset Design and Management Guide*. Transportation Association of Canada (TAC), 2013.

[11] D. Geiger, "Pavemenet Preservation Compendium II," tech. rep., U.S. Department of Transportation: Federal Highway Administration, 2006.

[12] D. J. Turner, "What's a Work-Zone?," *Federal Highway Administration*, vol. 62, no. 6, 1999.

[13] P. G. Gipps, "A model for the structure of lane-changing decisions," *Transportation Research Part B*, vol. 20, no. 5, pp. 403–414, 1986.

[14] A. Tarko, S. Kanipakapatnam, and J. Wasson, "Modeling and Optimization of the Indiana Lane Merge Control System on Approaches to Freeway Work Zones," Tech. Rep. FHWA/IN/JTRP-97/12, Joint Transportation Research Program, 1998.

[15] R. Herman, E. W. Montroll, R. B. Potts, and R. W. Rothery, "Traffic Dynamics: Analysis of Stability in Car Following," *Operations Research*, vol. 7, no. 1, pp. 86–106, 1959.

[16] A. D. May and E. M. H. Keller, "Non-Integer Car-Following Models," *Highway Research Record*, no. 199, pp. pp 19–32, 1967.

[17] H. Strobel, *Applied Systems Analysis Computer Controlled Transportation a Survey of Concepts, Methods, and International Experiences*. International Institute for Applied Systems Analysis, 1982.

[18] T. V. Mathew and K. Rao., "Fundamental relations of traffic flow," *The National Programme on Technology Enhanced Learning (NPTEL)*, pp. 1–8, 2006.

[19] Traffic Flow Theory and Characteristics Committee, "Transportation Research Circular - 75 Years of the Fundamental Diagram for Traffic Flow Theory: Greenshields Symposium," tech. rep., Transportation Research Board, Woods Hole, Massachusetts, 2011.

[20] M. Jabeena, "Comparative Study Of Traffic Flow Models And Data Retrieval Methods From Video Graphs," *Journal of Engineering Research and Applications*, vol. 3, pp. 1087–1093, 2013.

[21] N. M. Rouphail and G. Tiwari, "Traffic management in highway work zones and setting optimal maintenance levels and rehabilitation frequencies.," *Transportation Research Record*, no. 1035, pp. 50–58, 1985.

[22] W. Sarasua, W. Davis, D. Clarke, J. Kottapally, and P. Mulukutla, "Evaluation of Interstate Highway Capacity for Short-Term Work Zone Lane Closures," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1877, pp. 85–94, jan 2004.

[23] R. A. Krammes and G. O. Lopez, "Maintenance of the highway infrastructure," *Transportation Research Record*, no. 1442, pp. 49–56, 1994.

[24] T. Kim, D. J. Lovell, M. Hall, and J. Paracha, "A New Methodology to Estimate Capacity for Freeway Work Zones," *Transportation Research Record*, 2000.

[25] A. Al-Kaisy and F. Hall, "Guidelines for Estimating Freeway Capacity at Long-Term Reconstruction Zones," *Transportation Research Record*, 2002.

[26] Y. Chung, "Assessment of non-recurrent traffic congestion caused by freeway work zones and its statistical analysis with unobserved heterogeneity," *Transport Policy*, 2011.

[27] Z. L. Ma, H. L. Zhang, Y. Y. Zhang, and J. Wang, "Roadside accident frequency prediction model on expressway," *Chang'an Daxue Xuebao (Ziran Kexue Ban)/Journal of Chang'an University (Natural Science Edition)*, 2017.

[28] T. Hels and I. Orozova-Bekkevold, "The effect of roundabout design features on cyclist accident rate," *Accident Analysis and Prevention*, 2007.

[29] P. Nahidi, Seyedata; Golshani, Nima; Anastasopoulos, "An Empirical Exploratory Analysis of Factors Affecting Highway Segment Hazard-Level Likelihood Using Random Parameters Ordered Probit Regression," in *94th Transportation Research Board Conference*, (Washington D.C.), Conference Proceedings, 2015.

[30] B. Jones, L. Janssen, and F. Mannering, "Analysis of the frequency and duration of freeway accidents in Seattle," *Accident Analysis & Prevention*, vol. 23, pp. 239–255, aug 1991.

[31] S. P. Miaou and H. Lum, "Modeling vehicle accidents and highway geometric design relationships.," *Accident; analysis and prevention*, vol. 25, pp. 689–709, dec 1993.

[32] S. P. Miaou, "The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions," *Accident Analysis and Prevention*, vol. 26, no. 4, pp. 471–482, 1994.

[33] V. Shankar, F. Mannering, and W. Barfield, "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies.," *Accident; analysis and prevention*, vol. 27, pp. 371–89, jun 1995.

[34] M. Poch and F. Mannering, "Negative Binomial Analysis of Intersection-Accident Frequencies," *Journal of Transportation Engineering*, vol. 122, pp. 105–113, mar 1996.

[35] K. El-Basyouny and T. Sayed, "Accident prediction models with random corridor parameters," *Accident Analysis & Prevention*, vol. 41, no. 5, pp. 1118–1123, 2009.

[36] D.-G. Kim and S. Washington, "The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models," *Accident Analysis & Prevention*, vol. 38, pp. 1094–1100, nov 2006.

[37] N. V. Malyshkina and F. L. Mannering, "Zero-state Markov switching count-data models: An empirical assessment," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 122–130, 2010.

[38] S. Nahidi, S. Rahman, and S. Tighe, "Throughput Evaluation of Work zones on High Volume Highways of Ontario, Canada," in *12th International Transportation Specialty Conference, CSCE*, (Fredericton, NB), Canadian Society of Civil Engineering, Conference Proceedings, 2018.

[39] V. Shankar, J. Milton, and F. Mannering, "Modeling accident frequencies as zero-altered probability processes: An empirical inquiry," *Accident Analysis & Prevention*, vol. 29, pp. 829–837, nov 1997.

[40] J. Lee and F. Mannering, "Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis," *Accident Analysis & Prevention*, vol. 34, no. 2, pp. 149–161, 2002.

[41] Maycock, G., Hall, R.D., "Accidents at 4-Arm Roundabouts," tech. rep., TRRL Laboratory Report 1120, Transportation and Road Research Laboratory, Crowthorne, UK., 1984.

[42] C. R. Bhat, K. Born, R. Sidharthan, and P. C. Bhat, "A count data model with endogenous covariates: Formulation and application to roadway crash frequency at intersections," *Analytic Methods in Accident Research*, 2014.

[43] S. Chen, T. U. Saeed, M. Alinizzi, S. Lavrenz, and S. Labi, "Safety sensitivity to roadway characteristics: A comparison across highway classes," *Accident Analysis and Prevention*, 2019.

[44] B. McCullouch, K. Sinha, and P. Anastasopoulos, "Performance-Based Contracting for Roadway Maintenance Operations in Indiana," tech. rep., Joint Transportation Research Program, West Lafayette, IN United States, 2009.

[45] P. C. Anastasopoulos, V. N. Shankar, J. E. Haddock, and F. L. Mannering, "A multivariate tobit analysis of highway accident-injury-severity rates," *Accident Analysis & Prevention*, vol. 45, pp. 110–119, 2012.

[46] P. C. Anastasopoulos, F. L. Mannering, V. N. Shankar, and J. E. Haddock, "A study of factors affecting highway accident rates using the random-parameters tobit model," *Accident Analysis & Prevention*, vol. 45, pp. 628–633, 2012.

[47] G. F. Ulfarsson and F. L. Mannering, "Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents," *Accident Analysis & Prevention*, vol. 36, no. 2, pp. 135–147, 2004.

[48] A. Khattak, "Injury Severity in Multivehicle Rear-End Crashes," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1746, pp. 59–68, jan 2001.

[49] L.-Y. Chang and F. Mannering, "Predicting Vehicle Occupancies from Accident Data: An Accident Severity Approach," *Transportation Research Record*, vol. 1635, pp. 93–104, 1998.

[50] F. L. Mannering and C. R. Bhat, "Analytic methods in accident research: Methodological frontier and future directions," *Analytic Methods in Accident Research*, vol. 1, pp. 1–22, 2014.

[51] A. Shibata and K. Fukuda, "Risk factors of fatality in motor vehicle traffic accidents," *Accident Analysis and Prevention*, vol. 26, no. 3, pp. 391–397, 1994.

[52] M. Sarwar, S. Nahidi, and P. Anastasopoulos, "Analysis of Accident Injury-severities of Helmeted and Non-helmeted Motorcyclists," in *94th Transportation Research Board Conference*, (Orlando, FL), Conference Proceedings, 2015.

[53] V. Shankar and F. Mannering, “An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity,” *Journal of Safety Research*, vol. 27, no. 3, pp. 183 – 194, 1996.

[54] F. Ye and D. Lord, “Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models,” *Analytic Methods in Accident Research*, 2014.

[55] S. Nahidi, G. Fountas, S. P. Sarvani, M. T. Sarwar, and P. C. Anastasopoulos, “Project Discrepancies in Roadway Construction and Preservation: A Statistical Analysis of PublicPrivate Partnership Contract Types in the US,” *Frontiers in Built Environment*, vol. 3, mar 2017.

[56] S. Yasmin and N. Eluru, “Evaluating alternate discrete outcome frameworks for modeling crash injury severity,” *Accident Analysis and Prevention*, vol. 59, pp. 506–521, 2013.

[57] S. Jung, X. Qin, and D. A. Noyce, “Rainfall effect on single-vehicle crash severities using polychotomous response models,” *Accident Analysis and Prevention*, vol. 42, no. 1, pp. 213–224, 2010.

[58] C. Xu, A. P. Tarko, W. Wang, and P. Liu, “Predicting crash likelihood and severity on freeways with real-time loop detector data,” *Accident Analysis and Prevention*, vol. 57, pp. 30–39, 2013.

[59] J. Aguero-Valverde and P. Jovanis, “Analysis of Road Crash Frequency with Spatial Models,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2061, pp. 55–63, dec 2008.

[60] J. Ma and K. Kockelman, “Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1950, pp. 24–34, jan 2006.

[61] E. Park and D. Lord, “Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2019, pp. 1–6, dec 2007.

[62] P. C. Anastasopoulos and F. L. Mannering, “A note on modeling vehicle accident frequencies with random-parameters count models,” *Accident Analysis and Prevention*, vol. 41, no. 1, pp. 153–159, 2009.

[63] P. C. Anastasopoulos, G. Fountas, M. T. Sarwar, M. G. Karlaftis, and A. W. Sadek, "Transport habits of travelers using new energy type modes: A random parameters hazard-based approach of travel distance," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 516–528, 2017.

[64] N. Eluru and C. R. Bhat, "A joint econometric analysis of seat belt use and crash-related injury severity," *Accident Analysis and Prevention*, vol. 39, no. 5, pp. 1037–1049, 2007.

[65] H. Yang, K. Ozbay, O. Ozturk, and M. Yildirimoglu, "Modeling work zone crash frequency by quantifying measurement errors in work zone length," *Accident Analysis and Prevention*, 2013.

[66] Z. Nemeth and D. Migletz, "Accident Characteristics Before, During, and After Safety Upgrading Projects on Ohio's Rural Interstate System," *Transportation Research Record*, no. 672, 1978.

[67] B. T. Hargroves, "Vehicle Crashes in Highway Work Zones," *Journal of Transportation Engineering*, vol. 107, no. TE5, pp. 525–539, 1981.

[68] N. Rouphail, Z. Yang, and J. Fazio, "Comparative Study of Short- and Long-Term Urban Freeway Work Zones," *Transportation Research Record 1163*, no. 1163, pp. 4–13, 1988.

[69] J. W. Hall and M. Lorenz, "Characteristics of Construction-Zone Accidents," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1163, pp. 4–14, 1989.

[70] J. G. Pigman and K. R. Agent, "Highway Accidents in Construction and Maintenance Work Zones," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1270, pp. 12–21, 1990.

[71] M. Zhao and N. J. Garber, "Crash Characteristics at Work Zones," tech. rep., Mid-Atlantic Universities Transportation Center, 2001.

[72] T.-J. Ha and Z. A. Nemeth, *Maintenance management and safety.* National Academy Press, 1995.

[73] J. Bryden, L. Andrew, and J. Fortuniewicz, "Work Zone Traffic Accidents Involving Traffic Control Devices, Safety Features, and Construction Operations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1650, pp. 71–81, jan 1998.

[74] J. Daniel, K. Dixon, and D. Jared, "Analysis of Fatal Crashes in Georgia Work Zones," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1715, pp. 18–23, jan 2000.

[75] J. Chambless, A. M. Ghadiali, J. K. Lindly, and J. McFadden, "Multistate work-zone crash characteristics," 2002.

[76] S. D. Schrock, G. L. Ullman, a. S. Cothron, E. Kraus, and A. P. Voigt, "An Analysis of Fatal Work Zone Crashes in Texas," Tech. Rep. 2, Texas Transportation Institute, Austin, Texas, 2004.

[77] O. Salem, A. Genaidy, Heng Wei, and N. Deshpande, "Spatial distribution and characteristics of accident crashes at work zones of interstate freeways in Ohio," in *2006 IEEE Intelligent Transportation Systems Conference*, pp. 1642–1647, IEEE, 2006.

[78] D. Arditi, D.-E. Lee, and G. Polat, "Fatal accidents in nighttime vs. daytime highway construction work zones," *Journal of Safety Research*, vol. 38, no. 4, pp. 399–405, 2007.

[79] G. L. Ullman, M. D. Finley, J. E. Bryden, R. Srinivasan, and F. M. Council, *Traffic Safety Evaluation of Nighttime and Daytime Work Zones*. National Academies Press, 2008.

[80] T. G. Jin, M. Saito, and D. L. Eggett, "Statistical comparisons of the crash characteristics on highways between construction time and non-construction time," *Accident Analysis & Prevention*, vol. 40, no. 6, pp. 2015–2023, 2008.

[81] Y. Li and Y. Bai, "Highway Work Zone Risk Factors and Their Impact on Crash Severity," *Journal of Transportation Engineering*, vol. 135, no. 10, pp. 694–701, 2009.

[82] S. Dissanayake and S. R. Akepati, "Characteristics of Work Zone Crashes in the SWZDI Region: Differences and Similarities," in *Mid-Continent Transportation Research Symposium*, Transportation Research Board, 2009.

[83] J. Xing, H. Takahashi, and K. Iida, "Analysis of Bottleneck Capacity and Traffic Safety in Japanese Expressway Work Zones," in *Transportation Research Record*, 2010.

[84] G. Ullman, M. Finley, and L. Theiss, "Categorization of Work Zone Intrusion Crashes," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2258, pp. 57–63, dec 2011.

[85] E. Swansen and M. A. Knodler Jr., "Analysis of Linked Crash and Citation Data for Use in Determining Causation for Work Zone Crashes," *Transportation Research Board 91st Annual Meeting*, 2012.

[86] Y. Li, F. Cheng, and Y. Bai, "Characteristics of Truck-Related Crashes in Highway Work Zones," in *Sustainable Transportation Systems*, (Reston, VA), pp. 364–371, American Society of Civil Engineers, jun 2012.

[87] K. El-Rayes, L. Liu, and T. Elghamrawy, *Minimizing Traffic-Related Work Zone Crashes in Illinois Studying and Minimizing Traffic-Related Work Zone Crashes in Illinois*. University of Illinois at Urbana-Champaign, 2013.

[88] N. Garber and M. Zhao, "Distribution and Characteristics of Crashes at Different Work Zone Locations in Virginia," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1794, pp. 19–25, 2002.

[89] R. F. Benekohal and E. Shim, "Multivariate Analysis of Truck Drivers' Assessment Of Work Zone Safety.," *Journal of Transportation Engineering*, vol. 125, no. 5, p. 398, 1999.

[90] K. Vodden, D. Smith, F. Eaton, and D. Mayhew, "Analysis and Estimation of the Social Cost of Motor Vehicle Collisions in Ontario Final Report Ministry of Transportation," tech. rep., Transport Canada, 2007.

[91] A. J. Khattak, A. J. Khattak, and F. M. Council, "Effects of work zone presence on injury and non-injury crashes," *Accident Analysis and Prevention*, vol. 34, no. 1, pp. 19–29, 2002.

[92] R. Turochy, N. Jehn, and J. LAMondia, "Analysis of Work-Zone Crash Reports to Determine Factors Associated with Crash Severity," in *Transportation Research Record*, (Washington D.C., United States), p. 6, Transportation Research Board 97th Annual Meeting, 2018.

[93] Q. Meng and J. Weng, "Evaluation of rear-end crash risk at work zone using work zone traffic data," *Accident Analysis and Prevention*, vol. 43, no. 4, pp. 1291–1300, 2011.

[94] H. Yang, K. Ozbay, O. Ozturk, and K. Xie, "Work Zone Safety Analysis and Modeling: A State-of-the-Art Review," *Traffic Injury Prevention*, vol. 16, no. 4, pp. 387–396, 2015.

[95] P. W. Lin, K. P. Kang, and G. L. Chang, "Exploring the effectiveness of variable speed limit controls on highway work-zone operations," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 8, no. 3, pp. 155–168, 2004.

[96] S. Mohan and P. Gautam, "Cost of Highway Work Zone Injuries," *Practice Periodical on Structural Design and Construction*, vol. 7, no. 2, pp. 68–73, 2002.

[97] J. E. Bryden and L. B. Andrew, "Serious and Fatal Injuries to Workers on Highway Construction Projects," *Transportation Research Record*, vol. 1657, pp. 42–47, 1999.

[98] J. M. Wong, M. C. Arico, and B. Ravani, "Factors influencing injury severity to highway workers in work zone intrusion accidents," *Traffic Injury Prevention*, vol. 12, no. 1, pp. 31–38, 2011.

[99] Y. Bai and Y. Li, "Determining the drivers acceptance of eftcd in highway work zones," *Accident Analysis and Prevention*, vol. 43, no. 3, pp. 762 – 768, 2011.

[100] K. K. Mak, R. P. Bligh, and J. Lewis R. Rhodes, "Crash testing and evaluation of work zone traffic control devices," *Transportation Research Record*, vol. 1650, no. 1, pp. 45–54, 1998.

[101] S. P. Washington, M. G. Karlaftis, and F. Mannering, *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC, 2010.

[102] D. C. Montgomery, *Design and analysis of experiments*. John wiley & sons, 2017.

[103] T. A. Domencich, j. a. McFadden, Daniel, and C. R. Associates, *Urban travel demand : a behavioral analysis : a Charles River Associates research study*. Amsterdam : North-Holland Pub. Co. ; New York : American Elsevier, 1975. Includes index.

[104] W. H. Greene, *Econometric analysis*. Pearson Education India, 2003.

[105] J. H. Halton, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numerische Mathematik*, vol. 2, pp. 84–90, dec 1960.

[106] C. R. Bhat, "Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences," *Transportation Research Part B: Methodological*, vol. 37, no. 9, pp. 837–855, 2003.

[107] K. Train, *Discrete choice methods with simulation*. Cambridge University Press, 2003.

[108] J. C. Milton, V. N. Shankar, and F. L. Mannering, "Highway accident severities and the mixed logit model: An exploratory empirical analysis," *Accident Analysis & Prevention*, vol. 40, pp. 260–266, jan 2008.

[109] Transport Canada, " Canadian Motor Vehicle Traffic Collision Statistics:2017," 2018.

[110] W. G. Najm, R. Ranganathan, G. Srinivasan, J. S. Toma, E. Swanson, A. Burgett, and D. Smith, "Description of Light-Vehicle Pre-Crash Scenarios for Safety Applications Based On Vehicle-to-Vehicle Communications," tech. rep., John A. Volpe National Transportation Systems Center, Cambridge, MA, 2013.

[111] Royal Canadian Mounted Police, "Gazette / Royal Canadian Mounted Police.," 2018.

[112] D. Kong, X. Guo, B. Yang, and D. Wu, "Analyzing the Impact of Trucks on Traffic Flow Based on an Improved Cellular Automaton Model," *Discrete Dynamics in Nature and Society*, p. 14, 2016.

[113] American Traffic Safety Services Association, "Nighttime Lighting Guidelines for Work Zones A guide for developing a lighting plan for nighttime work zones," tech. rep., Federal Highway Administration, Fredericksburg, VA, 2013.

[114] R. L. Vecellio, P. E. John, and R. Mccarthy, "Lighting Specifications for Nighttime Construction Work Zones on Active Highways, Report 930-470," tech. rep., Auburn University, Auburn, Alabama, 2006.

[115] J. C. Paul Lutkevich, Don McLean, "FHWA Lighting Handbook," tech. rep., WSP, Boston, MA, 2012.

[116] Ministry of Transportation of Ontario (MTO), "Book 7 Ontario Traffic Manual Temporary Conditions," tech. rep., MTO, St. Catharines, Ontario, 2014.

[117] M. Corfitsen, "Tiredness and visual reaction time among young male nighttime drivers: A roadside survey," *Accident Analysis & Prevention*, vol. 26, no. 5, pp. 617 – 624, 1994.

[118] Organisation for Economic Co-Operation and Development, "Young Drivers: The Road to Safety," tech. rep., the Public Affairs Division, Public Affairs and Communications Directorate, Paris, France, 2006.

[119] S. L. Suen, "Advanced Accessibility Systems for Elderly and Disabled Travellers in Canada," *IFAC Proceedings Volumes*, 1994.

[120] T. Tervo, E. Räty, P. Sulander, J. M. Holopainen, T. Jaakkola, and K. Parkkari, "Sudden Death at the Wheel Due to a Disease Attack," *Traffic Injury Prevention*, 2013.

[121] E. Keskinen, "Education for older drivers in the future," *IATSS Research*, vol. 38, pp. 14–21, jul 2014.

[122] S. Perreault, "Impaired driving in Canada, 2015," tech. rep., Canadian Centre for Justice Statistics, 2016.

[123] K. Haslam, Bill; Schroer, John; Poole, "2016 Highway Safety Plan," tech. rep., Tennessee's Governor's Highway Safety Office, 2015.

[124] American Association of State Highway and Transportation Officials, *A Policy on Geometric Design of Highways and Streets, 2011*. Green book, American Association of State Highway and Transportation Officials, 2011.

[125] Road Safety Canada Consulting, *Road safety in Canada*. Ottawa, Ontario, Canada: Government of Canada, 2011.

[126] Federal Highway Administration (FHWA), "Alternative Fuels Data Center: Maps and Data - Vehicle Weight Classes & Categories."

[127] F. Simon and C. Corbett, "Road traffic offending, stress, age, and accident history among male and female drivers," *Ergonomics*, 1996.

[128] P. Marsh, "Sex differences in driving and insurance risk: An analysis of the social and psychological differences between men and women that are relevant to their driving behaviour.," tech. rep., The Social Issues Research Centre, 2004.

[129] M. R. Veall and K. F. Zimmermann, "PseudoR2's in the ordinal probit model," *Journal of Mathematical Sociology*, vol. 16, no. 4, pp. 333–342, 1992.

[130] R. E. Kass and A. E. Raftery, "Bayes Factors," *Journal of the American Statistical Association*, vol. 90, pp. 773–795, 1995.

[131] P. Refaeilzadeh, L. Tang, and H. Liu, *Cross-Validation*, pp. 532–538. Boston, MA: Springer US, 2009.

[132] P. de Leur and P. De, “Collision Cost Study Update FINAL Report,” tech. rep., Capital Region Intersection Safety Partnership, 2018.

APPENDICES

Appendix A

Initial List of Projects/Contracts for New Data Collection

Site No.	Project No.	Hwy	Location or Intersection	Length (km)	Project Type
1	2090-13-01	401	Wilson Ave Overpass WBL Coll.	37	Resurfacing
2	2264-14-01	401	Bayview Ave Overpass WB Coll.	37	Resurfacing
3	2316-15-01	401	Bathurst St Overpass WB Coll.	37	Resurfacing
4	2317-15-01	401	Hogg's Hollow Br. WB Coll.	37	Resurfacing
5	2082-13-01	427	S. B. Basket Weave S of Bloor St. (Br#5)	37	Brdg. Rehab
6	2430-13-01	427	Cloverdale Mall Access Ramp E-N	37	Brdg. Rehab
7	2038-13-01	401	Humber River Bridge #4 (W-N/S Ramp)	37	Brdg. Rehab

Table A.1 continued from previous page

Site No.	Project No.	Hwy	Location or Intersection	Length (km)	Project Type
8	2101-13-01	QEW	QEW/Hwy 420 Interchange , Bridge #2, NBL	34	Resurfacing
9	2102-13-01	QEW	QEW/Hwy 420 Interchange , Bridge #2, SBL	34	Resurfacing
10	2103-13-01	420	QEW/Hwy 420 Interchange	34	Resurfacing
11	2104-13-01	420	QEW/Hwy 420 Overpass Bridge #7 EBL	34	Resurfacing
12	2256-13-01	420	Bridge No. 4 Turning Rd. E-S	34	Resurfacing
13	2163-10-03	QEW	W-N Ramp Hwy 403 Under QEW	10	Unknown
14	2305-13-01	403	Dundas St. Over 403 EBL	10	Unknown
15	2206-10-02	QEW	W-N Ramp Hwy 403 Under Service Rd.	10	Unknown
16	2306-13-01	403	Dundas St. Over 403 WBL	10	Unknown
17	2318-13-01	QEW	Joshua's Creek Culvert	10	Unknown
18	2329-15-01	QEW	QEW EBL Over Ford Drive	10	Unknown
19	2182-13-01	406	St. Davids Rd Interchange Bridge #7	18	Unknown
20	2193-13-01	QEW	County Rd #20 Overpass EBL	18	Rehab
21	2194-13-01	QEW	County Rd #20 Overpass WBL	18	Rehab

Table A.1 continued from previous page

Site No.	Project No.	Hwy	Location or Intersection	Length (km)	Project Type
22	2199-13-01	406	406 Third Ave U/P	18	Rehab
23	2200-13-01	406	406 S-W Ramp Over South Service Rd	18	Rehab
24	2201-13-01	406	406 E-S Ramp Over South Service Rd	18	Rehab
25	2202-13-01	406	406 W-S Ramp Over South Service Rd	18	Rehab
26	2203-13-01	406	406/QEW E-S Ramp Bridge #14	18	Rehab
27	2204-13-01	406	406/QEW S-W Ramp Underpass Bridge#13	18	Rehab
28	2434-13-01	406	406 S-E Ramp SS Rd Bridge #12	18	Rehab
29	2435-13-01	406	406 S-W Third St Overpass Bridge #15	18	Rehab
30	2436-13-01	406	406 S-S Ramp Third St Overpass Br#18	18	Rehab
31	2323-10-01	403	Hurontario St Underpass (Br #41) NBL	24	Brdg. Rehab
32	2318-10-01	403	Erin Mills Parkway Underpass EBL	24	Brdg. Rehab
33	2318-10-02	403	Erin Mills Parkway Underpass WBL	24	Brdg. Rehab
34	2323-10-02	403	Hurontario St Underpass (Br #41) SBL	24	Brdg. Rehab

Table A.1 continued from previous page

Site No.	Project No.	Hwy	Location or Intersection	Length (km)	Project Type
35	2186-10-01	400	4 Th Line (Churchill) Underpass	30	Brdg. Replc.
36	2285-13-01	400	Churchill SdrRd Hwy 400 U/P	30	Brdg. Replc.
37	2350-09-01	401	Tremaine Rd. (Halton Rd) 22 Up	10	Deck Rehab
38	2185-13-01	400	Mount St Louis Road Underpass	30	Brdg. Rehab
39	2186-13-01	400	Simcoe City Road #19 Interchange	30	Brdg. Rehab
40	2187-13-01	400	Big Chute Road Underpass	30	Brdg. Rehab
41	2288-13-01	400	Medonte Concession 6 U/P	30	Brdg. Rehab

Appendix B

Work Zone Survey

1. Participant's Information

This survey is designed to get some information about the accidents occurred on work zones located at high volume highways in last 5 years. Each survey is designed for a single accident and participants are encouraged to fill these surveys for several accidents to enrich the database.

By returning a completed questionnaire this implies consent for the researchers to use the information provided for the study. The data from this survey will remain confidential and the details will not be presented to the public.

1. Name of the Company/Organization

2. Role in the Company/Organization

2. Accident Related Questions

1. Location of the Accident (e.g. Highway 401 close to exit 299)

2. Type of Accident?

- Fatalities
- Severe Injuries Minor Injuries
- Property Damage Only (PDO)
-

3. How many vehicles were involved in the accident?

- One
- Two
- More than two
- Other (please specify)

4. How many people were involved in the accident?

- One
- Two
- More than two
- Other (please specify)

5. Number of Fatalities, Severe Injuries, Minor Injuries, and PDO?

# of Fatalities

# of Severe Injuries

# of Minor Injuries

# of PDO

6. On which lane accident occurred?

- Outer right
- Inner Left
- Inner Right
- Outer Left
- Middle
- Other (please specify)

7. Did any of the on-site staff experience any type of harm? if yes what type?

- Yes
- No

If yes, what type? (e.g. Fatalities, Severe Injuries, etc.)

8. Type of vehicles involved in the accident?

- Passenger Car
- Small Truck
- Large Truck
- Bus

9. Cause of accident?

- Drink and Drive
- Lack of attention
- Technical failure of the vehicle
- Electronic devices
- Drowsiness
- Other (please specify)

3. Workzone and Highway Geometry Questions

1. which part of the workzone accident occurred?

- Advance Warning Area
- Approach Area
- Longitudinal Buffer Area
- Other (please specify)
- Work Area
- Termination area

2. Workzone was located on which side of the highway?

- Right
- Left

3. Number of closed lanes during the working period?

- One lane
- Two lanes
- More than two lanes
- Other (Please specify e.g. shoulder closure)

4. Was there any median in the location of the accident?

- Yes (Please specify the type of median)
- No

Median Type

5. What is the width of Median, Right Shoulder, and Left Shoulder at the location of the accident?

Median width

Right Shoulder Width

Left Shoulder width

6. Length of the Transition Area?

7. Length of workzone?

8. Length of the Termination Area?

9. Duration of workzone?

10. Type of barriers used for setting up the workzone? (e.g. Barrels, etc.)

- Barrels
- Concrete
- Other (please specify)

11. The total cost of the workzone? (Approximately)

4. Pavement and Related Activities

1. Type of pavement?

- ◯ Flexible Pavement (Asphalt)
- ◯ Rigid Pavement (Concrete)
- ◯ Composite
- ◯ Gravel
- ◯ Other (please specify)

2. Pavement Condition before workzone?

- ◯ Excellent
- ◯ Good
- ◯ Fair
- ◯ Poor
- ◯ Extremely poor

3. Pavement Condition after workzone?

- ◯ Excellent
- ◯ Good
- ◯ Fair
- ◯ Poor
- ◯ Extremely poor

4. Construction year of the Pavement?

5. Last major rehabilitation year?

Year

Pavement condition after major rehabilitation?

6. Last maintenance year?

Year

Pavement condition after maintenance?

7. International Roughness Index (IRI) before workzone?

[ ]

8. International Roughness Index (IRI) after workzone?

[ ]

9. IRI after major rehabilitation? (If Applicable)

[ ]

10. IRI after maintenance? (If Applicable)

[ ]

11. Activity type in the workzone?

[ ]

12. Type of work area?

- ○ Lane Closure
- ○ Lane shift/crossover
- ○ Moving work
- ○ Other (please specify)

[ ]

13. Did the work area have an influence on the occured crash?

- ○ Yes
- ○ No

5. General Information

1. Time of the Accident?

Date / Time

hh | mm | -

2. Date of the Accident?

Date / Time

MM/DD/YYYY

3. Presence of Police?

- Yes
- No

4. Posted speed limit?

5. Workzone speed limit?

6. Workzone signage type?

7. Number of signs used before the workzone?

- One
- Two
- Three
- More than Three
- Other (please specify)

8. Number of signs used after the workzone?

- ○ One
- ○ Two
- ○ Three
- ○ Other (please specify)

9. Percentage of trucks of the segment?

10. Average Annual Daily Traffic of the segment?

Appendix C

Occupant and Driver Data Collection Form

U.S. Department of Transportation
National Highway Traffic Safety Administration

EXAMPLE

STATE NUMBER (GSA CODES) [1]	CONSECUTIVE NUMBER [94]	VEHICLE NUMBER [1]	PERSON NUMBER [1]

AGE [18]

Actual Value Except
000-Less than One Year
001-120-Actual Age*
998-Not Reported
999-Unknown

SEX [1]

1-Male
2-Female
8-Not Reported
9-Unknown

PERSON TYPE [01]

01-Driver of Motor Vehicle in-Transport
02-Passenger of Motor Vehicle In-Transport
03-Occupant of a Motor Vehicle Not In-Transport
09-Unknown Occupant Type in a Motor Vehicle In-Transport

INJURY SEVERITY [2]

0-No Apparent Injury (O)
1-Possible Injury (C)
2-Suspected Minor Injury (B)
3-Suspected Serious Injury (A)
4-Fatal Injury (K)
5-Injured, Severity Unknown
6-Died Prior to Crash
9-Unknown/Not Reported

SEATING POSITION [11]

	Left	Middle	Right	Other	Unknown
Front Row Seats	11	12	13	18	19
2nd Row Seats	21	22	23	28	29
3rd Row Seats	31	32	33	38	39
4th Row Seats	41	42	43	48	49

50 - Sleeper Section of Cab (truck)
51 - Other Passenger in Enclosed Passenger or Cargo Area
52 - Other Passenger in Unenclosed Passenger or Cargo Area
53 - Other Passenger in Passenger or Cargo Area, Unknown Whether or Not Enclosed
54 - Trailing Unit
55 - Riding on Exterior of Vehicle
98 - Not Reported
99 - Unknown

RESTRAINT SYSTEM/HELMET USE [03]

01-Shoulder Belt Only Used
12-Booster Seat

POLICE REPORTED ALCOHOL INVOLVEMENT [9]

0-No (Alcohol Not Involved)
1-Yes (Alcohol Involved)
8-Not Reported
9-Unknown (Police Reported)

METHOD OF ALCOHOL DETERMINATION (By Police) [9]

1-Evidential Test (Breath, Blood, Urine)
2-Preliminary Breath Test (PBT)
3-Behavioral
4-Passive Alcohol Sensor (PAS)
5-Observed
8-Other (e.g., Saliva test)
9-Not Reported

ALCOHOL STATUS [00] **ALCOHOL TEST TYPE** [00] **ALCOHOL TEST RESULTS** [996]

STATUS:

0-Test Not Given
8-Not Reported
2-Test Given
9-Unknown if Tested

TYPE:

00-Test Not Given
01-Blood
02-Breath Test (AC)
03-Urine
08-Other Test Type
10-Preliminary Breath Test (PBT)
95-Not Reported
98-Unknown Test Type
99-Unknown if Tested

RESULTS:

Actual Value (Decimal Implied Before First Digit (0.xx))
Except:
000-939-Actual Value
940-.94 or Greater
996-Test Not Given
997-AC Test Performed, Results Unknown
998-Positive Reading with No Actual Value
995-Not Reported
999-Unknown if Tested

POLICE REPORTED DRUG INVOLVEMENT [1]

02-Lap Belt Only Used
03-Shoulder and Lap Belt Used
04-Child Restraint - Type Unknown
05-DOT-Compliant Motorcycle Helmet
08-Restraint Used - Type Unknown
10-Child Restraint - Forward Facing
11-Child Restraint - Rear Facing
16-Helmet, Other than DOT-Compliant Motorcycle Helmet
17-No Helmet
19-Helmet, Unknown if DOT Compliant
20 (None Used / Not Applicable)
29-Unknown if Helmet Worn
97-Other
98-Not Reported
99-Unknown

ANY INDICATION OF MIS-USE OF RESTRAINT SYSTEM OF HELMET USE [00]

0-No
1-Yes

AIR BAG DEPLOYED [20]

01-Deployed-Front
02-Deployed-Side(door,seatback)
03-Deployed-Curtain(roof)
07-Deployed-Other(knee,airbelt, etc)
08-Deployed-Combination
09-Deployed-Unknown Location
20-Not deployed
98-Not Reported
99-Deployment Unknown

EJECTION [0]

0-Not Ejected
1-Totally Ejected
2-Partially Ejected
3-Ejected-Unknown Degree
7-Not Reported
8-Not Applicable
9-Unknown if Ejected

EJECTION PATH [0]

0-Ejection Path Not Applicable
1-Through Side Door Opening
2-Through Side Window
3-Through Windshield
4-Through Back Window
5-Through Back Door/Tailgate Opening
6-Through Roof Opening (sunroof; convertible top down)
7-Through Roof (convertible top up)
8-Other Path (e.g., Back of pickup truck)
9-Ejection Path Unknown

EXTRICATION [0]

0-Not Extricated / Not Applicable
1-Extricated
9-Unknown

0-No (Drugs Not Involved)
1-Yes (Drugs Involved)
8-Not Reported
9-Unknown

METHOD OF DRUG DETERMINATION (By Police) [7]

1-Evidential Test(Blood, Urine)
2-Drug Recognition Expert (or Evaluator)(DRE) determination
3-Behavioral
7-Other
8-Not Reported

DRUG TEST

Status	Type1	Result1	Type2	Result2	Type3	Result3
0	0	000	0	000	0	000

Status:

0-Test Not Given
2-Test Given
8-Not Reported
9-Unknown if Tested

TYPE:

0-Test Not Given
1-Blood Test
2-Urine Test
3-Both: Blood & UrineTests
7-Unknown Test Type
6-Not Reported
8-Other Test Type
9-Unknown if Tested

RESULTS:

000-Test Not Given
001-Tested, No Drugs Found/Negative
997-Tested For Drugs, Results Unknown
998-Tested For Drugs, Drugs Found, Type Unknown/Positive
999-Unknown If Tested

See Instruction Manual for specific drug listing

TRANSPORTED TO FIRST MEDICAL FACILITY BY [5]

0-Not Transported
1-EMS Air
2-Law Enforcement
3-EMS Unknown Mode
4-Transported Unknown Source
5-EMS Ground
6-Other
8-Not Reported
9-Unknown

DIED AT SCENE/EN ROUTE [0]

0-Not Applicable
7-Died at Scene
8-Died En Route
9-Unknown

DEATH DATE (MMDDYYYY) [88888888]

MONTH/DAY

88-Not Applicable(Non-fatal)
99-Unknown

YEAR

8888-Not Applicable(Non-fatal)
9999-Unknown

DEATH TIME (HHMM) [8888]

Military Time Except: 8888-Not Applicable(Non-fatal) 9999-Unknown

RELATED FACTORS [00|00|00]

LOOKUP RELATED FACTORS-PERSON LEVEL CODES

Appendix D

Vehicle Data Collection Form

VEHICLE LEVEL

U.S. Department of Transportation
National Highway Traffic Safety Administration

STATE NUMBER (GSA CODES)	1	CONSECUTIVE NUMBER	94	VEHICLE NUMBER	1	NUMBER OF OCCUPANTS 01-98-Actual Value if Total Known 99-Unknown	2

DRIVER INFORMATION   PRE-CRASH INFORMATION

UNIT TYPE [1]

1-Motor Vehicle In-Transport(Inside or Outside the Trafficway)
2-Motor Vehicle Not In-Transport Within Trafficway
3-Motor Vehicle Not In-Transport Outside Trafficway
4-Working Motor Vehicle (Highway Construction, Maintenance, Utility Only)

HIT AND RUN [0]

0-No   1-Yes   9-Unknown

REGISTRATION STATE [48]

GSA CODES Except:
00-Not Applicable
91-Not Reported
92-No Registration
93-Multiple State Registration
94-U.S. Government Tags(includes military)
95-Cananda
96-Mexico
97-Other Foreign Country
98-Other Registration
99-Unknown

REGISTERED VEHICLE OWNER [2]

0-Not Applicable, Vehicle Not Registered
1-Driver (this crash) Registered Owner
2-Driver (this crash) Not Registered Owner (Other Private Owner Listed)
3-Vehicle Registered as Business/Company/Government Vehicle
4-Vehicle Registered as Rental Vehicle
5-Vehicle Stolen (Reported by Police)
6-Driverless/Motor Vehicle Parked/Stopped Off Roadway
9-Unknown

VEHICLE MAKE [49]

LOOKUP VEHICLE MAKE CODES

VEHICLE MODEL [401]

LOOKUP VEHICLE MODEL CODES

BODY TYPE [14]

LOOKUP BODY TYPE CODES

MODEL YEAR [2002]

Actual Value Except: 9998-Not Reported, 9999-Unknown

VEHICLE IDENTIFICATION NUMBER [JT3GN86R5202]

00000000000000000-No VIN Required
99999999999999999-Unknown
88888888888888888-Not Reported

VEHICLE TRAILING [0]

0-No Trailing Units
1-One Trailing Unit
2-Two Trailing Units
3-Three or More Trailing Units
4-Yes, Number of Trailers Unknown
5-Vehicle Towing Another Motor Vehicle - Fixed Linkage
6-Vehicle Towing Another Motor Vehicle - Non-Fixed Linkage
9-Unknown

TRAILER VEHICLE 777777777777, 777777777777, 777777777777

HAZARDOUS MATERIAL INVOLVEMENT/PLACARD

HM1(Involvement)	HM2(Placard)	HM3(Identification Number)	HM4(Class Number)	HM5(Released)
1	0	0000	0	0
Blank 1-No 2-Yes	Blank 0-Not Applicable 1-No 2-Yes 8-Not Reported	Blanks 0000-Not Applicable Actual 4-digit Number 8888-Not Reported	Blank 00-Not Applicable Actual 1-digit Number(with leading zero) 88-Not Reported	Blank 0-Not Applicable 1-No 2-Yes 8-Not Reported

BUS USE [0]

00-Not a Bus
01-School
04-Intercity
05-Charter/Tour
06-Transit/Commuter
07-Shuttle
08-Modified for Personal/Private Use
98-Not Reported
99-Unknown

SPECIAL USE [0]

00-No Special Use
01-Taxi
02-Vehicle Used for School Transport
03-Vehicle Used as Other Bus
04-Military
05-Police
06-Ambulance
07-Fire Truck
08-Non-Transport Emergency Services Vehicle
13-Incident Response
98-Not Reported
99-Unknown

EMERGENCY MOTOR VEHICLE USE [0]

0-Not Applicable
2-Non-Emergency, Non-Transport
3-Non-Emergency Transport
4-Emergency Operation, Emergency Warning Equipment Not in Use
5-Emergency Operation, Emergency Warning Equipment in Use
6-Emergency Operation, Emergency Warning Equipment in Use Unknown
8–Not Reported
9-Unknown

TRAVEL SPEED [999]

Actual Miles Per Hour Except:
000-Stopped Motor Vehicle In-Transport
001-151 Reported Speed up to 151 MPH
997-Greater than 151 MPH
998- Not Reported
999-Unknown

UNDERRIDE/OVERRIDE [0]

0-No Underride or Override Noted
UNDERRIDING A MOTOR VEHICLE IN-TRANSPORT
1-Underride (Compartment Intrusion)
2-Underride (No Compartment Intrusion
UNDERRIDING A MOTOR VEHICLE NOT IN-TRANSPORT
4-Underride (Compartment Intrusion)
5-Underride (No Compartment Intrusion
6-Underride Compartment Intrusion Unknown)

IDENTIFICATION NUMBER

00000000000000000 - No VIN Required
77777777777777777 - No Trailing Units
88888888888888888 - Not Reported
99999999999999999 - Unknown

JACKKNIFE [0]

0-Not an Articulated Vehicle
1-No
2-Yes, First Event
3-Yes, Subsequent Event

MOTOR CARRIER IDENTIFICATION NUMBER [00000000000]

GROSS VEHICLE WEIGHT RATING
GROSS COMBINATION WEIGHT RATING [0]

0- Not Applicable
1-10,000 lbs. or less
2- 10,001 - 26,000 lbs.
3- 26,001 lbs. or more
8-Not Reported
9-Unknown

VEHICLE CONFIGURATION [00]

00-Not Applicable
01-Single Unit Truck (Two Axles & GVWR of more than 10,000 pounds)
02-Single Unit Truck (Three or More Axles)
04-Truck Pulling/Trailer(s)
05-Truck Tractor (Bobtail)
06-Truck Tractor/Semi-Trailer
07-Truck Tractor/Double
08-Truck Tractor/Triple
10-Vehicle 10,000 pounds or less placarded for hazardous materials
19-Truck more than 10,000 lbs, cannot classify
20-Bus (seats 9-15 occupants, including driver)
21-Bus (seats for more than 15 occupants, including driver)
99-Unknown

CARGO BODY TYPE [00]

00-Not Applicable
01-Van/Enclosed Box
02-Cargo Tank
03-Flatbed
04-Dump
05-Concrete Mixer
06-Auto Transporter
07-Garbage / Refuse
08-Grain/Chips/Gravel
09-Pole - Trailer
10-Log
11-Intermodal Container Chassis
12-Vehicle Towing Another Motor Vehicle
22-Bus
96-No Cargo Body Type
97-Other
98-Unknown Cargo Body Type
99-Unknown

3-Underride Compartment Intrusion Unknown)
7-Overriding a Motor Vehicle In-Transport
8-Overriding a Motor Vehicle Not In-Transport
9-Unknown if Underride or Override

ROLLOVER [1]

0 - No Rollover
1 - Rollover, Tripped by Object/Vehicle
2 - Rollover, Untripped
9 - Rollover, Unknown Type

LOCATION OF ROLLOVER [3]

0 - No Rollover
1 - On Roadway
2 - On Shoulder
3 - On Median/Separator
4 - In Gore
5 - On Roadside
6 - Outside of Traffic way
7 - In Parking Lane/Zone
9 - Unknown

AREAS OF IMPACT - INITIAL CONTACT POINT [02]

00-Non-Collision
01-12-Clock Values
13-Top
14-Undercarriage
18-Cargo/Vehicle Parts Set-In-Motion
19-Other Objects Set-In-Motion
20-Object Set in Motion, Unknown if Cargo/Vehicle Parts or Other
61 – Left
62 – Left-Front Side
63 – Left-Back Side
81 - Right
82 – Right-Front Side
83 - Right-Back Side
98 – Not Reported
99 - Unknown

DAMAGED AREAS [02, 13]

(SELECT ALL THAT APPLY)
01 - 12-Clock Values
13 - Top
14 - Undercarriage
15 - No Damage
99 - Damage Areas Unknown

EXTENT OF DAMAGE [6]

0-No Damage
2-Minor Damage
4-Functional Damage
6-Disabling Damage
8-Not Reported
9-Unknown

VEHICLE REMOVAL [2]

2-Towed Due to Disabling Damage
3-Towed Not Due to Disabling Damage
5 - Not Towed
8-Not Reported
9-Unknown

SEQUENCE OF EVENTS [63, 59, 69, 64, 57, 01, 54]

LOOKUP SEQUENCE OF EVENTS CODES

MOST HARMFUL EVENT [01]

LOOKUP MOST HARMFUL EVENT CODES

RELATED FACTORS [00|00]

LOOKUP RELATED FACTORS - VEHICLE LEVEL CODES

FIRE OCCURRENCE [0]

0-No or Not Reported
1-Yes

Appendix E

Crash Data Collection Form

U.S. Department of Transportation
National Highway Traffic Safety Administration

CRASH LEVEL

STATE NUMBER (GSA CODES)	CONSECUTIVE NUMBER	Number of Forms Submitted for Persons Not in Motor Vehicles	Number of Vehicle Forms Submitted	Number of Motor Vehicle Occupant Forms Submitted
1	94	0	2	3

COUNTY 097
Actual GSA Code Except for:
000-Not Applicable
997-Other
998-Not Reported
999-Unknown

CITY 0000
Actual GSA Code Except for:
0000-Not Applicable
9997-Other
9898-Not Reported
9999-Unknown

CRASH DATE 03162017
Actual Month and Day Except for:

CRASH TIME 0720
Military Time:
9999-Unknown

TRAFFIC IDENTIFIER
Actual Posted Number, Assigned Number, or Common Name
(If No Posted or Assigned Number) Except: Nine-Fill if Unknown
I-10
GRAND BAY WILMER RD

ROUTE SIGNING 1
1-Interstate
2-U.S. Highway
3-State Highway
4-County Road
LOCAL STREET
5-Township
6-Municipality
7-Frontage Road
8-Other
9-Unknown

LANDUSE 01
LandArea
1-Rural
2-Urban
6-Trafficway Not in State Inventory
8-Not Reported
9-Unknown

FUNCTIONAL SYSTEM 01
Functional System
01-Interstate
02-Principal Arterial-Other Freeways and Expressway

TYPE OF INTERSECTION 1
01-Not an Intersection
02-Four-way Intersection
03-T-Intersection
04-Y-Intersection
05-Traffic Circle
06-Roundabout
07-Five Point, or More
10-L-Intersection
98-Not Reported
99-Unknown

RELATION TO TRAFFICWAY 04
01-On Roadway
02-On Shoulder
03-On Median
04-On Roadside
05-Outside Trafficway
06-Off Roadway - Location Unknown
07-In Parking Lane/Zone
08-Gore
10-Separator
11-Continuous Left-Turn Lane
98-Not Reported
99-Unknown

WORK ZONE 1
0-None
1-Construction
2-Maintenance
3-Utility
4-Work Zone, Type Unknown

LIGHT CONDITION 1
1-Daylight
2-Dark-Not Lighted
3-Dark-Lighted
4-Dawn
5-Dusk
6-Dark-Unknown Lighting
7-Other
8-Not Reported
9-Unknown

ATMOSPHERIC CONDITIONS 1 | 0
00-No Additional Atmospheric
04-Snow
10-Cloudy

https://www-fars.nhtsa.dot.gov/QueryTool/QuerySection/AccidentDisplayForm.aspx?ShowData=accform&CaseNum=94&StateNum=1&CaseYear=2017

03-Principal Arterial-Other
04-Minor Arterial
05-Major Collector
06-Minor Collector
07-Local
96-Trafficway Not in State Inventory
98-Not Reported
99-Unknown

OWNERSHIP 1

(See Instruction Manual)

NATIONAL HIGHWAY SYSTEM 1

0-This section is NOT on the NHS
1-This section is ON the NHS
9-Unknown if this section is on the NHS

LOOKUP NATIONAL HIGHWAY SYSTEM CODES

SPECIAL JURISDICTION 0

0-No Special Jurisdiction
1-National Park Service
2-Military
3-Indian Reservation
4-College/University Campus
5-Other Federal Properties
8-Other
9-Unknown

LOOKUP SPECIAL JURISDICTION CODES

MILEPOINT 00040

Actual to Nearest .1 Mile (Assumed Decimal)
Except: 00000-None 99998-Not Reported 99999-Unknown

GLOBAL POSITION

	Degrees	Minutes	Seconds
LATITUDE:	30	29	45.17
LONGITUDE:	088	20	18.23

FIRST HARMFUL EVENT 59

LOOKUP FIRST HARMFUL EVENT CODES

MANNER OF COLLISION 00

00-Not Collision with Motor Vehicle In-Transport
01-Front-to-Rear
02-Front-to-Front
06-Angle
07-Sideswipe-Same Direction
08-Sideswipe-Opposite Direction
09-Rear-to-Side
10-Rear-to-Rear
11-Other
98-Not Reported
99-Unknown

RELATION TO JUNCTION 01 00

Conditions
01-Clear
02-Rain
03-Sleet or Hail
05-Fog, Smog, Smoke
06-Severe Crosswinds
07-Blowing Sand, Soil, Dirt
08-Other
11-Blowing Snow
12-Freezing Rain or Drizzle
98-Not Reported
99-Unknown

SCHOOL BUS RELATED 0

0-No 1-Yes

RAIL GRADE CROSSING IDENTIFIER 0000000

NOTIFICATION TIME EMS 0722

Military Time Except:
8888-Not Notified 9998-Unknown if Notified 9999-Unknown

ARRIVAL TIME EMS 0732

Military Time Except:
8888-Not Notified 9998-Unknown if Arrived
9997-Officially Canceled 9999-Unknown

EMS TIME AT HOSPITAL 9999

Military Time Except:
8888-Not Applicable(Not Transported) 9998-Unknown if Transported
9996-Terminated Transport 9999-Unknown EMS Hospital Arrival Time
9997-Officially Canceled

RELATED FACTORS 14|0|0

LOOKUP ACCIDENT RELATED FACTORS CODES

SPECIFIC LOCATION	WITHIN INTERCHANGE AREA	
01-Non-Junction 02-Intersection 03-Intersection Related 04-Driveway Access 05-Entrance/Exit Ramp Related 06-Rail Grade Crossing 07-Crossover-Related 08-Driveway Access Related 16-Shared-Use Path Crossing 17-Acceleration/Deceleration Lane 18-Through Roadway 19-Other Location Within Interchange Area 20-Entrance/Exit Ramp 98-Not Reported 99-Unknown	0-No 1-Yes 8-Not Reported 9-Unknown	

CRASH LEVEL - CRASH EVENTS

Case Number: 94

Event No.	**Veh No.(This)**	**AOI (This)**	**SOE**	**Veh No.(Other)**	**AOI (Other)**
1	1		63		
2	1	2	59		
3	1		69		
4	1		64		
5	1	12	57		
6	1	0	1		
7	1	18	54	2	12

Event No.	**Veh No. (This)**	**AOI (This)**	**SOE**	**Veh No. (Other)**	**AOI (Other)**
1	1		Ran Off Roadway - Right		
2	1	2 Clock Point	Traffic Sign Support		
3	1		Re-entering Roadway		
4	1		Ran Off Roadway - Left		
5	1	12 Clock Point	Cable Barrier		
6	1	Non-Collision	Rollover/Overturn		
7	1	Cargo/Vehicle Parts Set-In-Motion	Motor Vehicle In-Transport Strikes or is Struck by Cargo, Persons or Objects Set-in-Motion from/by Another Motor Vehicle In Transport	2	12 Clock Point

Appendix F

Partial Effect of Ordered Probability Models

Table (F.1) Significant Factors' Partial Effect For Random Parameter Ordered Probit Model

Factors	Partial Effect	Elasticity	p-value
Partial effects on Prob [Y=00 or PDO] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	-0.029	-1.868	0.099
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	-0.082	-5.222	0.089
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	-0.057	-3.581	0.093
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	0.133	8.429	0.050
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	-0.013	-0.812	0.094
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	-0.018	-1.116	0.260
Traffic control devices (1 if used, 0 otherwise)	0.221	14.014	0.032
Manner of Collision (1 if rear-end, 0 otherwise)	0.042	2.641	0.171
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	-0.110	-6.961	0.058
Work zone signage (1 if warning sign is available, 0 otherwise)	0.148	9.404	0.128
Heavy truck involvement (1 if it is involved, 0 otherwise)	-0.241	-15.274	0.011

Table F.1 continued from previous page

Factors	Partial Effect	Elasticity	p-value
Roll over (1 if roll-over didn't occur, 0 otherwise)	0.045	2.881	0.119
Gender of the driver (0 if male, 1 if female)	0.032	1.999	0.191
Partial effects on Prob [Y=01 or Minor Injury] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	-0.341	-0.535	0.006
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	-0.165	-0.259	0.007
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	-0.161	-0.253	0.054
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	0.179	0.281	0.107
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	-0.753	-1.182	0.000
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	-0.441	-0.693	0.004
Traffic control devices (1 if used, 0 otherwise)	0.304	0.478	0.041
Manner of Collision (1 if rear-end, 0 otherwise)	0.345	0.541	0.016
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	-0.209	-0.329	0.057
Work zone signage (1 if warning sign is available, 0 otherwise)	0.201	0.315	0.096
Heavy truck involvement (1 if it is involved, 0 otherwise)	-0.118	-0.185	0.355
Roll over (1 if roll-over didn't occur, 0 otherwise)	0.753	1.183	0.000
Gender of the driver (0 if male, 1 if female)	0.282	0.442	0.008
Partial effects on Prob [Y=02 or Major Injury] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	0.262	0.995	0.005
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	0.185	0.704	0.004
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	0.163	0.618	0.010
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-0.233	-0.884	0.003
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	-0.184	-0.697	0.000
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	0.308	1.170	0.002
Traffic control devices (1 if used, 0 otherwise)	-0.374	-1.421	0.002

Table F.1 continued from previous page

Factors	Partial Effect	Elasticity	p-value
Manner of Collision (1 if rear-end, 0 otherwise)	-0.275	-1.043	0.009
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	0.237	0.899	0.003
Work zone signage (1 if warning sign is available, 0 otherwise)	-0.259	-0.982	0.013
Heavy truck involvement (1 if it is involved, 0 otherwise)	0.268	1.019	0.002
Roll over (1 if roll-over didn't occur, 0 otherwise)	-0.337	-1.279	0.000
Gender of the driver (0 if male, 1 if female)	-0.226	-0.858	0.005
Partial effects on Prob [Y=03 or Fatality] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	0.108	1.290	0.002
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	0.062	0.741	0.000
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	0.055	0.653	0.001
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-0.079	-0.946	0.000
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	0.949	11.324	0.000
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	0.151	1.799	0.017
Traffic control devices (1 if used, 0 otherwise)	-0.151	-1.801	0.002
Manner of Collision (1 if rear-end, 0 otherwise)	-0.112	-1.331	0.013
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	0.082	0.983	0.000
Work zone signage(1 if warning sign is available, 0 otherwise)	-0.090	-1.075	0.007
Heavy truck involvement (1 if it is involved, 0 otherwise)	0.091	1.080	0.000
Roll over (1 if roll-over didn't occur, 0 otherwise)	-0.462	-5.508	0.000
Gender of the driver (0 if male, 1 if female)	-0.087	-1.038	0.001

Table (F.2) Significant Factors' Partial Effect For Random Parameter Ordered Logit Model

Factors	Partial Effect	Elasticity	t-stat
Partial effects on Prob [Y=00 or PDO] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	-0.015	-2.223	-1.810
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	-0.025	-3.728	-1.840
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	-0.018	-2.617	-1.770
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	0.032	4.807	1.790
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	-0.009	-1.322	-1.700
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	-0.011	-1.689	-1.760
Traffic control devices (1 if used, 0 otherwise)	0.063	9.402	1.660
Manner of Collision (1 if rear-end, 0 otherwise)	0.009	1.412	1.600
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	-0.038	-5.723	-1.940
Work zone signage (1 if warning sign is available, 0 otherwise)	0.069	10.274	1.530
Heavy truck involvement (1 if it is involved, 0 otherwise)	-0.072	-10.721	-1.920
Roll over (1 if roll-over didn't occur, 0 otherwise)	0.019	2.763	1.890
Gender of the driver (0 if male, 1 if female)	0.010	1.554	1.740
Partial effects on Prob [Y=01 or Minor Injury] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	-0.694	-1.144	-7.610
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	-0.407	-0.671	-2.490
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	-0.366	-0.604	-2.360
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	0.444	0.732	2.500
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	-0.783	-1.291	-8.480
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	-0.769	-1.267	-15.280
Traffic control devices (1 if used, 0 otherwise)	0.683	1.126	4.110

Table F.2 continued from previous page

Factors	Partial Effect	Elasticity	t-stat
Manner of Collision (1 if rear-end, 0 otherwise)	0.385	0.634	2.060
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	-0.539	-0.888	-3.010
Work zone signage (1 if warning sign is available, 0 otherwise)	0.597	0.983	3.130
Heavy truck involvement (1 if it is involved, 0 otherwise)	-0.477	-0.787	-2.390
Roll over (1 if roll-over didn't occur, 0 otherwise)	0.862	1.420	24.570
Gender of the driver (0 if male, 1 if female)	0.437	0.721	3.100
Partial effects on Prob [Y=02 or Major Injury] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	0.587	1.616	10.010
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	0.407	1.120	2.650
Alcohol or Drug involvement (1 if any or both is/are not involved, 0 otherwise)	0.361	0.995	2.440
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-0.447	-1.231	-2.700
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	-0.201	-0.552	-2.040
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	0.450	1.238	3.080
Traffic control devices (1 if used, 0 otherwise)	-0.658	-1.812	-4.880
Manner of Collision (1 if rear-end, 0 otherwise)	-0.365	-1.006	-2.140
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	0.534	1.470	3.280
Work zone signage (1 if warning sign is available, 0 otherwise)	-0.607	-1.672	-3.580
Heavy truck involvement (1 if it is involved, 0 otherwise)	0.514	1.415	2.870
Roll over (1 if roll-over didn't occur, 0 otherwise)	-0.141	-0.389	-0.870
Gender of the driver (0 if male, 1 if female)	-0.412	-1.133	-3.180
Partial effects on Prob [Y=03 or Fatality] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	0.122	5.224	2.170
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	0.026	1.092	2.660
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	0.022	0.955	2.560

Table F.2 continued from previous page

Factors	Partial Effect	Elasticity	t-stat
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-0.029	-1.255	-2.590
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	0.993	42.464	389.510
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	0.330	14.123	2.020
Traffic control devices (1 if used, 0 otherwise)	-0.088	-3.760	-1.840
Manner of Collision (1 if rear-end, 0 otherwise)	-0.029	-1.237	-1.410
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	0.043	1.859	2.750
Work zone signage (1 if warning sign is available, 0 otherwise)	-0.059	-2.504	-1.940
Heavy truck involvement (1 if it is involved, 0 otherwise)	0.036	1.523	2.660
Roll over (1 if roll-over didn't occur, 0 otherwise)	-0.739	-31.604	-4.980
Gender of the driver (0 if male, 1 if female)	-0.036	-1.544	-2.120

Table (F.3) Significant Factors' Partial Effect For Random Parameter Ordered Arctangent Model

Factors	Partial Effect	Elasticity	t-stat
Partial effects on Prob [Y=00 or PDO] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	-0.018	-2.068	-2.060
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	-0.027	-3.108	-2.050
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	-0.020	-2.294	-1.990
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	0.034	3.936	1.990
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	-0.012	-1.329	-1.920
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	-0.014	-1.613	-2.000
Traffic control devices (1 if used, 0 otherwise)	0.065	7.544	1.830
Manner of Collision (1 if rear-end, 0 otherwise)	0.011	1.268	1.760
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	-0.041	-4.687	-2.160
Work zone signage (1 if warning sign is available, 0 otherwise)	0.073	8.399	1.680
Heavy truck involvement (1 if it is involved, 0 otherwise)	-0.074	-8.568	-2.090
Roll over (1 if roll-over didn't occur, 0 otherwise)	0.022	2.498	2.130
Gender of the driver (0 if male, 1 if female)	0.012	1.414	1.970
Partial effects on Prob [Y=01 or Minor Injury] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	-0.713	-1.173	-8.420
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	-0.405	-0.666	-2.160
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	-0.370	-0.609	-2.050
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	0.443	0.729	2.210
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	-0.793	-1.305	-8.870
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	-0.771	-1.269	-16.700
Traffic control devices (1 if used, 0 otherwise)	0.694	1.142	4.260

Table F.3 continued from previous page

Factors	Partial Effect	Elasticity	t-stat
Manner of Collision (1 if rear-end, 0 otherwise)	0.386	0.635	1.860
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	-0.546	-0.898	-2.780
Work zone signage (1 if warning sign is available, 0 otherwise)	0.613	1.008	3.120
Heavy truck involvement (1 if it is involved, 0 otherwise)	-0.484	-0.797	-2.210
Roll over (1 if roll-over didn't occur, 0 otherwise)	0.860	1.414	29.090
Gender of the driver (0 if male, 1 if female)	0.447	0.734	2.770
Partial effects on Prob [Y=02 or Major Injury] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	0.611	1.707	11.550
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	0.405	1.130	2.280
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	0.366	1.022	2.120
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-0.446	-1.247	-2.360
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	-0.186	-0.518	-1.920
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	0.480	1.339	3.230
Traffic control devices (1 if used, 0 otherwise)	-0.673	-1.878	-5.060
Manner of Collision (1 if rear-end, 0 otherwise)	-0.369	-1.030	-1.900
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	0.542	1.514	3.000
Work zone signage (1 if warning sign is available, 0 otherwise)	-0.625	-1.744	-3.550
Heavy truck involvement (1 if it is involved, 0 otherwise)	0.520	1.453	2.600
Roll over (1 if roll-over didn't occur, 0 otherwise)	-0.138	-0.384	-0.840
Gender of the driver (0 if male, 1 if female)	-0.423	-1.182	-2.820
Partial effects on Prob [Y=03 or Fatality] at means			
Collision Hour (1 if the collision occurred between 8 PM and 6 AM, 0 otherwise)	0.120	4.778	2.240
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	0.027	1.076	2.980
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	0.024	0.960	2.850

Table F.3 continued from previous page

Factors	Partial Effect	Elasticity	t-stat
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-0.031	-1.230	-2.890
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	0.990	39.433	339.330
Driver related factor (1 if Failing to keep in proper lane or run off the road or following improperly, 0 otherwise)	0.306	12.179	1.840
Traffic control devices (1 if used, 0 otherwise)	-0.087	-3.470	-1.980
Manner of Collision (1 if rear-end, 0 otherwise)	-0.028	-1.125	-1.500
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	0.045	1.775	3.050
Work zone signage (1 if warning sign is available, 0 otherwise)	-0.061	-2.430	-2.130
Heavy truck involvement (1 if it is involved, 0 otherwise)	0.038	1.520	2.990
Roll over (1 if roll-over didn't occur, 0 otherwise)	-0.744	-29.625	-4.900
Gender of the driver (0 if male, 1 if female)	-0.035	-1.413	-2.300

Appendix G

10 Fold Cross Validation Models

Table (G.1) Random Parameter Ordered Logit Model for Fold 1

List of statistically significant parameters	**Coefficient**	**p-value**
Constant	10.166	0.000
Collision Hour (1 if collision occurred between 8PM and 6 AM, 0 otherwise)	3.389	0.000
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	2.189	0.000
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	1.125	0.007
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-1.998	0.000
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	10.750	0.000
Driver related factor (1 if Failing to keep in proper lane or run-off-road or following improperly, 0 otherwise)	3.304	0.000
The month of collision (1 if the collision occurred between December and March, 0 otherwise)	-3.159	0.002
Traffic control devices (1 if used, 0 otherwise)	-4.880	0.000
Manner of Collision (1 if rear-end, 0 otherwise)	-1.815	0.007
Work zone signage (1 if a warning sign is available, 0 otherwise)	-2.860	0.003
Roll over (1 if roll-over didn't occur, 0 otherwise)	-4.693	0.000
(Std. dev. of parameter distribution) Normally distributed	*2.616*	*0.000*
Gender of the driver (0 if male, 1 if female)	-1.644	0.000
(Std. dev. of parameter distribution) Normally distributed	*4.637*	*0.000*
Heavy truck involvement (1 if it is involved, 0 otherwise)	3.216	0.000
(Std. dev. of parameter distribution) Normally distributed	*3.564*	*0.000*
Mu(01)	4.158	0.000
Mu(02)	6.651	0.000
Log-likelihood function	-247.571	
Likelihood at zero	-352.878	
ρ^2	0.298	
Adjusted ρ^2	0.245	
AIC	533.142	
BIC	598.3831	

Table (G.2) Random Parameter Ordered Logit Model for Fold 2

List of statistically significant parameters	**Coefficient**	**p-value**
Constant	6.525	0.000
Collision Hour (1 if collision occurred between 8 PM and 6 AM, 0 otherwise)	1.970	0.000
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	2.949	0.000
Road Geometry (1 if there is a junction around the work zone, 0 otherwise)	6.733	0.004
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	1.604	0.000
Speed at the time of the collision (1 if it is less than 50 mph, 0 otherwise)	-2.518	0.000
Driver related factor (1 if Failing to keep in proper lane or run-off-road or following improperly, 0 otherwise)	2.554	0.000
Manner of Collision (1 if rear-end, 0 otherwise)	-2.503	0.000
Roll over (1 if roll-over didn't occur, 0 otherwise)	-2.549	0.000
(Std. dev. of parameter distribution) Normally distributed	*4.070*	*0.000*
Gender of the driver (0 if male, 1 if female)	-2.128	0.000
(Std. dev. of parameter distribution) Normally distributed	*5.662*	*0.000*
Heavy truck involvement (1 if it is involved, 0 otherwise)	2.593	0.000
(Std. dev. of parameter distribution) Normally distributed	*1.892*	*0.000*
Mu(01)	4.394	0.000
Mu(02)	7.058	0.000
Log-likelihood function	-247.403	
Likelihood at zero	-330.918	
ρ^2	0.252	
Adjusted ρ^2	0.204	
AIC	526.806	
BIC	581.885	

Table (G.3) Random Parameter Ordered Logit Model for Fold 3

List of statistically significant parameters	**Coefficient**	**p-value**
Constant	14.371	0.000
Collision Hour (1 if collision occurred between 8 PM and 6 AM, 0 otherwise)	3.059	0.000
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	2.280	0.000
Road Geometry (1 if there is a junction around the work zone, 0 otherwise)	10.199	0.000
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	1.446	0.005
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-4.561	0.000
Surface Condition (0 if dry, 1 if wet, slushy, etc.)	-3.091	0.001
Traffic control devices (1 if used, 0 otherwise)	-3.271	0.000
Driver related factor (1 if Failing to keep in proper lane or run-off-road or following improperly, 0 otherwise)	6.319	0.000
The month of collision (1 if the collision occurred between December and March, 0 otherwise)	-5.861	0.000
Roll over (1 if roll-over didn't occur, 0 otherwise)	-7.500	0.000
(Std. dev. of parameter distribution) Normally distributed	4.483	0.000
Gender of the driver (0 if male, 1 if female)	-2.569	0.000
(Std. dev. of parameter distribution) Normally distributed	6.975	0.000
Heavy truck involvement (1 if it is involved, 0 otherwise)	4.361	0.000
(Std. dev. of parameter distribution) Normally distributed	6.088	0.000
Mu(01)	5.909	0.000
Mu(02)	9.356	0.000
Log-likelihood function	-244.082	
Likelihood at zero	-329.432	
ρ^2	0.259	
Adjusted ρ^2	0.204	
AIC	524.164	
BIC	586.049	

Table (G.4) Random Parameter Ordered Logit Model for Fold 4

List of statistically significant parameters	**Coefficient**	**p-value**
Constant	7.554	0.000
Collision Hour (1 if collision occurred between 8 PM and 6 AM, 0 otherwise)	2.924	0.000
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	1.532	0.000
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	0.964	0.019
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-2.600	0.000
Traffic control devices (1 if used, 0 otherwise)	-2.513	0.001
Driver related factor (1 if Failing to keep in proper lane or run-off-road or following improperly, 0 otherwise)	3.482	0.000
The month of collision (1 if the collision occurred between December and March, 0 otherwise)	-2.788	0.004
Work zone signage (1 if warning sign is available, 0 otherwise)	-1.943	0.013
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	1.563	0.000
Roll over (1 if roll-over didn't occur, 0 otherwise)	-4.580	0.000
(Std. dev. of parameter distribution) Normally distributed	2.880	0.000
Gender of the driver (0 if male, 1 if female)	-1.683	0.000
(Std. dev. of parameter distribution) Normally distributed	4.185	0.000
Heavy truck involvement (1 if it is involved, 0 otherwise)	1.975	0.000
(Std. dev. of parameter distribution) Normally distributed	2.831	0.000
Mu(01)	3.705	0.000
Mu(02)	5.937	0.000
Log-likelihood function	-247.201	
Likelihood at zero	-332.815	
ρ^2	0.257	
Adjusted ρ^2	0.203	
AIC	530.402	
BIC	592.443	

Table (G.5) Random Parameter Ordered Logit Model for Fold 5

List of statistically significant parameters	**Coefficient**	**p-value**
Constant	8.653	0.000
Collision Hour (1 if collision occurred between 8 PM and 6 AM, 0 otherwise)	2.681	0.000
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	2.122	0.000
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	1.868	0.000
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-1.646	0.002
Driver related factor (1 if Failing to keep in proper lane or run-off-road or following improperly, 0 otherwise)	4.617	0.000
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	1.911	0.000
The month of collision (1 if the collision occurred between December and March, 0otherwise)	-3.711	0.001
Work zone signage (1 if warning sign is available, 0 otherwise)	-2.301	0.008
Traffic control devices (1 if used, 0 otherwise)	-2.750	0.001
Manner of Collision (1 if rear-end, 0 otherwise)	-1.778	0.009
Roll over (1 if roll-over didn't occur, 0 otherwise)	-4.719	0.000
(Std. dev. of parameter distribution) Normally distributed	3.202	0.000
Gender of the driver (0 if male, 1 if female)	-1.184	0.007
(Std. dev. of parameter distribution) Normally distributed	*5.971*	*0.000*
Heavy truck involvement (1 if it is involved, 0 otherwise)	2.530	0.000
(Std. dev. of parameter distribution) Normally distributed	*2.924*	*0.000*
Road Geometry (1 if there is a junction around the work zone, 0 otherwise)	8.414	0.000
(Std. dev. of parameter distribution) Normally distributed	*5.780*	*0.004*
Mu(01)	4.465	0.000
Mu(02)	7.207	0.000
Log-likelihood function	-242.479	
Likelihood at zero	-329.502	
ρ^2	0.264	
Adjusted ρ^2	0.209	
AIC	526.958	
BIC	599.158	

Table (G.6) Random Parameter Ordered Logit Model for Fold 6

List of statistically significant parameters	**Coefficient**	**p-value**
Constant	8.253	0.000
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	1.224	0.002
Speed at the time of the collision (1 if it is less than 50 mph, 0 otherwise)	-2.571	0.000
Driver related factor (1 if Failing to keep in proper lane or run-off-road or following improperly, 0 otherwise)	3.304	0.000
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	2.752	0.000
Road Geometry (1 if there is a junction around the work zone, 0 otherwise)	8.192	0.000
Traffic control devices (1 if used, 0 otherwise)	-1.520	0.001
Manner of Collision (1 if rear-end, 0 otherwise)	-3.298	0.000
Rollover (1 if roll-over didn't occur, 0 otherwise)	-3.831	0.000
(Std. dev. of parameter distribution) Normally distributed	*4.835*	*0.000*
Gender of the driver (0 if male, 1 if female)	-3.141	0.000
(Std. dev. of parameter distribution) Normally distributed	*0.985*	*0.000*
Heavy truck involvement (1 if it is involved, 0 otherwise)	2.913	0.000
(Std. dev. of parameter distribution) Normally distributed	*5.232*	*0.000*
Mu(01)	4.287	0.000
Mu(02)	6.670	0.000
Log-likelihood function	-245.799	
Likelihood at zero	-329.432	
ρ^2	0.254	
Adjusted ρ^2	0.205	
AIC	523.598	
BIC	578.607	

Table (G.7) Random Parameter Ordered Logit Model for Fold 7

List of statistically significant parameters	**Coefficient**	**p-value**
Constant	9.385	0.000
Collision Hour (1 if collision occurred between 8 PM and 6 AM, 0 otherwise)	2.705	0.000
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	1.596	0.000
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	1.564	0.000
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-2.137	0.000
Traffic control devices (1 if used, 0 otherwise)	-4.810	0.000
Driver related factor (1 if Failing to keep in proper lane or run off road or following improperly, 0 otherwise)	3.230	0.000
Work zone signage (1 if warning sign is available, 0 otherwise)	-3.411	0.001
Month of collision (1 if the collision occurred between December and March, 0 otherwise)	-2.611	0.005
Location of the collision (1 if Median and Shoulder, 0 otherwise)	1.233	0.080
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	11.070	0.000
Roll over (1 if roll-over didn't occur, 0 otherwise)	-4.312	0.000
(Std. dev. of parameter distribution) Normally distributed	3.930	0.000
Gender of the driver (0 if male, 1 if female)	-1.203	0.004
(Std. dev. of parameter distribution) Normally distributed	4.679	0.000
Heavy truck involvement (1 if it is involved, 0 otherwise)	2.496	0.000
(Std. dev. of parameter distribution) Normally distributed	0.792	0.001
Mu(01)	3.935	0.000
Mu(02)	6.391	0.000
Log-likelihood function	-243.006	
Likelihood at zero	-328.307	
ρ^2	0.260	
Adjusted ρ^2	0.202	
AIC	524.012	
BIC	589.253	

Table (G.8) Random Parameter Ordered Logit Model for Fold 8

List of statistically significant parameters	**Coefficient**	**p-value**
Constant	9.038	0.000
Collision Hour (1 if collision occurred between 8 PM and 6 AM, 0 otherwise)	2.440	0.000
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	2.659	0.000
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	-2.097	0.000
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-2.806	0.000
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	12.425	0.000
Driver related factor (1 if Failing to keep in proper lane or run off road or following improperly, 0 otherwise)	5.343	0.000
Surface Condition (0 if dry, 1 if wet, slushy, etc.)	-2.269	0.009
Month of collision (1 if the collision occurred between December and March, 0 otherwise)	-4.541	0.000
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	3.157	0.000
Roll over (1 if roll-over didn't occur, 0 otherwise)	-6.818	0.000
(Std. dev. of parameter distribution) Normally distributed	*3.004*	*0.000*
Gender of the driver (0 if male, 1 if female)	-1.536	0.001
(Std. dev. of parameter distribution) Normally distributed	*6.716*	*0.000*
Heavy truck involvement (1 if it is involved, 0 otherwise)	2.740	0.000
(Std. dev. of parameter distribution) Normally distributed	*4.845*	*0.000*
Mu(01)	5.060	0.000
Mu(02)	8.182	0.000
Log-likelihood function	-246.201	
Likelihood at zero	-332.535	
ρ^2	0.260	
Adjusted ρ^2	0.205	
AIC	528.402	
BIC	590.443	

Table (G.9) Random Parameter Ordered Logit Model for Fold 9

List of statistically significant parameters	**Coefficient**	**p-value**
Constant	8.536	0.000
Collision Hour (1 if collision occurred between 8 PM and 6 AM, 0 otherwise)	2.467	0.000
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	1.671	0.000
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	1.136	0.007
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-1.754	0.000
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	10.684	0.000
Driver related factor (1 if Failing to keep in proper lane or run off road or following improperly, 0 otherwise)	3.452	0.000
Month of collision (1 if the collision occurred between December and March, 0 otherwise)	-2.961	0.006
The month of collision (1 if the collision occurred between April and July, 0 otherwise)	1.173	0.007
Work zone signage (1 if warning sign is available, 0 otherwise)	-1.844	0.041
Traffic control devices (1 if used, 0 otherwise)	-3.027	0.001
Manner of Collision (1 if rear-end, 0 otherwise)	-2.219	0.001
Roll over (1 if roll-over didn't occur, 0 otherwise)	-3.812	0.000
(Std. dev. of parameter distribution) Normally distributed	*3.817*	*0.000*
Gender of the driver (0 if male, 1 if female)	-1.173	0.005
(Std. dev. of parameter distribution) Normally distributed	*5.250*	*0.000*
Heavy truck involvement (1 if it is involved, 0 otherwise)	2.218	0.000
(Std. dev. of parameter distribution) Normally distributed	*0.921*	*0.000*
Mu(01)	4.083	0.000
Mu(02)	6.575	0.000
Log-likelihood function	-243.029	
Likelihood at zero	-329.432	
ρ^2	0.262	
Adjusted ρ^2	0.202	
AIC	526.058	
BIC	594.820	

Table (G.10) Random Parameter Ordered Logit Model for Fold 10

List of statistically significant parameters	**Coefficient**	**p-value**
Constant	4.866	0.000
Collision Hour (1 if collision occurred between 8 PM and 6 AM, 0 otherwise)	1.321	0.000
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	0.795	0.000
Alcohol or Drug involvement (1 if any or both is/are involved, 0 otherwise)	-0.942	0.000
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-1.748	0.000
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	5.867	0.000
Driver related factor (1 if Failing to keep in proper lane or run off road or following improperly, 0 otherwise)	1.791	0.000
Month of collision (1 if the collision occurred between December and March, 0 otherwise)	-1.648	0.002
Traffic control devices (1 if used, 0 otherwise)	-1.220	0.000
Manner of Collision (1 if rear-end, 0 otherwise)	-1.449	0.000
Roll over (1 if roll-over didn't occur, 0 otherwise)	-2.169	0.000
(Std. dev. of parameter distribution) Normally distributed	*2.795*	*0.000*
Gender of the driver (0 if male, 1 if female)	-0.730	0.004
(Std. dev. of parameter distribution) Normally distributed	*2.351*	*0.000*
Heavy truck involvement (1 if it is involved, 0 otherwise)	1.955	0.000
(Std. dev. of parameter distribution) Normally distributed	*1.240*	*0.000*
Mu(01)	2.564	0.000
Mu(02)	4.064	0.000
Log-likelihood function	-247.303	
Likelihood at zero	-330.631	
ρ^2	0.252	
Adjusted ρ^2	0.201	
AIC	530.606	
BIC	592.570	