

Traffic Conflict-based Road Safety Analysis: Data Requirements and Evaluation of Safety
Countermeasures

by

Jessica May Ting Keung

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil Engineering

Waterloo, Ontario, Canada, 2024

©Jessica May Ting Keung 2024

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the 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.

Statement of Contributions

Jessica Keung was the sole author for Chapters 1, 2, and 5 which were written under the supervision of Dr. Liping Fu and were not written for presentation or publication.

This thesis consists in part of two manuscripts written for presentation and publication. Exceptions to sole authorship of material are as follows:

Research presented in Chapters 3 and 4:

In Chapter 3: Sample Size Requirements for Conflict-based Studies, this research was conducted at the University of Waterloo by Jessica Keung under the supervision of Dr. Liping Fu. Dr. Chris Bachmann contributed to background information about the statistical methods described and Dr. Ting Fu provided input about the case study used. Jessica Keung wrote the draft manuscript, which all co-authors contributed intellectual input on.

The research in Chapter 4: Conflict-based Study for Evaluating the Effectiveness of Vision Zero Safety Countermeasures was conducted at the University of Waterloo under the supervision of Dr. Liping Fu. Dr. Qiangqiang Shangguan and Dr. Liping Fu were the primary co-investigators on the City of Toronto project. Jessica Keung designed the study with consultation from City of Toronto staff, coordinated the data collection with Transoft Solutions, conducted the exploratory data analysis, and drafted relevant portions of the manuscript. Lana Samara from Transoft Solutions deployed the cameras for data collection and processed the video data using the TrafSAFE software. Dr. Shangguan conducted the cross-sectional analysis, prepared the relevant portions for the draft manuscript, and prepared the manuscript for publication, which all co-authors contributed intellectual input on.

Citations:

Chapter 3: Keung, J., Bachmann, C., Fu, T., Fu, L. (2020). A Comprehensive Framework for Determining the Optimal Sample Size for Conflict-based Traffic Safety Analysis [Poster Presentation]. *The 99th Annual Meeting of the Transportation Research Board*, Washington, D.C.,

Chapter 4: Shangguan, Q., Keung, J., Fu, L., Samara, L., Wang, J., & Fu, T. (2023). Do traffic countermeasures improve the safety of vulnerable road users at signalized intersections? A combination of case-control and cross-sectional studies using video-based traffic conflicts. *Transportation Research Record: Journal of the Transportation Research Board*, 2678(1), 806–819. <https://doi.org/10.1177/03611981231172748>

I was responsible for conceptualizing study design, coordinating data collection and analysis, and drafting and submitting manuscripts. My coauthors provided guidance during each step of the research and provided feedback on draft manuscripts.

Abstract

Driven by the vision to eliminate road fatalities, Vision Zero initiatives have been widely adopted by many cities around the world, with significant investments of resources in various safety programs and countermeasures. Conflict-based traffic safety analysis is a burgeoning field, but many studies have failed to address the important question of how much data should be collected to make credible safety-related inferences and how the effectiveness of safety countermeasures could be quantified using conflict data. In this thesis research, a comprehensive framework based on power analysis is first proposed to determine the minimum sample size required for a conflict analysis study. Two case studies are investigated to illustrate how power analysis can be conducted for different types of conflict analysis study specifications, using the corresponding statistical tests. Power analysis is a well-established statistical tool used in many different scientific fields for determining an appropriate sample size. The power analysis exploits the significance criterion (α), power ($1-\beta$), and effect size (ES) such that the sample size is large enough to protect investigators from Type I and Type II errors to conventional thresholds of 95% and 80%, respectively. The minimum sample size is also the optimal sample size because it minimizes the observation period while maintaining acceptable protection from Type I and Type II errors.

A case study is then conducted to assess the safety benefits of three Vision Zero safety countermeasures using data from the City of Toronto. By applying a combination of case-control and cross-sectional studies, the research attempts to quantify the safety effects of three commonly applied Vision Zero countermeasures, namely, Leading Pedestrian Interval (LPI), No Right Turn On Red (NRTOR), and installation of a dedicated Bicycle Lane (BL). The traffic interactions between vehicles and vulnerable road users (VRUs) were extracted using a video data processing platform and two surrogate measures of safety, including post-encroachment time (PET) and conflict speed, were obtained and then used to classify the conflict severity into different levels. A comparative analysis using mixed-effects negative binomial regression was conducted to quantify the impacts of different treatments on the frequency of traffic conflicts under specific road weather and traffic conditions. The results show that these three types of traffic countermeasures can effectively reduce the frequency of high-risk and moderate-risk traffic conflicts, moderated by various, traffic exposure, weather and environmental conditions, and accessible pedestrian signals (APS). These findings could help road safety engineers and decision makers make better informed decisions on their road safety initiatives and projects.

Acknowledgments

I would like to thank my supervisor Dr. Liping Fu for supporting me through this long process.

Dedication

This thesis is dedicated to the millions who die or are injured in traffic-related collisions every year.

Table of Contents

Author's Declaration.....	ii
Statement of Contributions	iii
Abstract.....	v
Acknowledgments.....	vi
Dedication.....	vii
List of Figures	x
List of Tables	xi
Chapter 1 Introduction	1
1.1 Motivation.....	1
1.2 Research Purpose	2
1.3 Thesis Structure	3
Chapter 2 Road Safety Analysis Methods	5
2.1 Road Safety Management and Analysis	5
2.2 Collision-based Road Safety Analysis.....	9
2.3 Conflict-based Road Safety Analysis	12
2.3.1 Motivation for the Traffic Conflict Technique	14
2.3.2 Methodology of the Traffic Conflict Technique.....	15
2.3.3 Validity and Reliability of the TCT	16
2.4 Types of Surrogate Measures of Safety (SMoS).....	19
2.4.1 Time-to-Collision (TTC) Family	20
2.4.2 Post-Encroachment Time (PET) Family	24
2.4.3 Other surrogate measures of safety	27
Chapter 3 Sample Size Requirement for Conflict-based Studies.....	28
3.1 Traffic Conflict Studies: Population and Sample.....	28
3.2 Survey Design for Traffic Conflict Studies	30
3.3 Sample Size Requirements.....	32
3.3.1 Error Propagation.....	33
3.3.2 Sample Size Determination in Past Studies	34
3.4 Minimum Sample Size for Controlling Estimation Errors	35
3.4.1 Methodology and Conceptual Framework.....	35
3.4.2 Minimum Sample Size for Controlling Type II Errors - Power Analysis	37

3.5 A Priori versus Post Hoc Power Analysis	41
3.6 Case Studies: Post Hoc Power Analysis.....	42
3.6.1 Case Study 1: Pedestrian crosswalk safety at non-signalized crossings during nighttime (<i>Fu et al., 2013</i>).....	42
3.6.2 Case Study 2: An automated surrogate safety analysis at protected highway ramps using cross-sectional and before-after video data (<i>St-Aubin et al., 2013</i>).....	45
3.7 Application of Minimum Sample Size Framework: A Priori Power Analysis	47
Chapter 4 Conflict-Based Study for Evaluating the Effectiveness of Vision Zero Safety Countermeasures	52
4.1 City of Toronto Vision Zero Safety Countermeasures	52
4.2 Video Conflict Analysis Scenarios.....	53
4.3 Data Collection and Processing.....	55
4.4 Identification of Traffic Conflicts of Varying Levels of Severity.....	60
4.5 Exploratory Data Analysis of Traffic Conflicts	61
4.6 Safety Effectiveness of Alternative Countermeasures	65
Chapter 5 Conclusions.....	72
5.1 Summary of Main Findings.....	72
5.2 Recommendations and Future Research.....	73
Bibliography	75
Appendix A: Automated Data Collection for Video-based Traffic Conflict Technique.....	82

List of Figures

Figure 2-1: Roadway safety management process from the Highway Safety Manual (AASHTO, 2010)	6
Figure 2-2: Heinrich Triangle, Industrial Accident Prevention	12
Figure 2-3: Safety pyramid and continuum (Hyden, 1996)	14
Figure 2-4: Diagram of Time-To-Collision	20
Figure 2-5: Conflict Severity Diagram (Laureshyn & Varhelyi, 2018)	21
Figure 2-6 Illustration of PET (Laureshyn et al., 2016)	24
Figure 3-1: Stages of Experiment and Survey Design (Wu & Thompson, 2020)	31
Figure 3-2: Error propagation through a traffic conflict study	33
Figure 3-3 Statistical Test Specification and Input flowchart for Common Traffic Conflict Studies	36
Figure 3-4 Potential Errors in Statistical Hypothesis Testing	37
Figure 3-5 Sample size vs. power for different proportions of p_1 at du Fort site for a given p_2 and α	44
Figure 3-6 Sample size vs. power for different proportions of p_1 at Saint Laurent site for a given p_2 and α	44
Figure 3-7 Sample size vs. power for different effect sizes for Type A conflict at a given α	46
Figure 3-8 Sample size vs. power for different effect sizes for Type B conflict at a given α	47
Figure 3-9: Sample size vs. power with different effect sizes for F tests - ANOVA	48
Figure 3-10 Sample size vs. power with different effect sizes for χ^2 test	49
Figure 3-11 Sample size vs. power with different effect sizes for F test	50
Figure 4-1: Map of study sites and camera view at the intersection of Lake Shore Boulevard East and Parliament Street	53
Figure 4-2 Definition of Movements and Conflicts	54
Figure 4-3 Outcomes from the Video Analytics Platform (Transoft Solutions, n.d.)	55
Figure 4-4 Video-based Safety Monitoring (Transoft Solutions, n.d.)	55
Figure 4-5 Classification of Severity Levels for Vehicle and Vulnerable Road User Interaction (Zangenehpour et al., 2016)	60
Figure 4-6 Cumulative Frequency Distribution Curves and Density Histogram of PET Values under Different Traffic Countermeasures	61
Figure 4-7 Box-Violin Plot of Traffic Conflict Frequency with Different Risk Levels under Different Traffic Countermeasures	63
Figure A-1: Mobile video-based traffic conflict analysis system installed in Montreal, Quebec (Jackson et al., 2013)	83

List of Tables

Table 2-1: Summary of TTC Threshold Values	23
Table 2-2: Summary of Post-Encroachment Time Threshold Values	26
Table 2-3: Definitions of less commonly used Surrogate measures of safety	27
Table 3-1 Typical Effect Sizes for Various Statistical Tests (Cohen, 1992)	39
Table 3-2 Use of thermal cameras for conflict analysis - summary of data	43
Table 3-3 Summary Statistics for Site A20-E-E56-3	45
Table 3-4 Effect size, sample size, and observation period (F test - ANOVA).....	49
Table 3-5 Effect size, Sample Size and observation period (Chi-square test).....	50
Table 3-6 Effect size, Sample Size and observation period (F test)	51
Table 4-1 Selected Intersections with their respective Countermeasures and Sample Size	56
Table 4-2 Summary of Variables Descriptive Statistics	58
Table 4-3 Summary Description of the Processed Video Data for Intersections with Implemented Traffic Countermeasures	62
Table 4-4 Results of Kruskal-Wallis Test of Traffic Conflict Frequency with Different Risk Levels under Different Traffic Countermeasures	64
Table 4-5 Results of Traffic Conflict (Classified by PET Value and Conflict Speed) Frequency Modeling.....	68
Table 4-6 Results of Traffic Conflict (Classified by PET Value) Frequency Modeling	70

Chapter 1

Introduction

1.1 Motivation

Traditional road safety analysis predominantly utilizes car crashes as the primary indicator of safety (AASHTO, 2010). This methodology relies on post-hoc crash reports, which are often collected subjectively after incidents have occurred, and may lack the requisite detail for rigorous scientific investigation (Laureshyn & Varhelyi, 2018). Given the infrequency and randomness of collisions, extensive observation periods, often spanning years, are necessary to gather sufficient data for robust analysis.

To address these limitations, researchers are re-evaluating the Traffic Conflict Technique (TCT), a method that has been in existence for several decades. TCT employs near-miss interactions, or conflicts, as surrogate measures of safety (SMoS) to predict crash frequency and severity. The Swedish Traffic Conflict Technique manual defines a conflict as “an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged” (Laureshyn & Varhelyi, 2018). This definition underscores the significance of interactions among road users within shared spaces. Although not every conflict results in a collision, the spatial and temporal proximity of vehicles heightens the risk of such an event, suggesting that drivers may have executed evasive maneuvers to avert a collision.

Conflicts occur more frequently than vehicle collisions, enabling the collection of statistically significant data in a shorter time span compared to collision data. SMoSs offer a distinctive opportunity to assess road safety in scenarios where crash reports are unavailable or when a facility has been operational for a brief period (AASHTO, 2010). The primary motivation behind utilizing SMoSs is to identify potential hazards early, thus allowing road users to take preventive measures to avoid collisions.

While not all encounters are inherently hazardous, some conflicts can be severe if no evasive maneuvers are undertaken. The severity of a conflict is influenced by various factors, including speed, distance, and the type of maneuver required to avoid a collision. Understanding the severity of these conflicts can aid in predicting potential crash frequency and severity.

Many cities and municipalities are adopting Vision Zero or Vision Zero-inspired plans, aiming to reduce and ultimately eliminate fatalities and serious injuries on road transport systems. Traditional road safety analysis, which depends on the occurrence of injurious crashes, is counterproductive to the Vision Zero initiative. Conversely, conflict-based safety analysis, such as the TCT, shows promise in diagnosing and

evaluating safety trends. SMOs can complement traditional road safety analysis by providing additional insights for the road safety management process and in-service road safety review.

The widespread adoption of intersection cameras and advancements in computing power are driving significant changes in the analysis of SMOs. Despite the potential of SMOs, their validity and reliability remain subjects of debate among researchers. Once the concerns raised by skeptical researchers are addressed, SMO-based safety studies may become a viable alternative. These and other limitations will be discussed in subsequent chapters.

1.2 Research Purpose

According to the World Health Organization, over 1.35 million people die globally each year from traffic-related collisions, making it the leading non-disease cause of death worldwide and the foremost killer of individuals aged 5 to 29 years (Breene et al., 2018). Addressing traffic-related injuries and fatalities as a global health crisis necessitates applying the same analytical rigor in traffic safety studies as in health sciences and biostatistics. In clinical trial research, investigators are required to produce pre-study sample size calculations for ethical committees, grant-giving bodies, and funding agencies (Walters, 2009). In contrast, 18% of behavioral observation studies in conflict-based traffic safety did not include sample sizes or provide pre-study sample size calculations (van Haperen, 2016).

To our knowledge, no research has developed a framework for determining the minimum sample size for conflict-based traffic safety analyses. While sample size studies in road safety have been conducted, these primarily focus on collision analysis. The Highway Safety Manual (HSM) recommends calibrating models using a rule-of-thumb sample size of 30 to 50 locations with a minimum of 100 crashes per year. However, some argue that this range is insufficient for achieving desired accuracy (AASHTO, 2010).

Sample size is crucial because it impacts the estimation accuracy and statistical significance of experimental effects. During the study design phase, the required sample size can be determined through sample size calculations or power analysis. Sample size calculations allow investigators to control for estimation error with certainty, whereas power analysis manages statistical errors in hypothesis testing. Power analysis, a statistical tool, is commonly used to determine minimum sample sizes in experimental fields such as behavioral science (Cohen, 1992).

For conflict-based traffic safety studies, sample size calculations and power analysis can determine the minimum required sample size based on the study type and objective. Given that conflict analysis study objectives vary (e.g., before-and-after studies versus collision prediction model development), this thesis proposes a comprehensive framework to aid road safety practitioners in selecting the appropriate method for determining sample size specific to each study purpose.

This thesis will provide an overview of surrogate measures of safety (SMoS), review the current state of SMoS analysis, and offer insights into the advantages and disadvantages of surrogate safety analysis. It will also present a comprehensive framework for determining the minimum sample size required for an SMoS-based safety study, along with a case study evaluating the effectiveness of Vision Zero safety countermeasures using SMoS.

1.3 Thesis Structure

The sections of this thesis are as follows:

Chapter 2: Road Safety Analysis Methods

The literature chosen for this chapter is reference material and peer reviewed papers published in reputable journals. Many of the earlier papers and reference materials conduct the traffic conflict technique with human observers, this research focuses on automated video-based conflict analysis. The automated video-based traffic safety analysis process is described. The SMoS is formally defined based on the literature review. Different types of SMoS are defined and thresholds from existing studies are reviewed. This is done to illustrate the state of the field and to identify any gaps in standardization in definition or threshold values for surrogate safety analysis.

Chapter 3: Sample Size: Background, Methodology, and Framework

This chapter explores the concepts of populations and samples in relation to the TCT, discusses design of experiments, and highlights the importance of sample size calculations before conducting an experiment and the significance of the sample size calculation in SMoS-based safety studies. Important background information on sample size calculations and power analysis is provided and presents the proposed framework for conducting sample size calculations and power analysis for SMoS-based safety studies. Two case studies based on previous research are presented as instructive examples on how to apply the proposed framework.

Chapter 4: Conflict-Based Study for Evaluating the Effectiveness of Vision Zero Safety Countermeasures

This chapter presents a case study is conducted using video trajectory data from ten signalized intersections in the City of Toronto. This chapter presents the application of the proposed sample size determination framework before the data collection, the selection of the traffic safety countermeasures and study sites, and a summary of the data collected.

A combination case-control and cross-sectional study is conducted to quantify the safety effects of three commonly applied Vision Zero countermeasures, namely, Leading Pedestrian Interval (LPI), No Right Turn On Red (NRTOR), and installation of a dedicated Bicycle Lane (BL).

Two SMOs, including post-encroachment time (PET) and conflict speed, were obtained and then used to classify the conflict severity into different levels. A comparative analysis using mixed-effects negative binomial regression was conducted to quantify the impacts of different treatments on the frequency of traffic conflicts under specific road weather and traffic conditions. The results show that these three types of traffic countermeasures can effectively reduce the frequency of high-risk and moderate-risk traffic conflicts, moderated by various, traffic exposure, weather and environmental conditions, and accessible pedestrian signals (APS).

Chapter 5: Conclusion

This chapter presents the key findings and avenues for future research. These findings can help engineers and decision makers make better informed decisions on their road safety initiatives and projects.

Chapter 2

Road Safety Analysis Methods

There exist two general approaches for road safety and the evaluation of traffic safety countermeasures: collision-based analysis and conflict-based analysis. Due to the rare and infrequent nature of collision data, different approaches and techniques are employed to improve the accuracy of the analysis and control for bias inherent in the data. Due to the limitations of collision data, conflicts have become an attractive alternative for road safety analysis, namely for the evaluation of the effectiveness of traffic safety countermeasures. Conflict-based analysis is a promising technique though researchers remain skeptical of its validity and reliability. The different types of conflicts are defined and the thresholds of commonly used SMOs are presented.

2.1 Road Safety Management and Analysis

The most common approach to road safety analysis is based on the observed collision occurrences and the severity of these collisions. This approach has been widely applied by safety researchers and practitioners for quantifying the safety level of different road facilities for the purpose of network screening and the expected effects of various safety countermeasures. However, because collisions are largely random and occur infrequently, it may be impossible to collect sufficient data for analysis in a reasonable amount of time. This means that collision-based analysis could lead to inconclusive and unreliable results due to small sample size and the resulting safety improvement programs or actions could become reactive in nature or fail to address various safety issues in a timely manner.

Figure 2-1 illustrates the six step process used in road safety analysis as prescribed by the Highway Safety Manual (HSM) (AASHTO, 2010). The HSM outlines a systematic approach to improving road safety, which includes the following steps in network screening and diagnosis to identify and address problems in the road transport system.

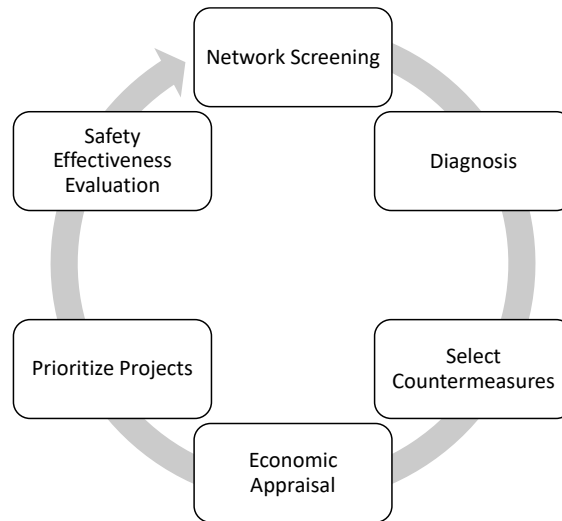


Figure 2-1: Roadway safety management process from the Highway Safety Manual (AASHTO, 2010)

Network Screening Phase

1. Reviewing and Analyzing the Network:

- Purpose: To identify locations with the highest potential for safety improvements.
- Process: The transportation network is systematically reviewed, assessing various road sections, intersections, and segments. This involves collecting and analyzing crash data and other relevant factors.
- Outcome: Sites are ranked based on their safety performance, prioritizing those with the most urgent needs for intervention.

Diagnosis Phase

2. Evaluating Data:

- Purpose: To understand the underlying causes of collisions at identified sites.
- Process: Detailed evaluation of data from the potential sites, including collision data, historical site data, and field conditions.
- Outcome: Identification of collision patterns and contributing factors, such as road design flaws or hazardous conditions.

3. Identifying Safety Problems:

- Purpose: To pinpoint specific issues that need to be addressed to improve safety.
- Process: Analysis of collision types, frequencies, and severities to identify recurring problems.

- Outcome: A clear understanding of the safety issues at each site, forming the basis for selecting countermeasures.

Safety Countermeasure Selection and Economic Appraisal

4. Selecting Countermeasures:

- Purpose: To develop strategies that will effectively reduce collisions.
- Process: Based on the diagnosis, potential safety countermeasures are identified. These may include engineering changes, enforcement strategies, or educational campaigns.
- Outcome: A list of proposed safety measures tailored to address the identified problems.

5. Economic Appraisal:

- Purpose: To ensure the cost-effectiveness of the proposed safety measures.
- Process: Assessing the economic benefits and costs of each countermeasure, considering factors such as implementation costs, expected reduction in collisions, and potential savings from reduced collisions.
- Outcome: Selection of economically justifiable projects that offer the best value for money.

Project Prioritization and Implementation

6. Project Prioritization:

- Purpose: To allocate resources efficiently and effectively.
- Process: Prioritizing safety projects based on their potential impact, cost, and alignment with stakeholder objectives (e.g., improving safety, enhancing mobility, minimizing environmental impacts).
- Outcome: A prioritized list of safety projects ready for implementation.

7. Implementation:

- Purpose: To execute the selected safety projects.
- Process: Developing detailed plans and timelines for implementing the prioritized safety measures, followed by their actual execution.
- Outcome: Improved road safety through the successful implementation of targeted interventions.

Safety Effectiveness Evaluation

8. Collecting Impact Data:

- Purpose: To measure the success of the implemented projects.

- Process: Gathering data on collision frequency and severity after the implementation of safety measures.
- Outcome: Quantitative data indicating the impact of the projects on road safety.

9. Evaluating Effectiveness:

- Purpose: To determine if the safety objectives have been met.
- Process: Analyzing the collected data to assess changes in collision frequency and severity.
- Outcome: A comprehensive evaluation of the effectiveness of the implemented safety measures, informing future safety planning and interventions.

The HSM defines road safety primarily through the lens of collision history, using a model called the Safety Performance Function (SPF). The SPF relates the expected number of crashes at a site to various influencing factors, such as traffic volume, road characteristics, and environmental conditions.

To improve the accuracy of safety performance predictions, the HSM employs the Empirical Bayes (EB) approach. This method combines historical crash data with SPFs to provide more reliable estimates of expected crash frequencies. The EB approach corrects for regression-to-the-mean bias and improves the identification of sites that truly have higher-than-expected crash rates, making it a powerful tool for safety analysis and planning.

By integrating SMoS into the HSM's framework, transportation agencies can enhance traditional safety analysis methods:

- Using Near Miss Data: SMoS data on near-miss incidents and conflicts can provide early indicators of potential safety issues.
- Complementing Collision Data: SMoS data can complement traditional collision data, offering a more comprehensive understanding of safety trends and helping identify potential hazards before actual collisions occur.

This integrated approach allows for a more proactive and effective strategy in improving road safety, aligning with the goals of Vision Zero and other safety initiatives aimed at eliminating fatalities and serious injuries on road transport systems.

There are several limitations inherent to collision data and introduce bias and affects its reliability when used to determine the effectiveness of a safety countermeasure or project. These limitations of collision data are as follows:

- Natural variability in collision frequency

- Regression-to-the-mean and regression-to-the-mean bias
- Variations in roadway characteristics and roadway geometry
- Conflict between collision frequency variability and changing site conditions

To address the inherent limitations of collision data, several statistical methods have been developed and can reliably estimate the expected average collision frequency on roadways: existing and future prior to its construction and use. The following statistical methods are used to estimate the expected collision frequency:

- Empirical Bayes method (EB method)
- Hierarchical Bayes method
- Full Bayes method

The roadway safety management process, as outlined in the HSM, uses the collision as a fundamental indicator of safety and is a reactive process. That means that the safety of road facilities can only be assessed after crashes occur and that safety issues at road facilities are treated when enough crashes constitute a trend. This raises the ethical issue of road safety. Road safety professionals want to prevent crashes from happening, but they must rely on crashes occurring to predict any future risk. These ethical issues and the reliance on inherently biased data is why the TCT can be a useful complement to the road safety management process. It can complement accident histories so that better decisions can be made for bigger impacts in the transportation system.

2.2 Collision-based Road Safety Analysis

Collision frequency can be used in different types of road safety studies and analysis. The most rudimentary method of determining the safety of a road entity is by considering the collision history. If a location experiences collisions more frequently than other locations, it is believed to be more dangerous than a neighboring location with fewer collisions. This does not account for traffic volumes, which is the largest contributing factor to collisions. There exist more opportunities for the vehicles to interact and potentially collide if there are more vehicles on the road. This can be managed by determining collision frequency on specific road entities and comparing the collision with other similar road entities. Due to the rare and random nature of collisions, road entities need to be observed for many years to determine the safety. Fluctuations and variability in year-to-year collisions require more years of data to determine the expected average collision frequency. Instead, the expected average collision frequency can be estimated using regression models. Safety performance functions (SPFs) are the regression models used to determine the

relationship between collision frequency and some predictors. These predictors can include traffic volume, weather, time of day, or speed of vehicles. Typically, the functional form of SPFs is based on the negative binomial distribution, which is well suited for modelling the variability of collision data. These models address regression-to-mean bias and can reliably estimate expected average collision frequency for existing roadway conditions but changes to existing conditions or new roadway designs prior to construction and use. The reliability of these models is based on how well the model fits the original data, the calibration of these models to local data, and the incorporation of historic collision for specific sites.

Safety effectiveness determines the impact of the safety countermeasure. There are three basic study designs that are used for safety effectiveness evaluation:

- Observational before /after studies
- Observational cross-sectional studies
- Experimental before /after studies

Observational studies can be conducted to determine or validate the efficacy of a safety countermeasure, in a time and resource efficient manner (Hauer, 1997). There are two types of observational studies: before/after and cross-sectional.

Observational before/after studies are commonly used in safety effectiveness evaluation. In collision-based road safety analysis, observational before /after studies use collision and traffic volume data for the time periods before and after the implementation of a treatment.

Before /after studies are often desirable but not feasible because of limitations in scheduling and timing. Instead, an observational cross-sectional study can be conducted. An observational cross-sectional study compares sites of similar characteristics such as geometry, land use and traffic volumes. Some of these sites will have the treatment or safety countermeasure installed and some will not. The safety effectiveness of a treatment can be determined through statistical modeling techniques. The data available such as collision frequency, traffic volume, and site characteristics are analyzed in a single model including an indicator variable to denote the presence or absence of the treatment. Then a model is developed to simulate the collision or conflict frequency of these sites; the difference in conflict frequency is then attributed to the treatment and is indicative of safety effects. In the context of the HSM, the model then simulates the site based on the variables through the following steps:

- 1) Determine safety countermeasures to be examined.
- 2) Determine road entities to observe and study.
- 3) Determine the type of statistical testing required and the minimum number of observations required.

- 4) Collect data on independent variables such as traffic volumes, weather, and site characteristics for the group of study sites (with and without treatment)
- 5) Collect collision data in the form of SMOs for selected sites.
- 6) Conduct analysis on collision data and independent variables.

The following data needs and inputs are recommended by the HSM:

- 10 to 20 treatment sites to evaluate a safety treatment
- 10 to 20 nontreatment sites for the nontreatment group
- 3 to 5 years of crash data for both the treatment and nontreatment sites

The data obtained for a cross-sectional study should be obtained for the same period of time for both the treatment and comparison sites.

Observational cross-sectional studies are useful when collecting data before the implementation of a treatment is not possible. However, it is difficult to assess the cause and effect or explain factors affecting the safety effectiveness.

Observational studies use observations to make inferences about changes in safety from the treatments. These treatments are implemented by highway agencies or municipalities as part of safety improvement projects and are not implemented to specifically observe the safety impacts of the treatments.

Experimental studies are designed specifically to observe the safety impacts of the safety treatments.

To conduct experimental studies, groups of sites with similar characteristics are selected. From this group of sites, the treatment is implemented at some sites and the rest are left untreated as a comparison group. The difference in crash frequency between the treatment and the comparison group is directly attributed to the efficacy of the safety countermeasure. Often, due to time and budget constraints, it is suggested to conduct an observational study rather than an experimental study. Experimental studies require the implementation of the same safety countermeasure at multiple sites and many years to collect sufficient collision data. Furthermore, confounding variables may obscure the impacts of the safety countermeasure due to changes to highway legislation, changes in travel demand or preferences, and road geometry. This process may be too time and resource intensive to determine efficacy of one potential safety countermeasure.

After estimating the treatment effectiveness as outlined in the previous sections, it is important to consider the precision of the treatment effectiveness. This ensures that when comparing the treatment sites to nontreatment sites, the differences observed between the two study groups is statistically significance and the results are meaningful. A hypothesis test, such as a t-test, can be used to assess the statistical significance of the estimate safety effectiveness.

2.3 Conflict-based Road Safety Analysis

The idea of SMOs originated from industrial accident prevention concepts. The Heinrich triangle, as seen in Figure 2-2, was developed in the 1950s. (AASHTO, 2010) The Heinrich Triangle is based on two concepts:

- 1) Events at the base of the triangle are proximate and precede the events close to the top.
- 2) Events at the base of the triangle occur more frequently and can be more reliably estimated.

Since the less severe events occur more frequently than the higher severity events (events at the top of the triangle), the less severe events can be more reliably estimated. Furthermore, it was hypothesized that the reduction in frequency of low severity incidents would result in the reduction in frequency of the more severe incidents near the top of the triangle. As a result, observations of highly frequent but low risk events could be used as a surrogate for less frequent but more severe events.



Figure 2-2: Heinrich Triangle, Industrial Accident Prevention

Observing conflicts as a surrogate for vehicle collisions for the purpose of evaluating road safety was first considered in the 1950's and was first practically applied in the 1960's by researchers at General Motors (Laureshyn & Varhelyi, 2018). Teams of trained human observers were sent to intersections to observe traffic events involving evasive maneuvers. These included changing trajectories, the use of brake lights, and the speeds of vehicles.

Initially, this method garnered interest from road safety researchers. The most appealing aspect of using surrogate measures of safety (SMoSs) is that conflict data could be gathered in much shorter periods of time compared to a crash and analysis would be less affected by time-dependent factors such as changes in

driving behavior, rules of the road, or geometry of the road facility. It became evident that this method was too laborious and time intensive to be a feasible road safety technique and subsequently, it fell out of favour. The Swedish Traffic Conflict Technique (TCT) was first developed in the 1970s at Lund University. In 1977, Amundsen and Hyden formally defined a traffic conflict. It is as follows (Amundsen & Hydén, 1977):

“A traffic conflict is an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged.”

Figure 2-3 illustrates the adapted Heinrich Triangle which was developed specifically for a road safety context. The Swedish TCT follows three distinct but basic concepts:

1. There must be a collision course in a conflict.
2. Conflict severity is defined on the onset of an evasive action.
3. There must be a distinction between severe and non-severe conflicts.

Severe conflicts can be broken down into more frequently observable less severe conflicts. By employing safety countermeasures to reduce the frequency of less severe conflicts with safety countermeasures, it is believed that the frequency of more severe interactions will be reduced as well.

It is often difficult to assess injury severity and the consequence of an avoided collision using conflicts and surrogate measures of safety. Different techniques ignore the potential consequences or use arbitrary rules to incorporate consequences into the final risk score. Instead, the Swedish TCT includes speed in the final grading of conflict severity. It is well accepted in the field of road safety that the higher the speed of the potential collision, the higher the risk of injury or death, which is discussed in Section 4.4.

The Swedish TCT is the oldest and most well-developed surrogate safety method, but it still faces many limitations, which will be discussed in Section 2.3.2. Of the other TCT developed, it is unique in that it is based on several large-scale validation studies.

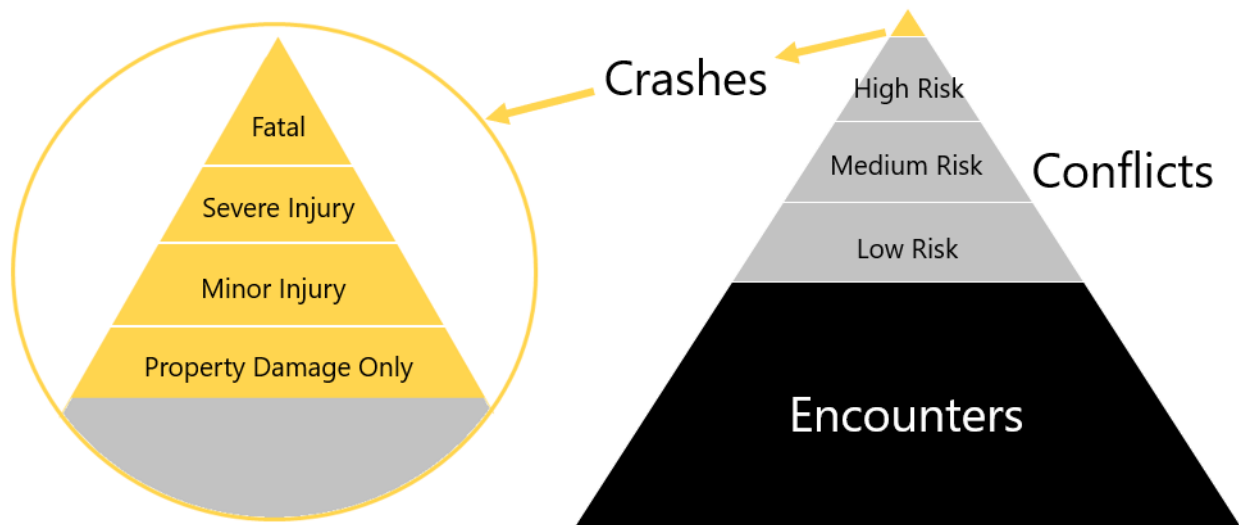


Figure 2-3: Safety pyramid and continuum (Hyden, 1996)

The concept is very similar to that of the Heinrich triangle. The severity of the encounter increases closer to the top of the pyramid. Severity is defined as the proximity to a collision and the potential consequences of the crash should it take place (Laureshyn & Varhelyi, 2018). The theory behind the safety pyramid and the traffic conflict technique is that a traffic process can be broken down into discrete and basic encounters.

The assumption in the safety pyramid is that there exists some relationship between the severity and frequency of events; less severe conflicts could lead to more severe crashes.

The lower portion of the pyramid represents the normal interactions that are safe and happen all the time. At the very top of the pyramid are the most injurious and dangerous events: the fatal and injury-causing crashes. These crashes are infrequent when compared to the total number of encounters represented at the bottom of the pyramid.

2.3.1 Motivation for the Traffic Conflict Technique

The goals of the TCT are:

- To diagnosis a site's safety.
- To identify contributing factors to a site's safety.
- To compare the effects and impacts of different road infrastructure, policy, rules, and safety countermeasures.
- To conduct before and after studies without having to wait for years.
- To monitor safety trends at different locations.

The most attractive aspect of using the TCT, as opposed to collision-based road safety analysis, is that it is a proactive process. Rather than collecting collision data over the span of many years, conflicts occur much more frequently and can help to identify high-risk areas before a collision occurs. If the logic of the safety pyramid holds true, these conflict events will precede and be proximate to any major and severe collisions. Also, by reducing the number of these conflicts, the number of collisions should be reduced as well. Analyzing conflicts also solves the ethical issue of traditional road safety analysis: having to observe collisions before safety issues can be addressed. These collisions may cause harm to road users, whether they are fatal, injury-causing or property damage only.

Because of the short amount of time required for the TCT to record statistically significant amounts of data, safety analysis can be conducted faster and can control for time-dependent factors. This includes changes in behavior due to policy and regulation changes, changes in technological trends, and exposure.

These benefits of the TCT are dependent of the validity and reliability of the technique, which is explored in Section 2.3.3.

2.3.2 Methodology of the Traffic Conflict Technique

The Swedish TCT offers two distinct methods for conducting observations: manual and video recording with analysis tools.

Manual Traffic Conflict Observation

Process:

- Trained human observers conduct manual observations over multiple days, typically for several hours at a time.
- Observers record conflicts, noting evasive actions taken, estimating speed and distances, sketching the conflict scene, and documenting other relevant details for later analysis.

Advantages:

- Requires minimal equipment (pen, recording sheet, watch).
- Flexibility to observe any intersection.
- Allows for proactive data collection.

Disadvantages:

- Data collection can be skewed.
- Observations typically occur in optimal weather conditions.
- Potential for observer effect, where road user behavior may change upon noticing observers.

Video Recording and Analysis Tools

Process:

- Uses video cameras to record traffic interactions.
- Analysis can be conducted in real-time or through review of recorded video clips post-observation.

Advantages:

- Provides more accurate data.
- Reduces observer effect.
- Allows for comprehensive analysis of conflicts.

Disadvantages:

- Limited perspective based on camera angle.
- May not capture all nuances of traffic behavior.

In summary, while manual observation offers flexibility and minimal equipment requirements, it may introduce biases and observer effects. Video recording, on the other hand, provides more accurate data and reduces observer impact but is constrained by camera angles and potential limitations in capturing real-time context. Both methods contribute to proactive road safety analysis, emphasizing the distinction between collisions and conflicts to enhance understanding and prevention strategies. A full description of how the data collection for video-based traffic conflict techniques can be automated and conducted.

2.3.3 Validity and Reliability of the TCT

Before camera technology was widely applied in road safety study, researchers were concerned about the reliability of conflict data collection methods. Limited to only using human observers, the traffic conflict technique faced two main challenges (Chin & Quek, 1997):

- Consistency: the variation in conflict data collection caused by the individual observer.
- Repeatability: the variability in interpretation and recording of the same situation between individual observers.

It has been proven that with proper training, it is possible for observers to detect conflicts and determine speeds and distances with consistency (Hyden, 1996). However, it is still difficult to maintain consistency throughout the observation period. Observation periods are typically eight hours long, where observers need to diligently assess the intersection for conflicts for one to two hours at a time before they could take a break. It is difficult for observers to maintain a constant level of alertness and perception throughout the observation period; skills deteriorate as the observer gets tired regardless of the level of training. By using human observers to collect data, some level of subjectivity is introduced into the data.

For the Swedish TCT to be reliable, the method of conflict detection and severity scoring should account for differences in their observer's mental acuity and perception, the weather, and the lighting. Video-based traffic analysis tools and software show promise in solving these problems.

Issues arise when choosing a surrogate safety measure to use. Very few validation studies have been conducted to relate observed conflicts with actual accidents at the same site.

According to Chin et al. (1997), the validity of a TCT is judged by the correlation between observed conflicts and the collision history. Many conflict studies have reported challenges in demonstrating an adequate correlation between conflicts and collisions, due to changes in data collection technology and methodologies and long-term variability of traffic conditions, road infrastructure, and driver behaviour.

For example, Sayed et al. (1994) found the strongest correlations between collisions and conflicts (where the threshold for a critical event is $t_c = 1.5s$) in intersections with R^2 values ranging from 0.64 to 0.81.

Sayed et al. (1999) created models to aggregate conflict data into the number of average hourly vehicle conflicts. The conflict data included: observed traffic conflicts, land use type (rural, urban, suburban), and traffic volumes. Based on the average hourly vehicle conflicts (where the threshold for a critical conflict is 1.5s), linear regression analysis was conducted to model the relationship between collisions and conflicts. The best correlation between collisions and conflicts was for signalized intersections with R^2 at a 95% confidence interval ranging from 0.70 to 0.77. Correlation was worse at unsignalized intersections. (Sayed & Zein, 1999).

Gettman et al. (2008) developed the following model to relate conflicts (where the threshold for a critical conflict is 1.5s) and collisions:

$$Crashes = 0.119Conflicts^{1.419} \quad (1)$$

The R^2 value for this model is 0.41 (Gettman et al., 2008). With more data and longer period of observations, researchers hope to develop models that will show a clear relationship between conflicts and collisions.

It has been claimed that these issues of validity arise from inaccurate and unreliable conflict data collection and underreporting of collisions (Hakkert et al., 1987). Others questioned the usefulness of the technique and called for a reassessment of the entire concept. Hauer et al, 1986, proposed a different way of approaching the validity of the traffic conflict technique by proposing that the operational definition of validity should be (Hauer & Garder, 1986):

“a technique for the estimation of safety is ‘valid’ if it produces unbiased estimates of the variance, of which is deemed to be ‘satisfactory’”.

The technique is deemed “satisfactory” depending on the variance of its estimates and on the cost of estimate (e.g., labor and time). Therefore, the validity of the safety technique should be based on the variance of the safety estimates which it produces rather than the model’s ability to predict collisions based on conflict data.

Hauer and Garder (1986) defines safety as “the expected number of accidents (by severity) occurring on the entity per unit of time”. The problem with trying to predict the number of accidents, regardless of using conflicts or not, is that accidents are rare and random events.

Rather than trying to establish a statistical relationship between conflicts and collisions, the validity of the TCT is best judged by comparing the variance of the estimates in rates in conflicts and collisions rather than comparing the estimates themselves. The most valid method would not be one that can accurately predict accidents. Rather, the most valid method according to Hauer and Garder, (1986) would be the one that produces unbiased estimates with the smallest variance and that the relationship between the conflict and the collision should be as follows:

$$\lambda = c \times \pi$$

$$\left[\begin{array}{c} \text{Number of accidents expected} \\ \text{to occur on a road entity during} \\ \text{a certain period} \end{array} \right] = \left[\begin{array}{c} \text{Number of conflicts} \\ \text{occurring on a} \\ \text{road entity in that time} \end{array} \right] \times \left[\begin{array}{c} \text{Collision – to – conflict} \\ \text{ratio for that road entity} \end{array} \right] \quad (2)$$

where λ is the number of accidents expected to occur, π is the number of conflicts occurring, and c is the collision-to-conflict ratio.

The variance of π would allow for the determination of the accuracy with which safety can be measured through conflicts. The validity of the conflict method is based on the difference between the actual value of π and the expected value of π ($\hat{\pi}$). For c to be a reliable and stable property of the conflict event, c must be consistent and nearly the same on similar entities given a sufficiently large data set. Furthermore, by ensuring that $\pi - \hat{\pi}$ is sufficiently small, it ensures a small variance of λ . As a result, the variation in the expected number of conflicts and collisions do not vary much from entity to entity given a large enough data set and similar entity types.

This would validate the TCT. At the time which the Hauer and Garder (1986) paper was written, it is important to note that computer vision, machine learning, and ubiquitous use of high-quality cameras at intersections were non-existent. Concerns about the reliability and validity of the TCT stemmed from the onerous and time intensive TCT methods that required human observers. With improvements in computing

power and a push for data-driven decision making, large conflict datasets can be collected to come to conclusive results of its validity.

2.4 Types of Surrogate Measures of Safety (SMoS)

The American Association of State Highway and Transportation Officials (AASHTO) released the Highway Safety Manual (HSM) in 2010. The HSM describes SMoS as events that are closest in relation to, and usually precede, actual collisions. They are the quantification of conflicts and near miss encounters described in the TCT. SMoS presume the existence of a causal relationship between the conflict and expected crash frequency. In a transportation safety context, it is important that SMoS meet two requirements (Tarko et al., 2009):

- SMoS should be based on an observable non-crash interaction that is causally related to collisions in predictably and reliably.
- There exists a practical method for converting SMoS to collision frequencies and severities.

There are many different types and variants of SMoS; variants of specific types can be grouped into “families”. According to the HSM, the following can be used in conflict-based safety analysis (AASHTO, 2010):

- Speed
- Encroachment time
- Gap time
- Deceleration rate
- Proportion of stopping distance
- Post-encroachment time (PET)
- Initially attempted post-encroachment time
- Time to collision (TTC)

Many other types of SMoS exist, including any measures that can be used to estimate the average crash frequency and severity. Examples include speed, number of intoxicated drivers, exposure (traffic volume), and proportion of belted occupants. Although these all contribute to crash occurrence and/or severity, these measures are difficult to quantify with human observers or cameras in the field. Thus, these measures will not be discussed.

In literature, the most used SMOs are variants of time to collision (TTC) and post-encroachment time (PET). Because of this, TTC and PET will be explored in detail. TTC and PET are very similar. Both are measured in units of time, typically seconds, and represent the margin of safety that road users had before a collision. The lower the TTC/PET, the less time available for road users to correct the error, and the greater the safety risk. The difference is that TTC considers what could happen whereas PET considers what does happen.

2.4.1 Time-to-Collision (TTC) Family

Time-to-collision (TTC) is the predicted time remaining for two road users (where road user 1 is the leading vehicle and road user 2 is the following vehicle) who are on a collision course to collide had an evasive action not been taken by either of them, as illustrated in Figure 2-4. It describes the time remaining for the road user to successfully perform an evasive action.

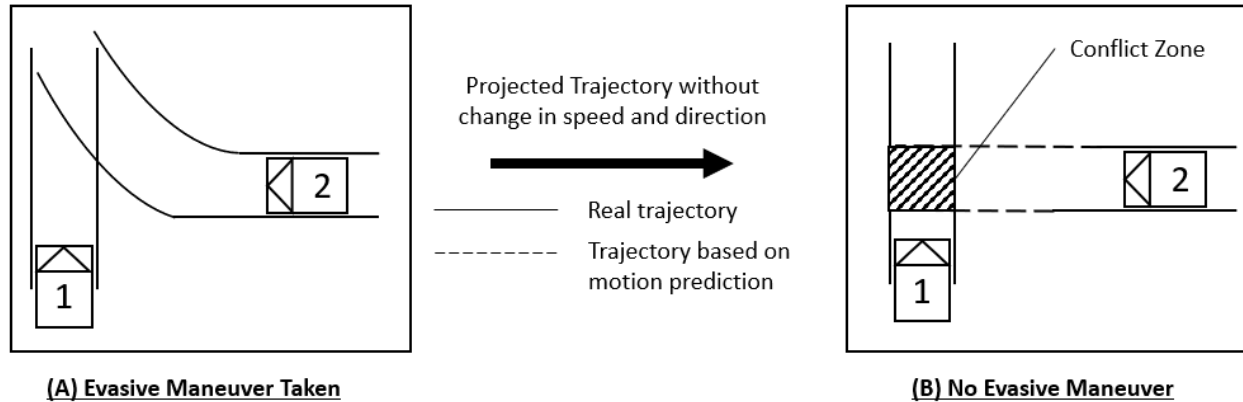


Figure 2-4: Diagram of Time-To-Collision

To determine TTC, it is necessary to first define a collision course. A collision course consists of road users' paths, where unless road users take an evasive action, a collision will occur. The criticality of a conflict is defined at the moment one of the road users take an evasive action (Laureshyn & Varhelyi, 2018).

Rear-end collision occurs when the speed of the following vehicle is greater than the lead vehicle and no evasive maneuver is taken. The greater the difference in speed or the smaller the physical gap, the shorter the TTC. Therefore, TTC can quantify the severity of the conflict based on the magnitude. As TTC decreases, the severity increases. It is important to note that TTC is a hypothetical measure; it is not possible to observe a TTC. Rather, it is calculated based on the future motion prediction (St-Aubin et al., 2013). It would be possible to find a TTC for every pair of road users.

Automated video-based conflict software has been developed to track the trajectory of every road user and future trajectories based on current motion patterns (St-Aubin et al., 2015b). Currently, the models used to predict trajectories for TTC are being improved and further developed. These models assume constant velocity and trajectory, which is unrealistic because drivers would make an evasive maneuver. It has been suggested that through machine learning, models of natural motion can be built and applied to reduce the error on assumptions of these potential interactions (St-Aubin et al., 2015a).

As discussed in the previous section, the Swedish TCT incorporates speed into the use of TTC as it is highly associated with the expected severity of the collision had a collision result from the conflict (Laureshyn & Varhelyi, 2018). Figure 2-5 illustrates how speed, on the y-axis, would change the criticality of conflict depending on the TTC, on the x-axis. The red line shows the threshold of speeds vs. TTC for interactions to be considered severe and non-severe.

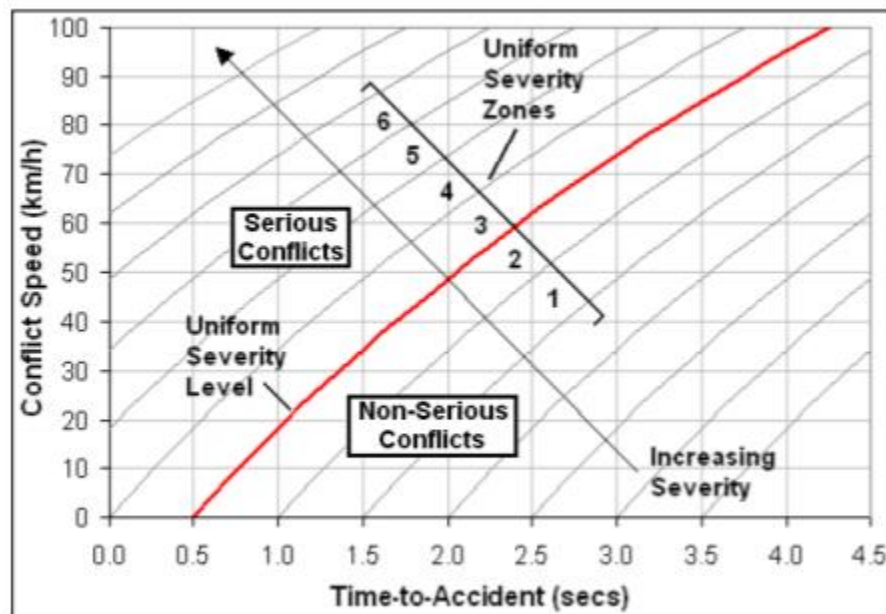


Figure 2-5: Conflict Severity Diagram (Laureshyn & Varhelyi, 2018)

Variants of TTC have been proposed (Hauer, 1996), like the reciprocal of TTC, the minimum TTC, and unique centile of TTC have been used, but the thresholds are only applicable to TTC.

2.4.1.1 Threshold values for TTC

Based on the review of literature, the threshold values for TTC are summarized in Table 2-1, which also include information about the characteristics of the studies, such as the number of sites and their respective

locations. The purpose of using a threshold for Time-to-Collision (TTC) in traffic conflict analysis is to distinguish between different levels of risk associated with conflicts between vehicles. These TTC values represent the ad hoc threshold between severe and non-severe conflicts. Thresholds can be used to identify critical conflicts (e.g.: interactions between vehicles that are at higher risk of collision). The justification of certain threshold values vary between studies: some are based on the perception time of 1.5 s for drivers (St-Aubin et al., 2015a) or thresholds established in previous studies (Sayed & Zein, 1999). One of the earliest justifications of a critical threshold of 1.5 s is based on watching video-taped conflicts and finding that road users would consciously take evasive maneuvers as late as possible based on their opinion of safe behavior (Amundsen & Hydén, 1977). This table is not exhaustive but shows that in the current state of the method, threshold TTC ranges from 1.5 to 5 seconds depending on speed, road entity type, and surrounding land use. Some of the studies and results are described in further detail below, these studies are explained further because of novel approaches for applying the TCT, validating the results from using the TCT and providing insight into why certain thresholds are chosen for TTC.

Table 2-1: Summary of TTC Threshold Values

<u>Study</u>	<u>Number of Sites</u>	<u>Location</u>	<u>TTC Threshold</u>	<u>Threshold Justification</u>
Amundsen & Hydén, 1977	115 intersections	Stockholm and Malmo, Sweden	1.5 seconds	Observation of conflicts
Lord, 1996	8 signalized intersections	Ontario, Canada	1.5 seconds	Based on literature (Amundsen & Hydén, 1977))
Autey et al., 2012	3 channelized right turn lanes	Penticton, British Columbia	3 seconds	Based on literature (Sayed & Zein, 1999) and (Sayed et al., 1994)
Sayed et al., 2013	Burrard Bridge Ramps	Vancouver, British Columbia	3 seconds	Based on literature (Amundsen & Hydén, 1977)
Souleyrette et al., 2012	High speed rural expressway J-turn intersection, simulated in VisSim 25 times with 25 random seeds	Floyd, Iowa	1.5 seconds for urban low speed (25 to 30 mph) signalized intersections 5 seconds for rural areas	Based on literature (Hyden, 1996)
St-Aubin et al., 2015a	40 roundabout weaving zones	Quebec, Canada	1.5 seconds	Based on literature (Hyden, 1996)
Ismail et al., 2010	Pedestrian scramble	Oakland, California	5 seconds	Based on the minimum TTC observed

2.4.2 Post-Encroachment Time (PET) Family

The second most used SMoS in literature is post-encroachment time (PET). Figure 2-6 illustrates the definition of PET. PET is the amount of time required for the trailing vehicle, vehicle 2 in Figure 2-6, to pass the same point that the leading vehicle, vehicle 1, had passed earlier. (Laureshyn & Varhelyi, 2018) This is the quantification of a near miss, the small gap time between the trajectories between two road users who were on a collision course.

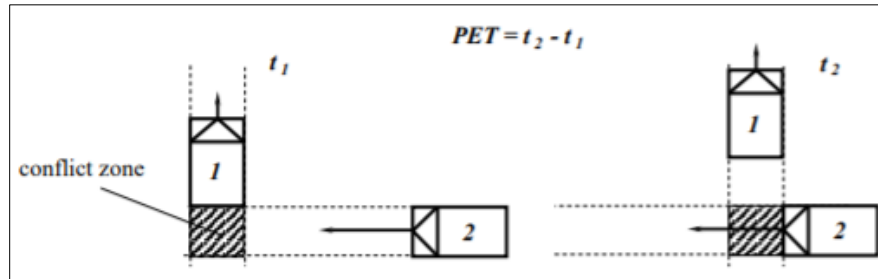


Figure 2-6 Illustration of PET (Laureshyn et al., 2016)

PET is widely favored in safety analysis for several reasons compared to other measures like Time-to-Collision (TTC):

- **Direct Measure of Near Misses:** Unlike TTC, which focuses on potential collision times, PET directly measures the time gap between vehicles after narrowly avoiding a collision, providing clearer insight into near-collision proximity.
- **Reflects Evasive Actions:** PET inherently reflects the effectiveness of evasive maneuvers taken by drivers to avoid collisions. It offers a practical indicator of how well drivers navigate potential hazards in real-world scenarios.
- **Broader Applicability:** PET can be applied across various traffic situations and environments, making it versatile for assessing safety across different road types and conditions.

By quantifying near-miss events with PET, researchers and transportation agencies gain valuable insights into potential collision risks and the efficacy of safety measures. This makes PET a valuable tool for improving road safety strategies and understanding driver behavior in critical situations.

There are several variations of post-encroachment time (PET) used in traffic conflict analysis. Gap time and encroachment time are closely related concepts to PET, focusing on the temporal dynamics of vehicle interactions in conflict zones.

Gap time measures the duration between the rear bumper of one vehicle passing a reference point and the front bumper of the following vehicle passing the same point. It quantifies the temporal spacing between vehicles entering a conflict zone or intersection. Similarly, encroachment time measures the duration between vehicles' bumpers entering a conflict zone or intersection. It specifically indicates the overlap or temporal proximity of vehicles in the same space, highlighting potential conflict situations where vehicles encroach upon each other's path.

The main difference between is that gap time focuses on the temporal spacing between two vehicles whereas encroachment time focuses on the overlap between two vehicles in a conflict zone.

Another variant of PET is negative PET (nPET), also known as pre-encroachment time. Ismail et al. (2010) consider nPET particularly in contexts involving interactions between vehicles and pedestrians. In vehicle-only scenarios, the order of entry into the conflict zone is crucial, whereas in pedestrian-vehicle interactions, temporal proximity between the pedestrian and vehicle is more critical for assessing severity. Predicted PET (pPET) is another variation used in intersections, akin to Time-to-Collision (TTC), incorporating motion prediction alongside PET measurements (St-Aubin et al., 2013). pPET is especially valuable for analyzing interactions involving sharp angles, where conventional measures may not fully capture the dynamics of potential conflicts.

These variations of PET provide nuanced insights into vehicle interactions, helping to identify and mitigate collision risks in diverse traffic scenarios.

2.4.2.1 Threshold values for PET

Through the literature review, the following values of PET are determined to be typical threshold values, as summarized in Table 2-2. PET threshold values range from 0 to 9.95 seconds. While many of these studies suggested their threshold values based on the minimum PET observed in the field, a study bases the threshold values on AASHTO values for gap times (Souleyrette & Hochstein, 2012). Similarly, to Section 2.4.1.1, some studies are summarized below because of their approach and application of PET in road safety analysis.

Table 2-2: Summary of Post-Encroachment Time Threshold Values

<u>Study</u>	<u>Number of Sites</u>	<u>Location</u>	<u>PET Threshold</u>	<u>Threshold Justification</u>
Alhajyaseen, 2014	5 signalized intersections	Nagoya City, Japan	5 seconds	Assume vehicle speed is 50 km/hr. In 5 seconds, a vehicle can travel approximately 70 m. Therefore, two consecutive vehicles are not likely to existing within an intersection within 5 seconds.
Songchitruksa & Tarko, 2006	16 sites	Lafayette, Indiana	6.5 seconds	Provided the best fitted models for the site-aggregated counts, given short observation period.
Zangenehpour, 2015	23 intersections with cycle tracks	Montreal, Quebec	Discretized into 4 severity levels: $t \leq 1.5$ seconds $1.5 < t \leq 3$ seconds $3 < t \leq 5$ seconds $t > 5$ seconds	$PET \leq 1.5$ seconds is chosen as the threshold for very dangerous interactions based on literature review (Sayed et al. 2013c).
Kassim et al., 2014	2 signalized intersections with over 384 recorded conflict events	Ottawa, Ontario	$0 < t \leq 3$ seconds	Selected based on literature review, not intended to suggest a critical PET threshold.
Souleyrette & Hochstein, 2012	High speed rural expressway J-turn intersection, simulated in	Floyd, Iowa	5.0 seconds and A maximum threshold value of 9.95 seconds	Paper states that the reason for the 5.0 second threshold was unclear; 9.95 seconds maximum

	VisSim 25 times with 25 random seeds			threshold is based on maximum input in VisSIM
AASHTO Green Book (American Association of State Highway and Transportation Officials, 2011)			$5.5 < t \leq 12.0$ seconds	Depending on intersection traffic control, minor road design vehicle, and the desired maneuver of the minor road vehicle.

2.4.3 Other surrogate measures of safety

Many other SMOs have been developed and used in past research. There are many more SMOs that are currently being investigated, such as angle of collision and the difference in speeds between conflicting road users. Some examples of other SMOs and what they represent are summarized in Table 2-3.

Table 2-3: Definitions of less commonly used Surrogate measures of safety

<u>Surrogate Safety Measure</u>	<u>Definition</u>
Deceleration-to-Safety Time (DST) (Allen et al., 1977)	Deceleration is required to reach non-negative PET if road users continue the collision course. This measure is particularly useful for left turn movements.
Conflicting Speed (CS) (Laureshyn & Varhelyi, 2018)	The speed of the road user that makes the evasive maneuver when the evasive maneuver is taken. The higher the CS, the more severe the conflict and thus the more severe the potential collision.
Congestion (Laureshyn et al., 2016)	Because of high density, congestion can cause conflicts in urban intersections.
Intersection Conflict Index (Sayed & Zein, 1999)	An aggregation of ratios based on average hourly conflict ratios and hourly entering volumes. This concept is used to summarize the level of service and quantify the conflict risk at the intersection based on exposure.

Chapter 3

Sample Size Requirement for Conflict-based Studies

This chapter defines the population and sample of traffic conflict studies, discusses the significance of errors in these studies and explores the significance and requirements of sample size calculations for conflict-based safety studies. A proposed framework for determining the minimum sample size for conflict-based safety studies is presented, which is applicable to multiple types of studies. Three case studies are presented as illustrative examples of how to determine the minimum sample size for conflict-based safety studies.

3.1 Traffic Conflict Studies: Population and Sample

In a study or survey, the population refers to the actual or hypothetical group of units with characteristics that can be represented using random variables and their distributions. These characteristics are identified from the individual units, called observation units, within the survey population. Survey sampling involves real-world populations, which are considered a finite set of observable units.

The survey population can therefore be represented by the set of N labels:

$$U=\{1,2,3\ldots N\} \quad (3)$$

where U refers to the population or equivalently the universe and N is the population size (Wu & Thompson, 2020).

Each unit labelled i ($i=1,2,\ldots,N$) represents a unique observable unit in the population. Surveys can target various populations and depending on the study's objective, the survey population may be a subset of a larger population.

The population for the TCT includes all possible conflicts between road users. This includes all drivers, pedestrians, cyclists, or any other individuals who interact with the road entity of interest and other road users. The survey population of the TCT is the specific subset of the entire population and is defined using eligibility criteria. Eligibility criteria are a specific set of characteristics that define the survey population; it is the inclusion or exclusion of observable units based on the eligibility chosen for the survey. This ensures that observations are conducted under homogeneous conditions, specifying factors such as time of day, observation periods, traffic control measures, number of lanes, and specific conflict zones (e.g.,

intersections, merging zones). Observations focus on conflicts occurring between vehicles (veh-veh) and between vehicles and pedestrians (veh-ped).

The survey sample is denoted by S and is a subset of the survey population U :

$$S = \{i_1, i_2, \dots, i_n\} \subseteq U \quad (4)$$

where S is the set of units observed and n is the number of observations recorded (the sample size).

There are two primary types of survey sampling: probability sampling and non-probability sampling. Probability sampling is more common and forms the basis of modern survey sampling theory and practice. However, the use of non-probability sampling is increasing due to its time efficiency and cost-effectiveness. Probability sampling selects units for the survey sample based on a probability measure over all possible samples. This probability measure is the likelihood of each experimental outcome occurring. By considering the probability of each outcome when selecting sample units, probability sampling eliminates potential biases through randomization. This method allows for statistically rigorous statements about the population, with well-defined margins of error that are controlled by the sample size.

Non-probability sampling obtains data from readily accessible parts of the population, such as choosing video footage from intersections or road segments that are believed to be most representative of a city. . Non-probability sampling is useful for exploratory studies, such as pilot studies, to gather preliminary information about the population. The validity of statistical inferences based on non-probability samples relies on model assumptions and available information about the survey populations. While challenging, data analysis using non-probability samples is possible with the correct methods

There are also two specific types of probability sampling: stratified sampling and cluster sampling. Stratified sampling divides the population into different groups, called strata. Samples are then collected from each stratum. Cluster Sampling also divides the population into groups, but data is only collected from specified clusters. The main difference between these two methods lies in how the survey sample is selected from the groups.

For example, when collecting traffic conflict data for a road safety study in the City of Toronto, stratified sampling would involve collecting data from multiple districts across the city. Each district's data would then be weighted and combined to represent the entire city, ensuring that all districts are proportionally represented in the analysis. In contrast, cluster sampling would involve selecting a few specific districts that are deemed representative of the entire city. Data would then be collected solely from these selected districts, with the assumption that the data from these districts can generalize to the entire city.

Sampling frames are a list of all the individual units that can be sampled from the population. One of the major sources of errors in surveys is due to an incomplete sampling frame where it is not an accurate and complete representation of the population.

For example, the sampling frame for a traffic conflict study would be all the observed conflicts in a traffic conflict at a specific intersection. However, based on the definition of the population, it may be an incomplete sampling frame if certain time periods are omitted, if data for some conflicts are not collected because the camera view was obstructed or was misclassified, or if the one intersection was used to represent the entire City.

There are three main justifications for using a survey sample rather than a census (where every individual or unit within the population is surveyed):

- **Cost:** Survey samples can provide sufficient and reliable information at far less cost than attempting to collect data from an entire population.
- **Time:** Survey samples can be collected relatively quickly, and results can be made available in a timely fashion.
- **Accuracy:** Estimates based on a well-designed and well-executed survey sample are often more accurate than results from a poorly conducted census. By using high quality survey sample data, appropriate sampling methods, and suitable sample sizes, reliable statistical inferences and conclusions can be made to any desired level of accuracy.

3.2 Survey Design for Traffic Conflict Studies

Traffic conflict analyses using the TCT vary depending on the objective of the study. Some may try to use the technique to conduct network screening to determine potential locations and corridors of high risk. Others may want to use the TCT to determine the effectiveness of safety countermeasures and various safety interventions. Regardless of the type of safety study being conducted, these surveys generally follow a process like the one illustrated in Figure 3-1.

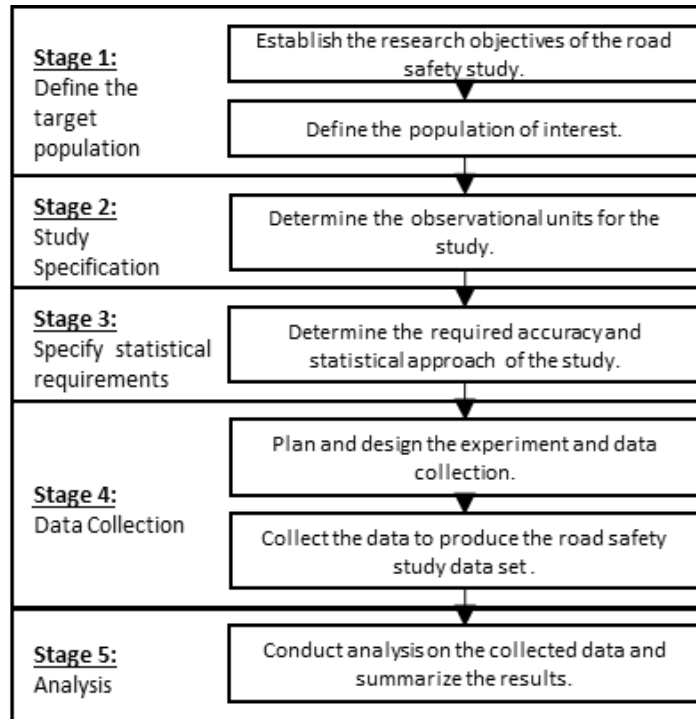


Figure 3-1: Stages of Experiment and Survey Design (Wu & Thompson, 2020)

The first stage involves defining the target population by clearly stating the survey objectives and identifying the population to be sampled. For road safety studies, this means defining the study's objectives (e.g., network screening, diagnosis, before-and-after studies) and determining the units from which the study aims to draw conclusions. For example, the study might focus on all conflicts in the City of Toronto or all conflicts before and after implementing a safety countermeasure.

After defining the research objectives and population, the next step is to select the specific observation units for the study and identify the relevant data required. In a road safety study, these observation units could be collisions or conflicts. The data collected can vary widely, including PET, TTC, vehicle speed, speed differences, and more.

The third stage requires the researcher to determine the necessary accuracy of the estimates and the statistical approach for the study. The required accuracy depends on the study's objectives, financial and time constraints, and the chosen statistical approach. For instance, a pilot study in Toronto to identify traffic conflict characteristics at various intersections might have lower accuracy requirements than developing a model to predict traffic conflict frequency based on speed. This stage also involves establishing the method of statistical analysis, such as hypothesis testing, ANOVA testing, or various regression analyses.

Sensitivity thresholds for statistical tests, such as detecting the impact of a safety countermeasure at an intersection, should be set at this stage.

Once the study objectives and population are defined and an acceptable level of accuracy is determined, the data collection and experiment can be planned. This includes choosing the sampling method and organizing the fieldwork for data collection.

After collecting the data, it should be aggregated into a dataset, and summaries should be produced. The data can then be analyzed, and reports can be written to present the findings.

3.3 Sample Size Requirements

When designing a safety study, investigators should carefully determine the necessary sample size based on several factors. These include the study's budget constraints, the cost associated with collecting each unit of data and conducting required measurements, and the level of accuracy needed for planned statistical inferences (Wu & Thompson, 2020)

It is crucial to define the performance measures and focus of the study to accurately represent the population or sample entity. The focus may center on specific events such as collisions or near misses, which directly impact the population of interest. For example, the population could encompass all observations of conflicts between pedestrians and vehicles that pose potential risks, distinguished by spatial and temporal parameters. This requires establishing metrics based on measurable units within these entities.

Determining the sample size depends on the variation within the population and sampling strategy. Conducting a pilot study can help assess this variation across different locations and conditions, providing insights into how many observations are necessary to reliably capture the desired data. Variations in traffic patterns and environmental conditions underscore the need for tailored observations over specific time periods to understand frequency and severity of conflicts.

For example, the entity of interest could focus on the average frequency of conflicts during peak traffic hours in a designated zone. By observing and quantifying conflicts such as PET or speed, researchers aim to estimate both the frequency and severity of potential collision risks over time. This approach helps address the research question of how many observations are required to minimize statistical error and evaluate the effectiveness of safety countermeasures.

In summary, defining the population, sample size, and entities of interest provides a structured approach to studying safety. This includes setting clear thresholds for what constitutes a conflict of interest and aligning observations with the goal of correlating conflicts to actual collisions within specified conditions and contexts.

3.3.1 Error Propagation

Understanding how error can propagate in conflict-based traffic safety studies provides opportunities to improve the design of the study, ensuring reliable results. Errors propagate through any scientific study through random variation, sampling error, and error due to lack of model fit. Error of measurement is defined as the sampling error, which is the inaccuracy when determining the magnitude of a measurement. Error due to lack of model fit, or error of specification, arises from the simplification imposed by a model of an effect, omission of a variable, or the general lack of fit the model has when compared to reality (Alonso, 1967). Random error and variation are the randomness inherent in the observations themselves. Researchers have worked to improve the validity of traffic conflict studies by reducing measurement error and error in specification. Contributions to reducing error of measurement and specification are summarized in Figure 3-2.

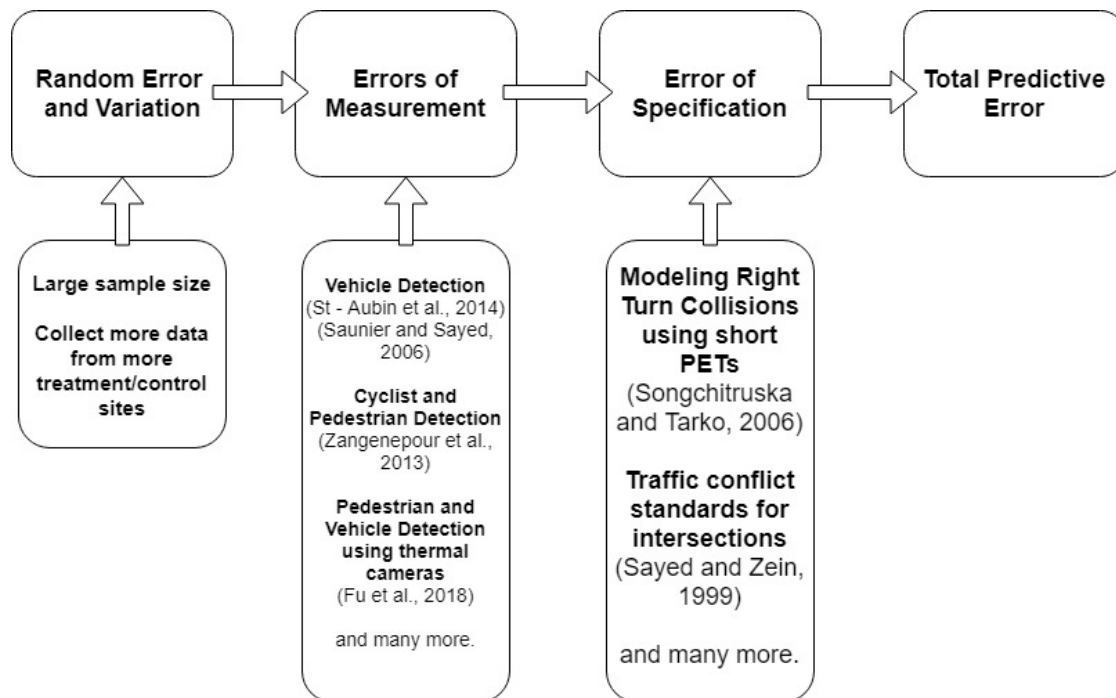


Figure 3-2: Error propagation through a traffic conflict study

Video-based traffic conflict analysis can automatically and accurately collect traffic conflict data from video data and has shown great promise in tracking and classifying vehicles, pedestrians and cyclists. It aims to solve the lack of correlation between observed conflicts and collision frequency due to inaccurate conflict data collection. Researchers have proposed an automated system for proactive road safety analysis using a large video dataset recorded from nearly 40 roundabout weaving zones across 20 different

roundabouts in Quebec, Canada. This automated system achieved a vehicle tracking accuracy ranging from 85 to 95% (St-Aubin et al., 2015b). Similarly, others have proposed a methodology to automate road user classification and tracking of road users at intersections with heavy pedestrian and cyclist volumes using video data. These road users included pedestrians, cyclists, and vehicles. Their methodology achieved greater than 88% overall classification accuracy (Zangenehpour et al., 2015). Further work has adapted the video-based classification methodology in Zangenehpour et al. (2015) for the use of thermal video cameras. Their study developed a methodology to evaluate pedestrian crosswalk safety at non-signalized intersections at night, and the classification algorithm achieved an accuracy of 95% (Fu et al., 2016).

Many predictive models between conflicts and collisions have been developed. These models aim to minimize variance of the unbiased safety estimates from the TCT. Researchers investigated the relationship between conflicts and right-angle collisions at intersections by collecting post encroachment time (PET) and collision data from 16 signalized intersections in Indiana. The study found that PET frequency is a key determinant of right-angle collisions and can indicate safety levels (Songchitruksa & Tarko, 2006). Sayed and Zein (1999) conducted a conflict survey at 94 intersections in British Columbia, Canada and found that the conflict-collision models resulted in R^2 values of up to 0.77

3.3.2 Sample Size Determination in Past Studies

Using SMOs as indicators of safety would reduce the required observation period, since conflicts occur more frequently but are of lower severity. It is still important to consider the sample size for safety studies based on SMOs to control for absolute or relative estimation errors. When designing a safety study, investigators should consider how large should the sample size be. This depends on the budget of the study, the cost of collecting one unit of data and taking all required measurements, and the accuracy required for the planned statistical inferences (Chin & Quek, 1997; Cohen, 1992; Hauer, 1978).

Most studies collect as many observation samples of traffic conflicts as possible to ensure the validity of the research results. However, few studies on traffic conflicts discuss the determination of the sample size in detail. For example, Rajeswaran et al. (2023) modeled 91 four-legged signalized intersections, and then used VISSIM simulation software to conduct a one-hour traffic simulation for each intersection to collect traffic conflict data including TTC and PET, to build the crash-conflict models. Hauer (1978) assumed that the expected conflict rate obeys the negative binomial distribution, and then calculated the required sample size for traffic conflict research under different confidence intervals. Alluri et al. (2016) estimated the minimum sample sizes to get the reliable Highway Safety Manual (HSM) calibration factors using Empirical Bayes (EB) method, and the results showed that a general one-size-fits-all approach of collecting

a sample size of 30 to 50 sites is not appropriate because different facility types require different sample sizes.

In summary, it is essential to adopt a statistical method to determine the optimal sample size according to different research objectives, different data distribution and different expected effects in traffic conflict-based studies, to avoid waste of resources while ensuring the effect of the model.

3.4 Minimum Sample Size for Controlling Estimation Errors

3.4.1 Methodology and Conceptual Framework

This section introduces a framework based on statistical power analysis for determining the minimum sample sizes for conflict analysis studies. As shown in Figure 3-3, the framework can be broken down into 5 main stages:

- 1) Project initialization – the project is initiated with the required background information including project objectives and scope and constraints;
- 2) Objective specification – determine the main objective of the traffic conflict analysis;
- 3) Study Specification – there are many different types of studies that employ traffic conflict analysis, four of the most common of which are included in the framework;
- 4) Statistical test specification – five of the most common statistical tests in traffic safety analysis have been included in the framework (for others, see a statistics reference such as Cohen (2013) for how to conduct an associated power analysis); and
- 5) Determine the inputs for power analysis.

As shown in Figure 3-3, the inputs required for power analysis (stage 5) depend on the statistical test employed in the study. The appropriate statistical test (stage 4) depends on the study specification, which generally falls into four categories: before and after studies, cross-sectional studies, conflict and collision modelling, or general-purpose monitoring and methodological development. Finally, the study specification (stage 3) depends on the study objective, which is often to either determine a safety impact or develop a model between conflicts and collisions. Hence, using the a priori power analysis framework, an investigator can determine the appropriate sample size for their traffic conflict study.

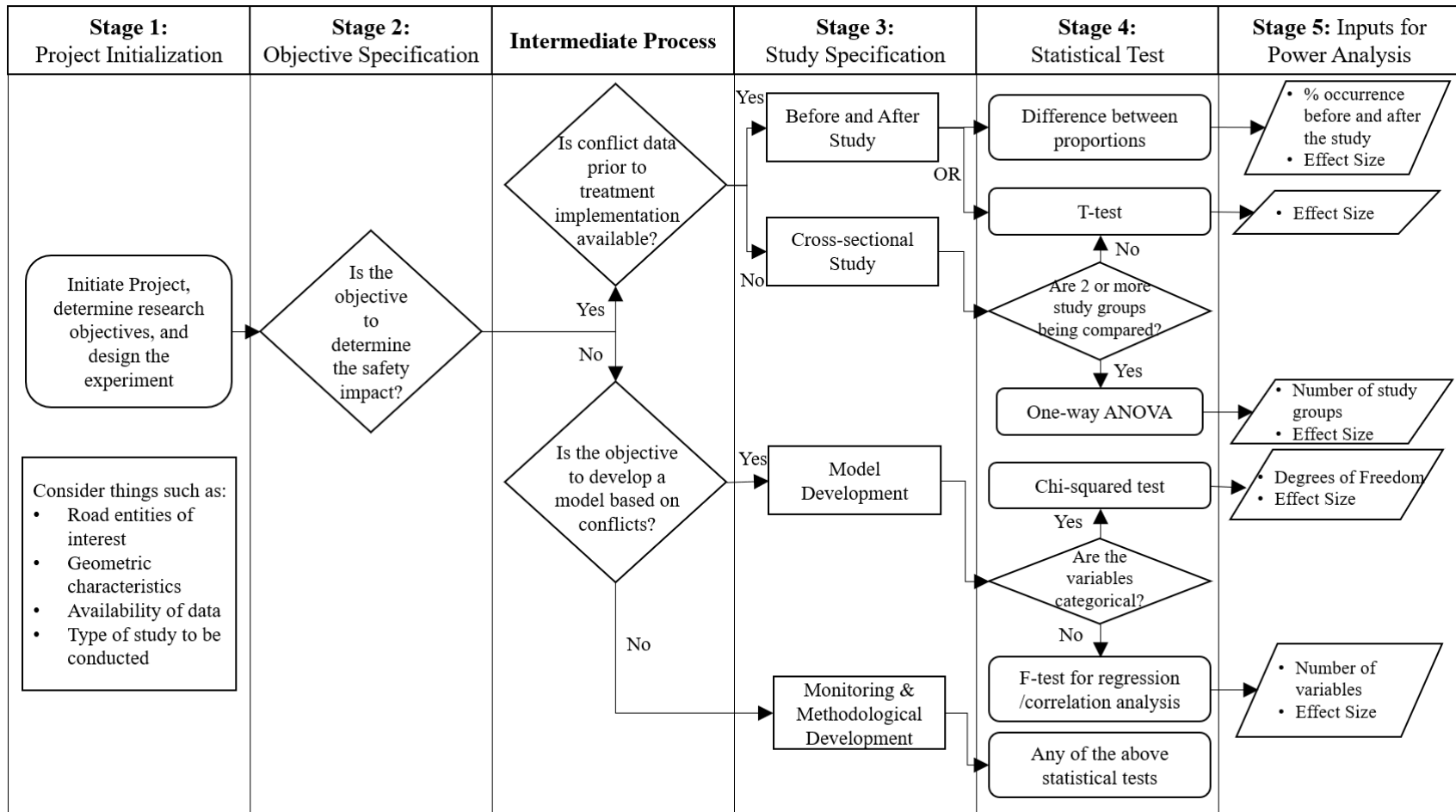


Figure 3-3 Statistical Test Specification and Input flowchart for Common Traffic Conflict Studies

3.4.2 Minimum Sample Size for Controlling Type II Errors - Power Analysis

In hypothesis testing, there are two types of errors that can occur, as shown in Figure 3-4. A Type I error is the error of rejecting a true null hypothesis (H_0), commonly referred to as a “false positive”. The probability of Type I error is often denoted as α (Sheskin, 2003). A Type II error is defined as when the hypothesis test fails to reject a false null hypothesis (H_0), commonly referred to as a “false negative”. The probability of Type II error is often denoted as β (Sheskin, 2003).

	H_0 Rejected	Fail to reject H_0
H_0 True	Type I Error	Correct
H_0 False	Correct	Type II Error

Figure 3-4 Potential Errors in Statistical Hypothesis Testing

Power analysis is based on hypothesis testing and exploits three variables: statistical power ($1-\beta$), significance criterion (α), and population effect size (ES) (Cohen, 1992). Power analysis allows investigators to find the minimum sample size required such that Type I and Type II errors are controlled for while considering how “noticeable” the effect will be. By convention, $\alpha = 0.05$ and $\beta = 0.2$ are used as the maximum allowable risk of Type I and II errors, respectively (Hoenig & Heisey, 2001). The sample size (N) is the minimum number of observations required for the results of the experiment to be statistically significant from the power analysis. The sample size is the minimum number of observations required for each study group. Statistical power (denoted by $1-\beta$) is the probability that a false null hypothesis (H_0) will be rejected, or the probability that the experiment will not be subject to a Type II error. It is convention that $1-\beta = 0.80$ or the probability of Type II error being committed is 0.20 (Cohen, 1992). The significance level or criterion, as denoted by α , is the risk of mistakenly rejecting a true null hypothesis (H_0), and thus committing a Type I Error. Hence, the significance criterion is the maximum allowable risk of committing a Type I error. Literature suggests the conventional value of α is 0.05 (Cohen, 1992). In formulating a hypothesis, the Alternative to the Null (H_a) is often expressed in terms of a variable or measure with respect to a pre-specified value called effect size such as the difference in frequency of a given type of conflicts between before and after an intervention is introduced at a site (Cohen, 2013). In general terms, the population effect size (ES) is calculated through a transformation such that the effect size is dependent on

the scale to which the quantities of interest deviate rather than the difference in magnitude. Transformations vary with different statistical tests and are outlined in Cohen (2013).

For road safety researchers and practitioners, the ES of a treatment to be detected and used in a hypothesis testing is difficult to determine prior to investigation, especially for new types of treatments that have been rarely studied in the past. For example, the results of the treatments may have counterintuitive impacts, and hence estimating the ES of a treatment before it is implemented is difficult. The key guideline for determining the detectable ES is its practical importance or relevance. Depending on the scale or significance of the treatment projects, three levels of effects have conventionally been used to define ES, including small, medium, and large. As defined in Cohen (1992), “Medium ES represents an effect likely to be visible to the naked eye of a careful observer. Small ES is to be noticeably smaller than medium but not so small as to be trivial, and large ES is set to be the same distance above medium as small was below”. These definitions are subjective but help put into perspective the different scales of effects possible. The different effect sizes provide a guideline for the sample sizes required for experiments and may require expert opinion and experience to hypothesize the effect size of a treatment.

If the statistical power is low or the sample size is too large, then the investigators have an opportunity to reevaluate the experiment and fix the problem before conducting an underpowered experiment (Cohen, 2013). Table 3-1 summarizes the effect sizes for different tests required for power analysis (Cohen, 1992).

Table 3-1 Typical Effect Sizes for Various Statistical Tests (Cohen, 1992)

<u>Test</u>	<u>Applications</u>	<u>Effect Size</u>		
		<u>Small</u>	<u>Medium</u>	<u>Large</u>
Difference between proportions	Before and after study	0.20	0.50	0.80
t-test for difference in standardized means	- Before and after study - Cross sectional Study	0.20	0.50	0.80
One-way ANOVA	Cross sectional study	0.10	0.25	0.40
Chi-square for goodness of fit	Development of predictive models (using categorical variables)	0.10	0.30	0.50
F-test	Development of predictive models (using continuous variables)	0.02	0.15	0.35

The ES for the power analysis of each type of the above statistical tests can be performed using the following formulas as follows:

The difference of proportions test is used to determine whether there is a significant difference between the proportions of a certain characteristic in two independent groups. For a two-sample comparison of proportions, the effect size h is typically calculated using the following formula:

$$h = \varphi_1 - \varphi_2 \quad (5)$$

where

$$\varphi_1 = 2 \arcsin \sqrt{P_1} \quad (6)$$

$$\varphi_2 = 2 \arcsin \sqrt{P_2}$$

where P_1 and P_2 are the proportions of the two groups being compared, φ_1 and φ_2 are the arcsin transformation of P_1 and P_2 .

T-tests can be used to determine if there exists a difference between the means of two groups. The effect size d for difference between standardized means can be computed from:

$$d = \frac{|\mu_1 - \mu_2|}{\sigma} \quad (7)$$

Where d is the effect size, μ_1 and μ_2 are the means of the two populations of interest, and σ is the common standard deviation.

To determine the common standard deviation, the following equation is recommended if the standard deviations are found to be similar (Cohen, 2013):

$$\sigma = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}} \quad (8)$$

where σ_1^2 and σ_2^2 are the variances of each study group.

One-way Analysis of Variance (ANOVA) is used to determine if there is a statistically significant difference between the means of three or more independent groups. One-way ANOVA tests use the effect size f and it can be represented as:

$$f = \sqrt{\frac{\sum_{i=1}^k P_i \times (\mu_i - \mu)^2}{\sigma^2}} \quad (9)$$

Where $P_i = n_i/N$; n_i is the sample size of group i ; N is the total number of sample size; μ_i is the mean value of group i , μ is the mean value of total groups; and σ^2 is the error variance within groups, k is the number of groups

Chi-Square Goodness of Fit test is used to determine whether a sample matches a population with a specific distribution. For Chi-Square tests, the effect size w uses the formula:

$$w = \sqrt{\sum_{i=1}^k \frac{(p_{0i} - p_{1i})^2}{p_{0i}}} \quad (10)$$

where p_{0i} is the cell probability in i th cell under H_0 ; and p_{1i} is the cell probability in i th cell under H_1 and k is the number of cells

The F-test is used to compare variances of two or more groups to determine if they are significantly different from each other and is commonly used in the context of ANOVA and regression analysis. For F-tests, the effect size f^2 uses the formula:

$$f^2 = \frac{R^2}{1 - R^2} \quad (11)$$

where R^2 is the population squared value of multiple correlation.

3.5 A Priori versus Post Hoc Power Analysis

A priori power analysis can be conducted in two ways: to determine the power for a given sample size, α , and ES; or to determine the sample size given the α , power, and ES. The first method allows researchers to check if the experiment results will be underpowered or not statistically significant. The second method, which aligns with the objective of this study, allows researchers to find the minimum sample size such that the experiment results will be statistically significant.

Post hoc power analysis is where the data have already been collected and a researcher wants to check the *observed* power, sample size based off the *observed* effect size, and the *detectable* effect size. Post hoc power analysis is conducted in two of the case studies in this paper. However, it is not generally advisable to conduct post hoc power analysis. Consider post hoc power analysis from two perspectives: intuitively and mathematically.

When conducting a priori power analysis, the significance criterion, power, and effect size generates the minimum sample size that corresponds to the user-specifications. The post hoc power analysis will generate results that correspond with the user-specifications. Statistically significant inputs in the post hoc power analysis would result in a statistically significant output and vice versa, inputs that are not statistically significant will result in an output that is not statistically significant. Post hoc power analysis would only confirm the significance of the user-specifications and provide no new insights; this is known as the “power approach paradox” (PAP) (Hoenig & Heisey, 2001). Hence, power analysis is most useful in the research planning stage to ensure protection from Type I and II error.

This phenomenon is explained with more mathematical rigor. Through many simulations, researchers concluded that “the observed power from post hoc power analysis is almost always a biased estimator of the true power” (Yuan & Maxwell, 2005). The post hoc power analysis skews the observed power to confirm the results of the experiment more readily.

Rather than employing post hoc power analysis, different strategies can be used to check the significance of an experiment. The investigator can check the p -values of the experiment or construct a confidence interval. Confidence intervals are useful because they can be used to determine the range of possible results within the required level of confidence without falling victim to the PAP.

In this research, we conducted the power analysis using two statistical analysis packages available in R: ‘pwr’ and ‘pwr2’ (Champerly et al, 2018; Lu et al., 2017). The ‘pwr’ and ‘pwr2’ packages are available through the Comprehensive R Archive Network (CRAN). By providing the corresponding variables required for power analysis, the minimum sample size required to achieve a conventional or user-specified significance level and statistical power is determined.

3.6 Case Studies: Post Hoc Power Analysis

Two case studies are used to illustrate the application of the proposed framework using real-world conflict data. These instructive examples use post hoc analyses which demonstrate how power analysis is conducted for different statistical tests. The first case study is based on Fu et al., 2016 and illustrates the use of power analysis for a difference in proportions test. The second case study is based on St-Aubin et al., 2013 and illustrates the use of power analysis for a difference in means test. Recall it was established in the previous section that post hoc power analysis should not be done in practice. It is carried out in these case studies for instructive purposes only. Rather, a priori power analysis should be conducted in the research planning stages; an example of a priori power analysis is conducted in Chapter 4.

3.6.1 Case Study 1: Pedestrian crosswalk safety at non-signalized crossings during nighttime (Fu et al., 2013).

Fu et al. (2013) conducted a conflict-based study on pedestrian safety at crosswalks using data collected by thermal video cameras. The study compared the pedestrian-vehicle conflict rates during the day and night. The sample size was not directly provided in the study but the observation period and the hourly number of interactions with PET less than 20 seconds was provided. Conflict rates are defined as the number of interactions with PETs of less than 5s, divided by the number of interactions with PETs of less than 20 s. The arcsin transformation (equation (6)) is applied, using the conflict rates as proportions, to determine the ϕ . Hence, the product of these two values is considered as the sample size. The sample size and conflict rates are summarized in Table 3-2.

For this case study, using the R package ‘pwr’, $\alpha = 0.05$, $1 - \beta = 0.8$, and the sample sizes provided (Table 3-2), the detectable effect size was determined to be approximately 0.156 and 0.029, for the du Fort and Saint Laurent sites, respectively. These values are summarized in Table 3-2.

Table 3-2 Use of thermal cameras for conflict analysis - summary of data

<u>Location</u>	<u>Sample Size in Study</u>		<u>Conflict Rate (%)</u>		φ		<u>Observed Effect Size</u>	<u>Detectable Effect Size</u>
	Day	Night	Day	Night	Day	Night	<u>(h_o)</u>	<u>(h_d)</u>
du Fort	416	1400	3.16	5.97	0.357	0.494	0.136	0.156
Saint Laurent	16137	22351	0.97	0.99	0.197	0.199	0.002	0.029

At the du Fort location, the observed effect size is lower than the detectable difference effect size, which means that the difference in daytime and nighttime conflict rates are not detectable at the current sample size. Similarly, the observed effect size at Saint Laurent is lower than the detectable effect size, suggesting the difference in daytime and nighttime conflict rates at this location are not detectable at the current sample size. Recall that nothing can be said definitively using post hoc power analysis, but the case study nevertheless illustrates the arcsin transformation associated with the difference in proportions test for the purposes of conducting a power analysis. To better determine the optimal sample size in this study, Figure 3-5 and Figure 3-6 illustrate the sample size vs. power for different proportion p_1 at du Fort and Saint Laurent sites for a given p_2 and α . From the results of power analysis shown in Figure 3-5, it can be concluded that the required daytime and nighttime traffic conflict sample sizes at the current effect size level, given power value, and significance criteria are 459 and 1,545, respectively. For Saint Laurent site, as shown in Figure 3-6, the required daytime and nighttime traffic conflict sample sizes are 2,584,189 and 3,579,102, respectively. This is due to the observed effect size being very small, requiring more observations to detect this impact.

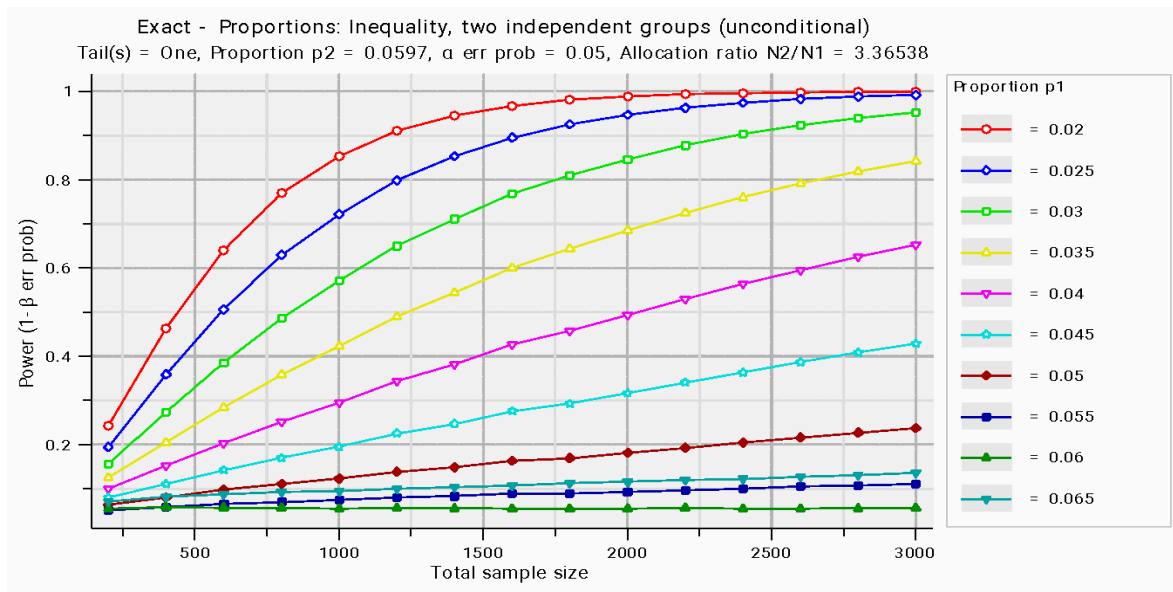


Figure 3-5 Sample size vs. power for different proportions of p_1 at du Fort site for a given p_2 and α

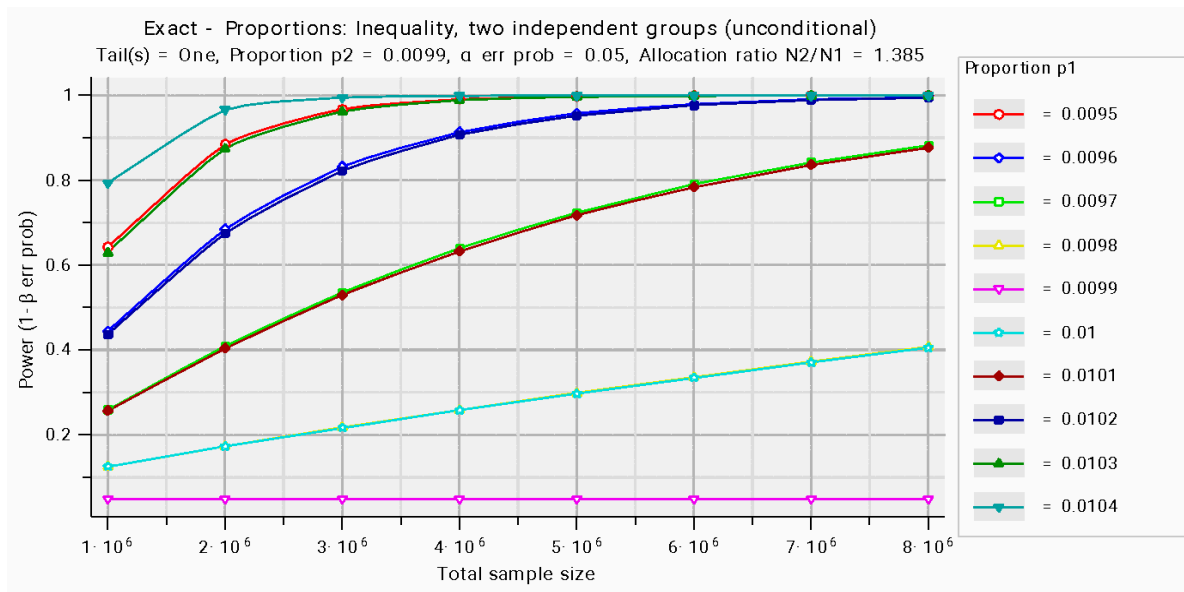


Figure 3-6 Sample size vs. power for different proportions of p_1 at Saint Laurent site for a given p_2 and α

3.6.2 Case Study 2: An automated surrogate safety analysis at protected highway ramps using cross-sectional and before-after video data (St-Aubin et al., 2013).

St-Aubin et al. (2013) conducted a conflict-based study evaluating the impact of lane-change bans on highway entrance and exit ramps in the City of Montreal. The main conflict indicator used is TTC and the traffic conflicts were categorized as Type A and Type B. Type A are rear-end interactions and Type B are the diagonal interactions experienced when vehicles change lanes (i.e., lane change conflicts). In terms of study specification, a before and after study is considered. The researchers conducted a before-after comparison using weighted interaction density maps and found no clear evidence that the treatment had an effect on the two types of conflicts. This study was selected because the before and after summary statistics (including: the sample size, the observed mean, and the observed standard deviation) are provided for one location (Site A20-E-E56-3).

For this case study, a t-test is conducted on the difference in means (TTC). The observed effect sizes are calculated through equations (7) and (8), and the detectable effect size are calculated through post hoc power analysis. Table 3-3 summarizes the case study's results.

Table 3-3 Summary Statistics for Site A20-E-E56-3

<u>Type of Interaction</u>	<u>Reported Sample Size</u>		<u>Mean TTC</u>		<u>Standard Deviation of TTC</u>		<u>Common Standard Deviation (σ)</u>	<u>Observed Effect Size</u>	<u>Detectable Effect Size</u>
	<u>Before</u>	<u>After</u>	<u>Before</u>	<u>After</u>	<u>Before</u>	<u>After</u>			
A – Rear End	247143	268635	10.7	11.2	9.3	11.1	10.24	0.049	0.008
B – Diagonal Interaction	178407	299546	21.1	20.8	12.6	12.5	12.55	0.024	0.008

The observed effect size is greater than the detectable difference in effect for both types of interactions, indicating that the before and after effects of the lane change ban treatment could be captured with this sample size. Again, nothing can be said definitively using post hoc power analysis, but the case study nevertheless illustrates how to conduct a t-test on a difference of means for the purposes of conducting a power analysis. In addition, using power analysis, the sample size vs. statistical power for different effect sizes for two types of traffic conflicts are shown in Figure 3-7 and Figure 3-8. It can be seen from Figure

3-7 that according to the effect size and data ratio of the rear-end conflict obtained from the before and after observations in this case, as well as the given power value, the sample sizes required for the before and after observations obtained by the power analysis are 4,945 and 5,375 respectively. Similar results can also be obtained from Figure 3-8, where the required sample size for the before and after observations for diagonal interactions analysis obtained by the power analysis are 17,125 and 28,757 respectively.

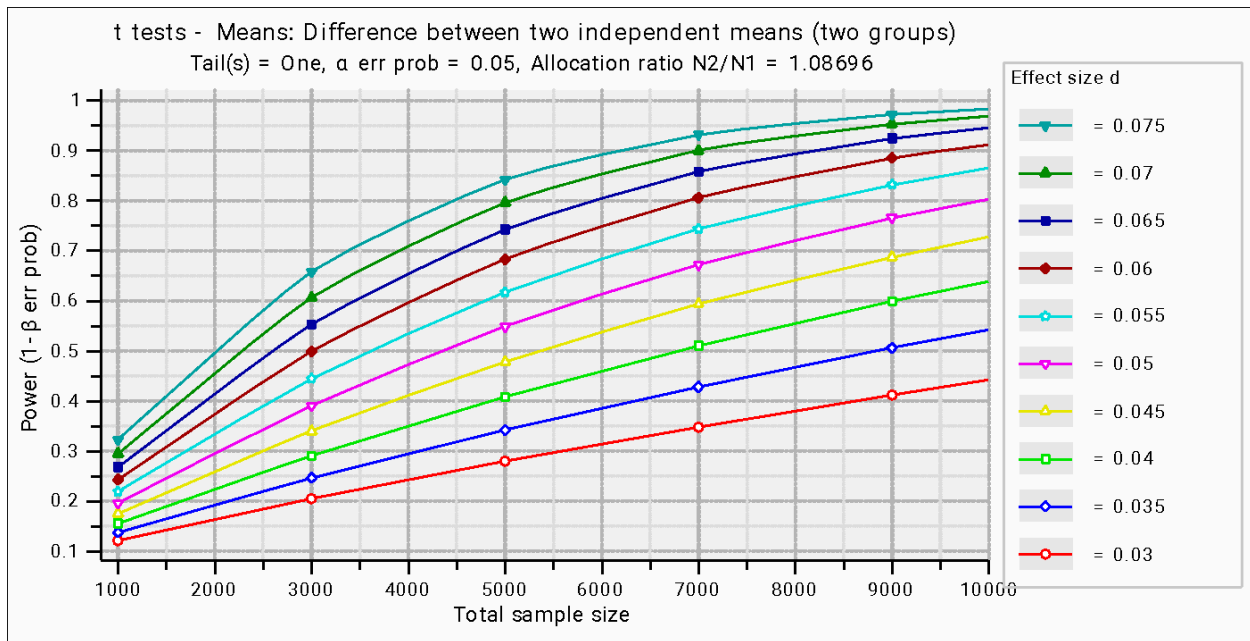


Figure 3-7 Sample size vs. power for different effect sizes for Type A conflict at a given α

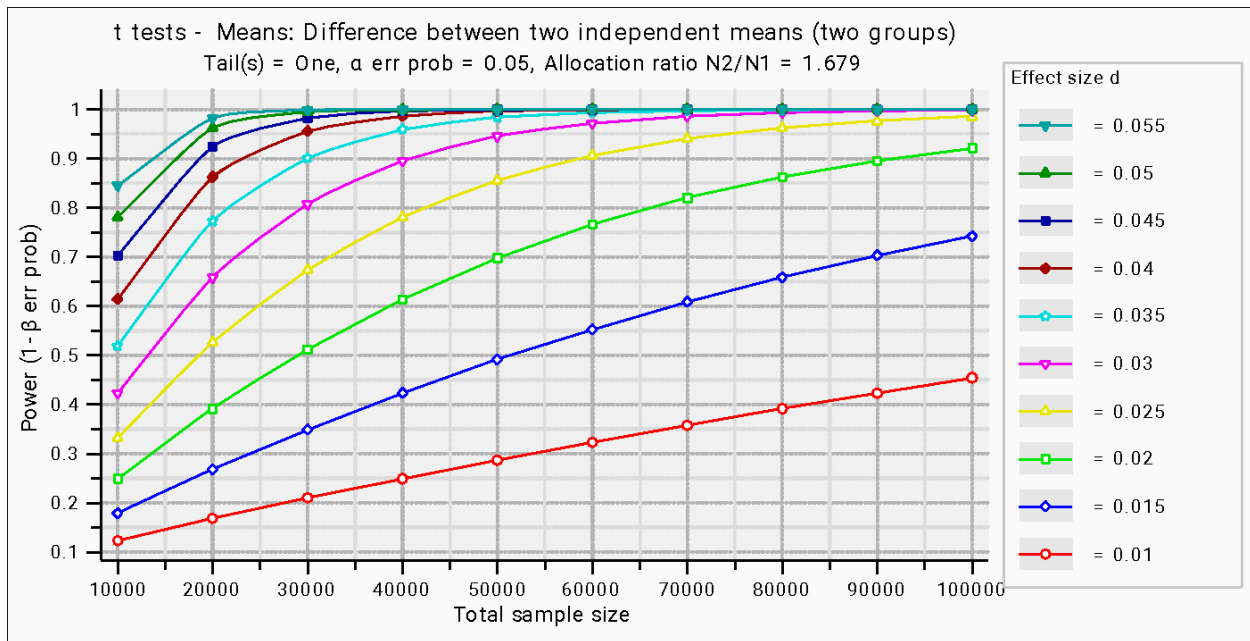


Figure 3-8 Sample size vs. power for different effect sizes for Type B conflict at a given α

3.7 Application of Minimum Sample Size Framework: A Priori Power Analysis

This section illustrates the application of the proposed power analysis method for determining the minimum sample size for a real-world project conducted in the City of Toronto. The project was to investigate the application of a new video-based platform for real-time road safety data collection and the evaluation of traffic safety countermeasures. One of the objectives of this project was to conduct a cross-sectional study to investigate the safety impacts of No Right Turn On Red (NRTOR) and Leading Pedestrian Interval (LPI) policies on vulnerable road users. During the data collection design phase, the proposed framework to determine minimum sample size for a traffic conflict analysis study is used.

It was assumed that a one-way ANOVA test will be conducted to determine the safety impact of each treatment (Fu, 2018). The power analysis considers different effect sizes using the R package ‘pwr’ at $\alpha = 0.05$ and $1 - \beta = 0.8$. Figure 3-9 compares the effect of sample size on the statistical power of this experiment for different effect sizes from 0.1 to 0.55.

It is important to note the sample size difference between small and medium and between medium and large effect sizes. The magnitudes between small and medium and between medium and large effect size are the same but the small effect size requires 14 times as many observations as the large effect size. By following the same logic as Figure 3-9 it would be expected that the sample size would be larger for a smaller than

expected effect size. It is possible more observations should be considered if the effect size is indeed much smaller than 0.1. By using this comprehensive framework, researchers can reevaluate the experiment and ensure protection from Type I and II errors.

Table 3-4 summarizes the optimal sample size for each effect size level. From the collected data, it was determined that the average daily pedestrian-vehicle conflict frequency at the intersection of Lake Shore Boulevard East and Parliament Street intersection is 7 conflicts per day. Based on the power analysis and the average conflict frequency, the project will require 3 to 40 days of observations (Table 3-4). It is operationally feasible to study these intersections for 40 days or approximately two months.

In addition to ANOVA test, a Chi-square test can also be applied for this project to check whether there is a relationship between two categorical variables (with countermeasure vs. without countermeasure, conflict vs. non-conflict). Chi-square test requires an additional parameter – Degree of freedom (Df). In this case, instead of analyzing the data using a test of independent proportions, the chi-square statistic could be calculated in the 2×2 contingency table. Figure 3-10 compares the effect of sample size on the statistical power of Chi-square test for different effect sizes from 0.1 to 0.55 when Df equals 1.

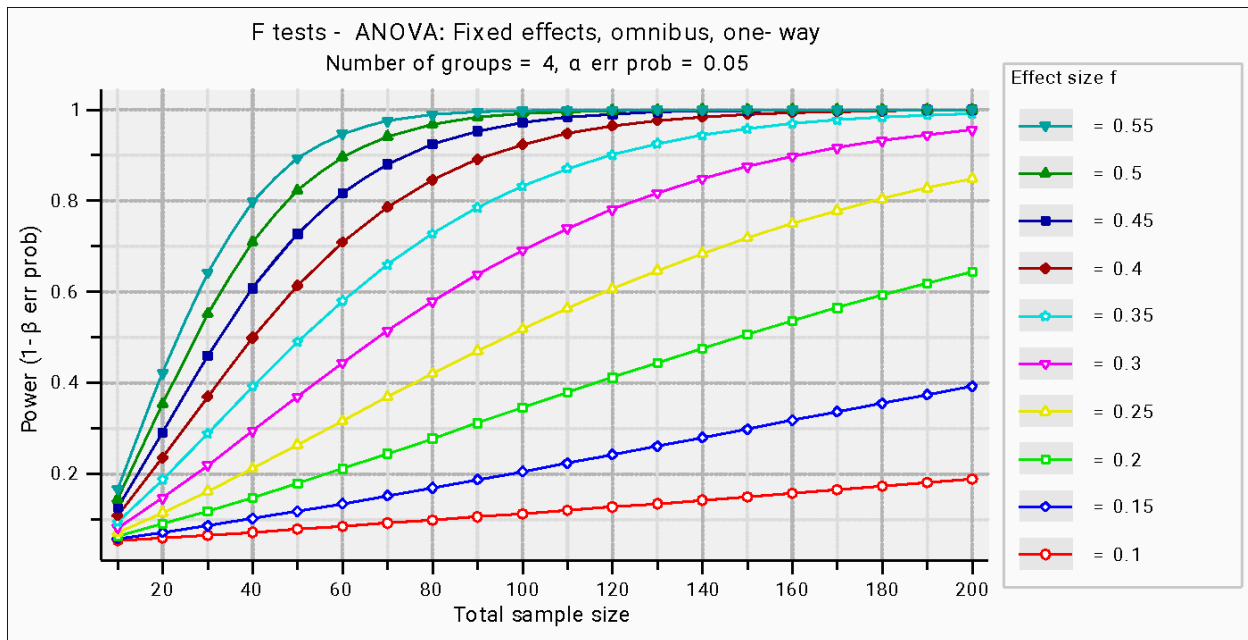


Figure 3-9: Sample size vs. power with different effect sizes for F tests - ANOVA

Table 3-4 Effect size, sample size, and observation period (F test - ANOVA)

<u>Anticipated Effect Size</u>	<u>Effect Size</u> (Cohen, 1992)	<u>Minimum Sample Size</u>	<u>Estimated</u> <u>Observation period</u> <u>(days)</u>
Small	0.10	274	40
Medium	0.25	45	7
Large	0.40	19	3

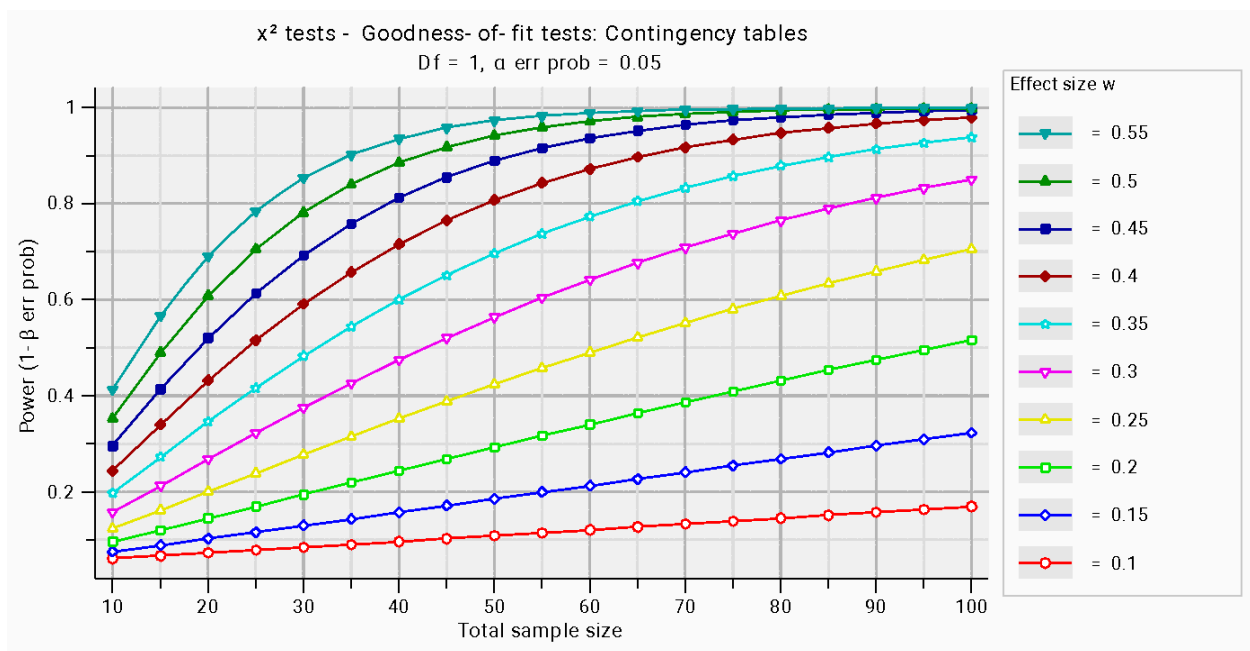


Figure 3-10 Sample size vs. power with different effect sizes for χ^2 test

Based on the three different effect size levels for Chi-square test recommended by Cohen (1992), we can conclude that the required sample sizes for the small, medium, and large effect size levels are 56 days, 7 days, and 3 days, respectively, as shown in Table 3-5.

Table 3-5 Effect size, Sample Size and observation period (Chi-square test)

<u>Anticipated Effect Size</u>	<u>Effect Size</u> (Cohen, 1992)	<u>Minimum Sample Size</u>	<u>Estimated</u> <u>Observation period</u> <u>(days)</u>
Small	0.10	392	56
Medium	0.30	44	7
Large	0.50	16	3

Lastly, if the researcher develops predictive models using the continuous traffic conflict variables, the F-test from the power analysis can be used to determine the optimal sample size at a given α level, power, number of predictors and effect size. Assume that in this experiment, α is 0.5, power is 0.8, the number of tested predictors is 1, and the number of total predictors is 5. Figure 3-11 compares the effect of sample size on the statistical power of this experiment for different effect sizes from 0.02 to 0.38. Also, based on the suggested effect sizes, it can be concluded that the required sample size for the small, medium, and large effect size levels are 57 days, 8 days, and 4 days, respectively, as shown in Table 3-6.

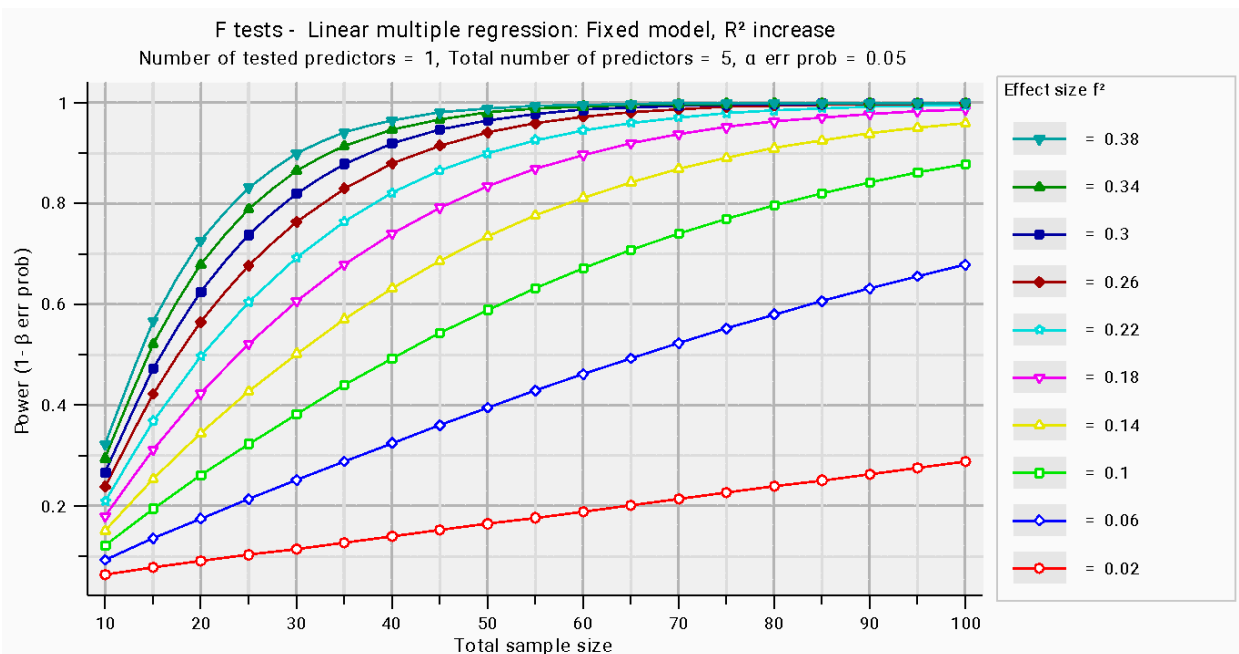


Figure 3-11 Sample size vs. power with different effect sizes for F test

Table 3-6 Effect size, Sample Size and observation period (F test)

<u>Anticipated Effect Size</u>	<u>Effect Size</u> (Cohen, 1992)	<u>Minimum Sample Size</u>	<u>Estimated Observation period (days)</u>
Small	0.02	395	57
Medium	0.15	55	8
Large	0.35	25	4

Chapter 4

Conflict-Based Study for Evaluating the Effectiveness of Vision Zero Safety Countermeasures

This chapter presents the analysis of video-based traffic conflict data to determine the effectiveness of safety countermeasures implemented in the City of Toronto. The following sections present the selection of the safety countermeasures and study sites, a summary of the conflict variables, and the evaluation of the effectiveness of the proposed traffic countermeasures.

4.1 City of Toronto Vision Zero Safety Countermeasures

This study aims at quantifying the safety effect of several specific safety countermeasures widely implemented by the City of Toronto. These countermeasures were implemented as part of Toronto's Vision Zero Road Safety Plan - a comprehensive five-year (2017-2021) action plan which focused on reducing traffic-related fatalities and serious injuries on the streets of Toronto. Three traffic countermeasures are considered in this study: Leading Pedestrian Interval (LPI), No Right Turn On Red (NRTOR), and the installation of dedicated Bicycle Lane (BL).

LPIs are safety countermeasures implemented at signalized intersections. With LPI, pedestrians begin crossing the street before vehicles are given the green light (Guo et al., 2020). Pedestrians, who may be in the blind spot of an approaching vehicle, can walk into the intersection and appear in the driver's field of view before the driver can begin their movement. LPI increases the visibility of pedestrians to drivers and may reduce conflicts with turning vehicles. Furthermore, this safety countermeasure is also cost-effective to implement, as it only requires the retiming of the traffic signals.

NRTOR, as a safety intervention, prevents vehicles from making a right-turn movement during a red light phase (Ellis & Van Houten, 2009). Allowing vehicles to turn right on red lights can have an impact on the safety of pedestrians crossing the streets, especially at locations with high pedestrian traffic and areas near schools, parks, and senior homes. In areas with high pedestrian traffic, turning right on red may not be possible due to the constant flow of pedestrians crossing the street; essentially rendering right turns on red useless. One related safety issue that arises is that drivers may fail to stop before the crosswalk and look both ways before making a right turn on a red. Prohibition of right turn on red addresses some of these issues despite delaying vehicles that want to make a right turn.

To quantify the safety effect of these countermeasures, this research applies a combination of case-control and cross-sectional studies. Treatment sites (with a countermeasure of interest) and control sites (without

any countermeasures) are first identified and safety data from these sites are then collected. Figure 4-1 illustrates the 10 selected signalized intersections in the City of Toronto with the three different types of traffic countermeasures being implemented: three intersections with NRTOR group, two intersections with LPI, one intersection with a bike lane on the shoulder of the roadway, and four intersections as part of the control group. All intersections were chosen in collaboration with City of Toronto staff with the following two considerations: i) selected locations with safety countermeasures of interest; ii) the selected intersections have similar geometric characteristics and traffic and environmental conditions.

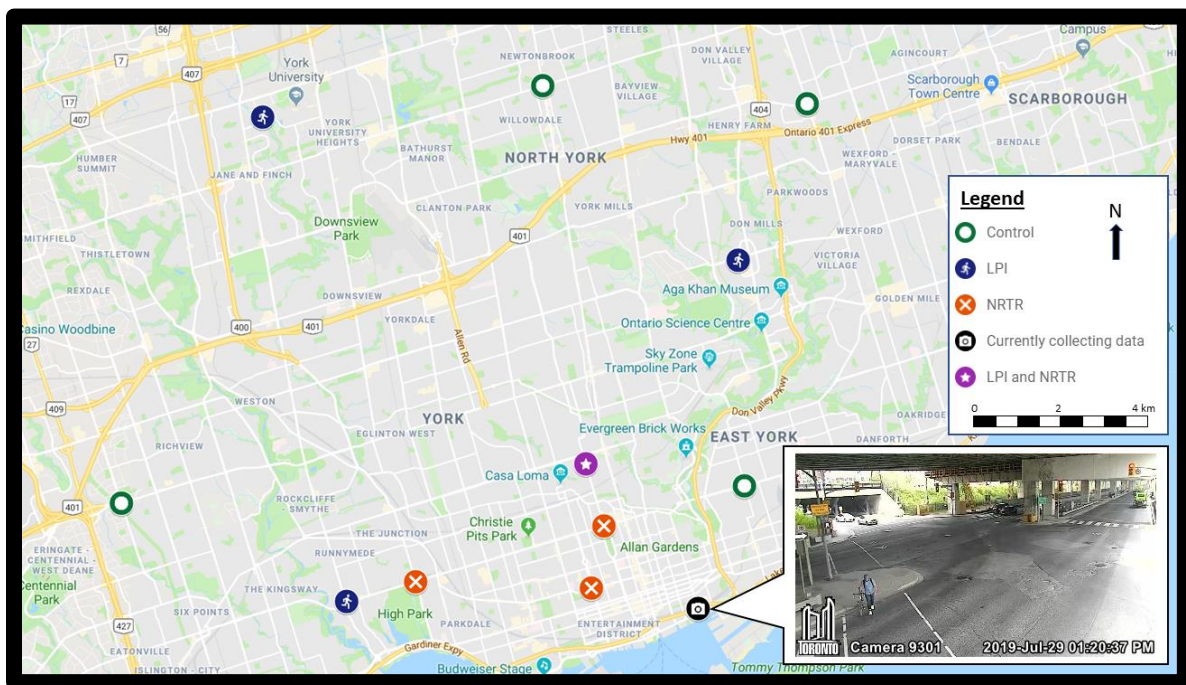


Figure 4-1: Map of study sites and camera view at the intersection of Lake Shore Boulevard East and Parliament Street

4.2 Video Conflict Analysis Scenarios

Once the study sites and the cameras at these locations were identified, sample video clips were collected and used to define the underlying conflict scenarios, each of which represents an interaction between two types of road users (or movements). As illustrated in Figure 4-2, two conflicting scenarios can be identified in the video from a single camera: one between right-turning vehicles and pedestrians and the other between

through movements and left-turning vehicles. Furthermore, non-conflict movements could also be defined as additional scenarios of interest. Examples include jaywalking or bicycle red light violations.

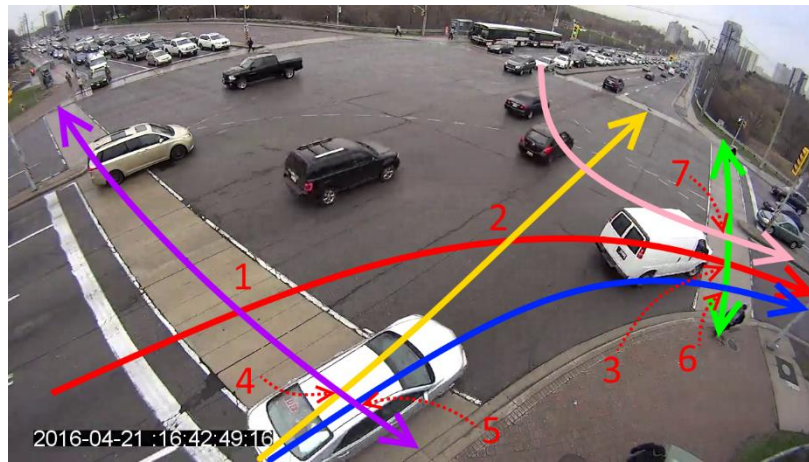


Figure 4-2 Definition of Movements and Conflicts

Once the video scenarios are identified, they must be calibrated to accurately mask the video and define the movements to be captured. As part of the process, the homographic parameters are generated based on camera coverage and aerial images of the intersection, as shown in Figure 4-3. This is done at the beginning of the implementation for each scenario to establish the coverage area parameters and should be revised periodically if necessary.

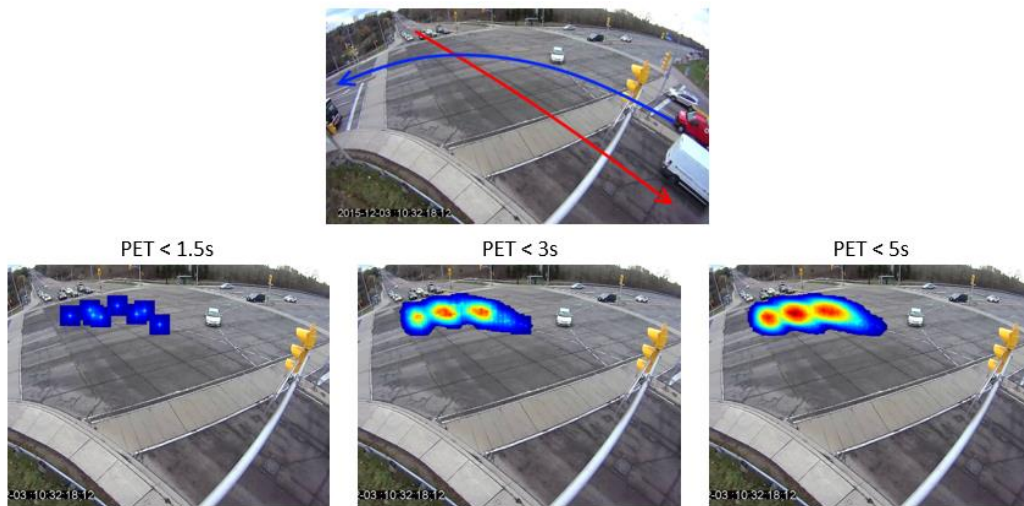


Figure 4-3 Outcomes from the Video Analytics Platform (Transoft Solutions, n.d.)

4.3 Data Collection and Processing

A total of 1118 hours of video data at the 10 selected intersections were collected, or 331, 237, 118, 432 hours for the NRTOR, LPI, BL and Control groups, respectively. All video data were then processed using a commercial software TrafXSAFE provided by Transoft to extract road user trajectories and conflict data to perform conflict analyses. Figure 4-4 shows the general architecture and flow of the data collection and processing.

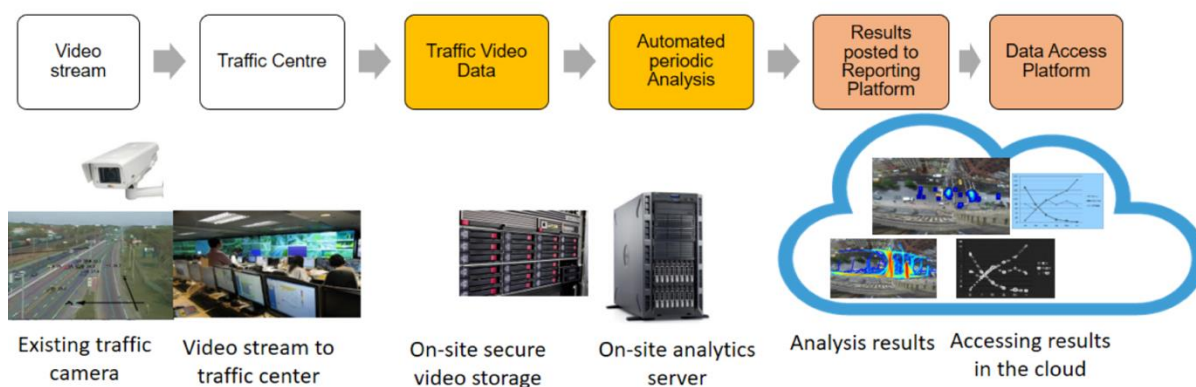


Figure 4-4 Video-based Safety Monitoring (Transoft Solutions, n.d.)

The video data processing procedure includes four steps: i) detecting and tracking the road users in the video; ii) classifying the detected road users into different types, i.e., vehicle, pedestrian, and cyclist; iii) selecting the interactive road user pairs with traffic conflict; and iv) calculating the surrogate measures of safety for extracted traffic conflict (Transoft Solutions, n.d.).

This software records data for each conflict, including conflicting traffic (e.g., vehicle type and pedestrian), post-encroachment time (PET), and speed. To prevent the cameras from missing some of the conflicts that occurred, the locations of the cameras must meet certain requirements, such as video resolution, field of view, and height/distance from the intersection. All traffic conflicts with PET less than 10s were recorded for this study. Figure 4-3 also shows the conflicts by severity for a conflicting movement and a scenario obtained from the data analysis platform.

A total of 101,364 traffic conflicts (i.e., PET<10s) at 10 selected intersections were recorded and used in this study, as shown in Table 4-1.

Table 4-1 Selected Intersections with their respective Countermeasures and Sample Size

Intersection	PX (ID)	Safety Countermeasure	Number of Observations (PET<10s)
Sheppard and Victoria Park	748	Control	12790
Finch and Yonge	128	Control	11623
Eglinton and Martin Grove	786	Control	3134
Danforth and Pape	345	Control	14196
Spadina and Dundas	277	NRTOR	8979
Bloor and Parkside	328	NRTOR	6804
Bloor and Avenue	86	NRTOR	15560
Don Mills and Lawrence	402	LPI	9207
Bloor and Jane	333	LPI	12000
King and Strachan	538	BL	11095

Table 4-2 shows the summary statistics of the collected variables for the evaluation of the effectiveness of traffic countermeasures. Since the selected intersections are all signalized, this paper mainly focuses on the traffic conflicts between right turning vehicles and pedestrians or cyclists on the crosswalk. The conflict

variables include conflict speed and the PET value for that conflict. Road user 1 represents the conflicting vehicle, which includes cars, work vans, buses, motorcycles and Road user 2 is a pedestrian or a bicycle. In addition, climate data including temperature, humidity, wind speed, visibility, and weather condition, were retrieved from Environment Canada's historical weather database, using the Toronto City Centre Climate Station. The climate condition at each hour was associated with each observation, this is used to determine if poor weather condition was a factor affecting the occurrence and severity of traffic conflicts. Furthermore, accessible pedestrian signals (APS) timings for each selected intersection were also collected and included in our study.

Table 4-2 Summary of Variables Descriptive Statistics

Conflict Variables				
Variables	Mean	Std. Dev	Min	Max
Road User 1 Conflict Speed (km/h)	18.297	8.675	0.00	59.99
Road User 2 Conflict Speed (km/h)	7.070	3.911	0.00	16.79
PET (s)	5.223	1.949	0	9.98
Traffic Volume Variables				
Variables	Mean	Std. Dev	Min	Max
Right Turn Volume per hour	31.628	24.488	1	146
Crosswalk Volume per hour	59.423	48.105	1	347
Climate Variables				
Variables	Mean	Std. Dev	Min	Max
Temperature (°C)	-1.658	3.654	-10.3	8.8
Humidity (%)	72.616	10.982	39	93
Wind Speed (km/h)	20.847	9.384	0	46
Visibility (km)	13.960	4.426	1.6	16.1
Variables	Dummy variable		Count	
Weather condition	Sunny		71,293	
	Snow		25,034	
	Rain		2,333	
	Fog		556	
	Haze		2,148	
Temporal Variables				
Variables	Dummy variable		Count	
Time-of-day	Daytime		81,253	
	Nighttime		20,111	

<i>Traffic Signal Variables</i>		
Variables	Dummy variable	Count
Accessible Pedestrian Signals (APS)	No APS Signal	25,553
	Pushbutton Actuated	56,119
	Fixed Time	19,692
<i>Countermeasures Variables</i>		
Variables	Dummy variable	Count
Countermeasure	Control	39697
	LPI	20698
	NRTOR	30075
	BL	10894

4.4 Identification of Traffic Conflicts of Varying Levels of Severity

Traffic conflicts are defined on the basis of two safety surrogate measures jointly, namely, road user conflict speed and post-encroachment time (PET), following the methodology by Navarro et al. (2022) and Zangenehpour et al. (2016). The threshold values for road user conflict speed were selected on the basis of a study by European Traffic Safety Council (ETSC), which states that the probability of pedestrian fatality is 5%, 45% and 85%, respectively when hit by a vehicle at 32km/h, 48km/h and 64km/h (ETSC, 2007).

Zangenehpour et al. (2016) classified conflicts into four categories based on PET:

- Very dangerous interaction: $PET \leq 1.5s$;
- Dangerous interaction: $1.5s < PET \leq 3s$;
- Mild interaction: $3s < PET \leq 5s$; and
- No interaction: $PET > 5s$.

As a result, four severity levels are defined for interactions or conflicts between vehicles and VRUs based on conflict speed and PET, as shown in Figure 4-5.

In this research, two traffic conflict severity classification methods (i.e., PET-based, PET and conflict speed-based) were also compared in the traffic countermeasures evaluation.

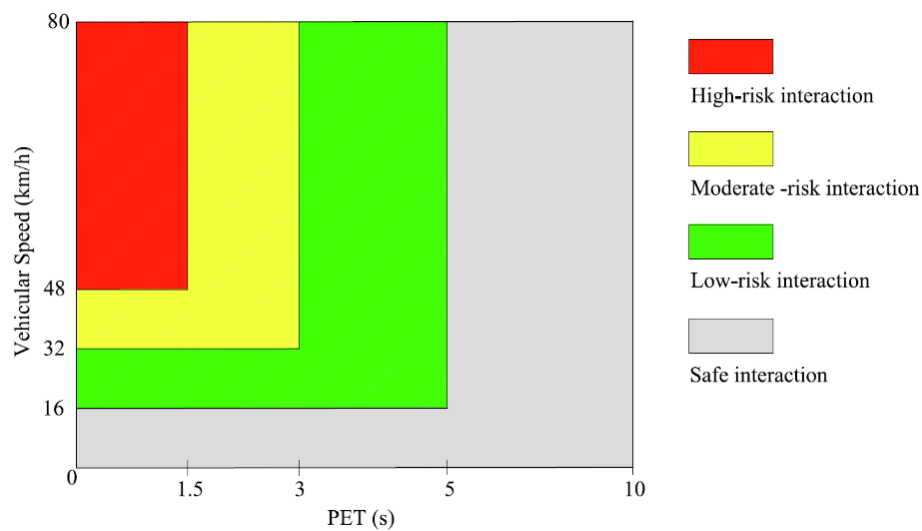


Figure 4-5 Classification of Severity Levels for Vehicle and Vulnerable Road User Interaction
(Zangenehpour et al., 2016)

4.5 Exploratory Data Analysis of Traffic Conflicts

An initial investigation of the conflict data collected by the video equipment was conducted and the PET values of individual traffic conflicts for each treatment group were analyzed. Figure 4-6 shows the cumulative frequency distribution curves and density histogram of the PET values under the four comparison cases. The mean PET values of traffic conflicts corresponding to the four traffic countermeasures of Control, LPI, NRTOR, and BL are 5.4s, 5.18s, 4.99s, and 5.3s, respectively.

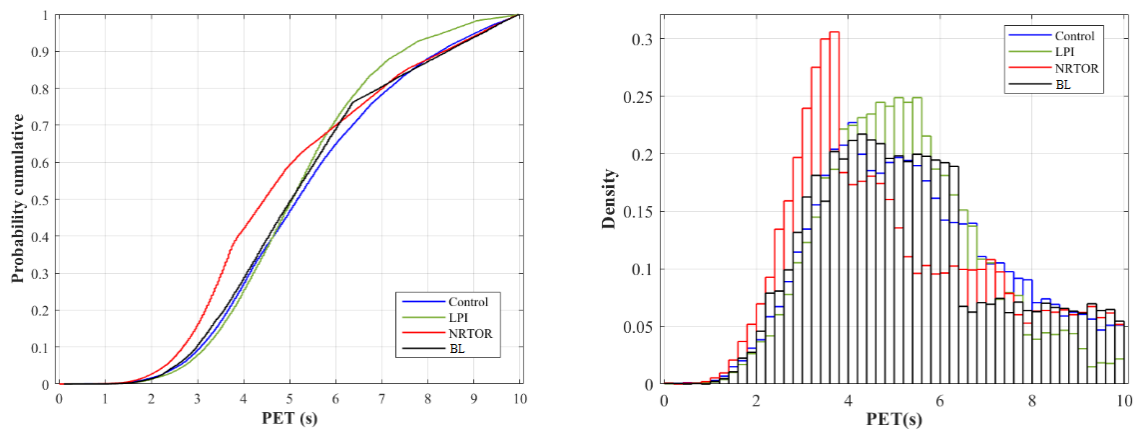


Figure 4-6 Cumulative Frequency Distribution Curves and Density Histogram of PET Values under Different Traffic Countermeasures

Table 4-3 shows the summary description of the processed video data for the selected intersections with different countermeasures. The frequency of traffic conflicts (i.e., total traffic conflicts per hour) at different severity levels is counted and provided in Table 4-3 and Figure 4-7. Furthermore, Kruskal-Wallis test, a non-parametric analysis method which has the advantage that the samples do not need to satisfy a specific distribution, was used to test the difference in traffic conflicts between different countermeasure groups (McKight & Najab, 2010). Table 4-4 provides the results of Kruskal-Wallis test of traffic conflict frequency of different risk levels under different traffic countermeasures.

Table 4-3 Summary Description of the Processed Video Data for Intersections with Implemented Traffic Countermeasures

	Intersection	Total Traffic Conflicts Per Hour						Turn speed (km/h)		Cross speed (km/h)		Volume (per hour)	
		High-risk	Moderate-risk	Low-risk	Very dangerous PET≤1.5s	Dangerous 1.5s<PET≤3s	Mild 3s<PET≤5s	Mean	Std	Mean	Std	Turn	Cross
Control	Sheppard/Victoria Park	0.05	1.79	33.60	0.47	7.78	36.07	22.70	10.01	7.76	4.55	460.4	289.4
	Yonge/Finch	0.06	0.82	32.63	0.49	13.29	48.88	18.23	8.82	7.22	4.27	226.6	751.2
	Danforth/Pape	0.01	0.64	24.40	0.32	9.34	48.18	15.75	8.88	6.28	3.15	201.3	655.4
	Eglinton/Martin Grove	0	0.13	7.49	0.21	1.95	8.17	19.80	6.48	7.28	5.14	320.2	60.8
LPI	Lawrence/Don Mills	0	0.87	26.24	0.32	6.28	27.02	22.21	9.07	7.08	4.19	426.1	255.4
	Bloor/Jane	0	0.23	28.81	0.16	6.82	43.04	16.48	6.46	6.40	2.54	385.9	243.9
NRTOR	Spadina/Dundas	0.02	0.30	21.28	0.23	9.98	39.29	15.10	7.02	7.11	3.77	145.8	838.4
	Bloor/Parkside	0	0.22	23.63	0.19	5.02	28.66	19.26	6.79	6.00	3.17	239.6	177.4
	Bloor/Avenue	0.04	2.15	50.39	1.14	27.69	52.97	18.83	9.78	8.00	4.40	286.1	743.7
BL	King/Strachan	0	0.14	26.51	0.27	9.48	37.51	16.57	5.85	7.23	3.63	138.3	385.4
Total	Control	0.03	0.87	25.13	0.38	8.30	36.22	18.79	9.49	7.05	4.12	302.1	439.2
	LPI	0	0.51	27.70	0.23	6.58	36.09	18.91	8.18	6.69	3.36	403.6	249.0
	NRTOR	0.02	0.86	31.09	0.50	14.10	40.50	17.85	8.62	7.30	4.05	219.8	626.3
	BL	0	0.14	26.51	0.27	9.48	37.51	16.57	5.85	7.23	3.63	138.3	385.4

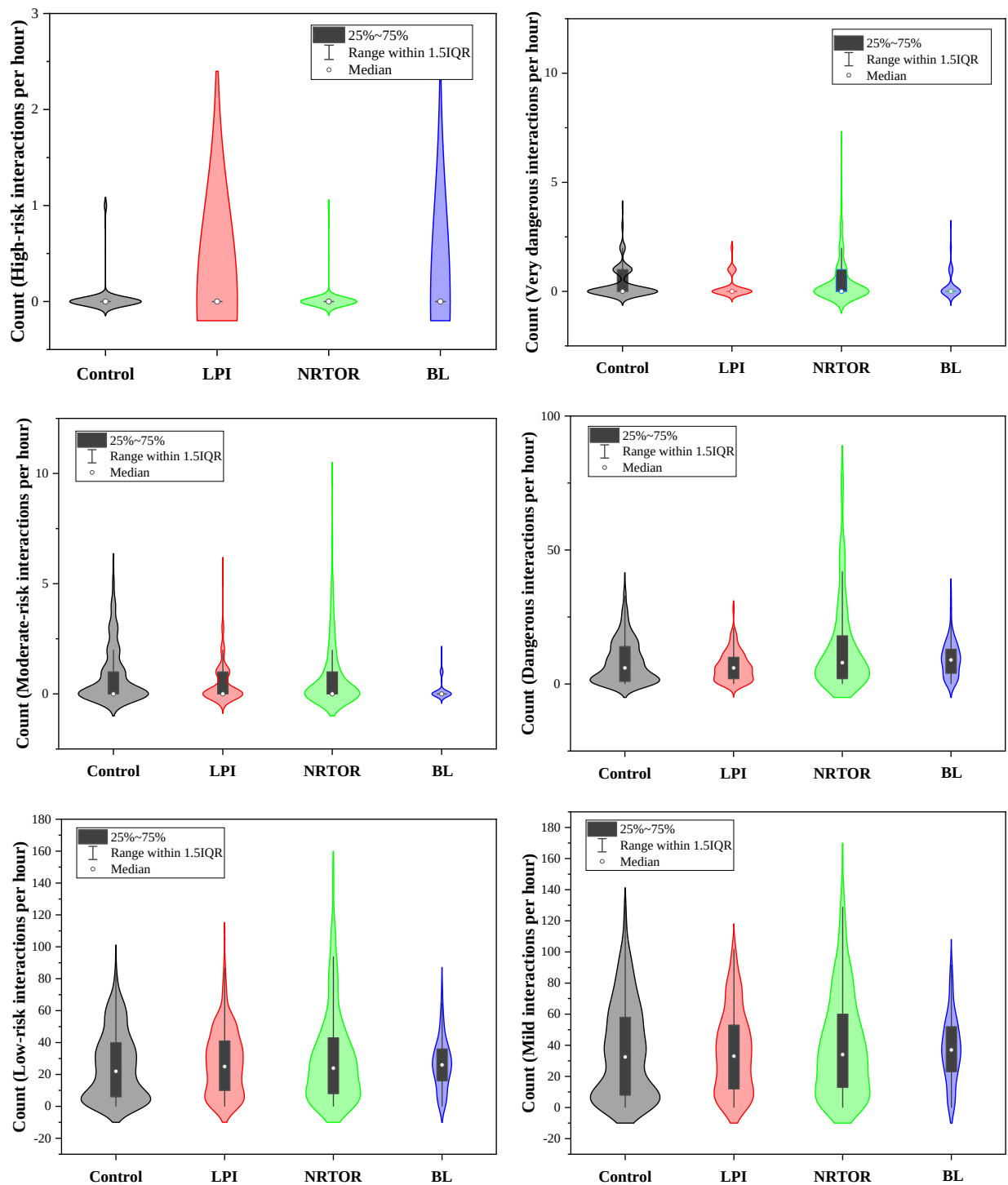


Figure 4-7 Box-Violin Plot of Traffic Conflict Frequency with Different Risk Levels under Different Traffic Countermeasures

Table 4-4 Results of Kruskal-Wallis Test of Traffic Conflict Frequency with Different Risk Levels under Different Traffic Countermeasures

Group1	Group2	High-risk Interaction		Moderate-risk Interaction		Low-risk Interaction	
		χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>
Control	LPI	6.94	0.008***	8.83	0.003***	3.34	0.068*
Control	NRTOR	1.48	0.224	1.83	0.176	3.53	0.060*
Control	BL	3.38	0.066*	36.1	<0.001***	2.1	0.147
LPI	NRTOR	3.64	0.056*	2.20	0.138	0.02	0.898
LPI	BL	--	--	17.19	<0.001***	0.03	0.871
NRTOR	BL	1.77	0.183	24.39	<0.001***	0.02	0.875
Total		10.33	0.016**	39.15	<0.001***	5.43	0.143
Group1	Group2	Very dangerous Interaction		Dangerous Interaction		Mild Interaction	
		χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>
Control	LPI	5.44	0.020**	1.79	0.181	0.44	0.505
Control	NRTOR	0	0.987	11.66	<0.001***	2.27	0.132
Control	BL	1.91	0.167	6.03	0.014**	1.66	0.197
LPI	NRTOR	4.48	0.034**	15.57	<0.001***	0.53	0.466
LPI	BL	0.14	0.704	16.09	<0.001***	0.88	0.348
NRTOR	BL	1.62	0.202	0	0.969	0.09	0.758
Total		7.11	0.068*	24.91	<0.001***	3.20	0.361

Note: *, **, *** refer to significance at 10%, 5%, 1% level.

4.6 Safety Effectiveness of Alternative Countermeasures

To evaluate the effectiveness of the three countermeasures (i.e., LPI, NRTOR, and BL), separate conflict frequency models were first developed. The hourly traffic conflict frequency of different severity levels was chosen as the measure of safety level of the selected intersections. As discussed in the literature review, various statistical methods have been applied to model the traffic conflict frequency over the past several years, such as Poisson regression model and negative binomial regression model (Shahdah et al. (2015); Ma et al., (2017)) However, these traditional Poisson and negative binomial regression models assume that the traffic conflicts at intersections are independent, thus ignoring the heterogeneity characteristics of different intersections. In this research, a mixed-effects negative binomial regression model is introduced, which is known to be effective in modelling varied effects between intersections with repeated measurements (Stapleton et al., 2019).

The mixed-effects negative binomial model incorporates a random intersection-specific effects term into the relationship between the expected numbers of traffic conflicts at a given time period t at a given intersection i , $\tilde{\mu}_{it}$, and its explanatory variables, as shown below:

$$\tilde{\mu}_{it} = \mu_{it} \delta_i \quad (12)$$

where δ_i represents the random location-specific effects term.

Similar to the negative binomial model, $\tilde{\mu}_{it}$ is assumed to be an exponential function of some covariates:

$$\tilde{\mu}_{it} = \mu_{it} \delta_i = \exp(\beta X_{it} + \mu_i) \quad (13)$$

where β is the estimated coefficient vector of the applied model; $\exp(\mu_i)$ is a gamma-distributed term with mean 1 and variance k , where k is the overdispersion parameter in the negative binomial model.

Therefore, the probability density function of the mixed-effects negative binomial regression model can be represented as:

$$f(y_{it}|X_{it}) = \frac{\Gamma(a+b)\Gamma(a+\sum_T \mu_{it})\Gamma(b+\sum_T \mu_{it})}{\Gamma(a)\Gamma(b)\Gamma(a+b+\sum_T \mu_{it}+\sum_T y_{it})} \prod_T \frac{\Gamma(\mu_{it}+y_{it})}{\Gamma(\mu_{it})+\Gamma(y_{it}+1)} \quad (14)$$

where y_{it} is the number of traffic conflicts at intersection i during given period t ; a and b are distributional parameters, which can be estimated using the maximum likelihood method.

The likelihood ratio (LR) test was applied in this study to compare the performance of the mix-effects negative binomial model and the traditional fixed-effects negative binomial model. The incidence-rate ratio (IRR) for each explanatory variable was also calculated to represent the effect of changes in that variable on the model results. All coefficients in the proposed mixed-effect negative binomial model were estimated using Stata/MP 16.0.

Table 4-5 and Table 4-6 present the modelling results for conflict frequencies of different severity classes defined by the two approaches introduced previously, namely PET-based and PET with conflict speed-based. The mixed-effects model performed better than the fixed-effects model in most comparison groups. All insignificant variables were eliminated in the final model.

When PET and conflict speed are combined to classify conflict severity, as shown in Table 4-5, it can be found that right-turn traffic volume and crosswalk volume have a positive impact on the frequency of traffic conflict at any severity level. Furthermore, the results of high-risk interaction frequency modeling showed that only the NRTOR group significantly reduced the frequency of high-risk interactions between vehicle and VRUs, and the high-risk interaction frequency of the NRTOR group decreased by 83.4% compared with the control group. As for the results of moderate-risk interaction frequency modeling, it can be found that the frequency of moderate-risk conflicts is significantly reduced at night compared to daytime, the same finding in the results of low-risk interaction frequency modeling. In addition, compared with the control group, the three traffic countermeasures, LPI, NRTOR and BL, all significantly reduced the moderate-risk interaction frequency by 45.7%, 51.9% and 60.6%, respectively. As shown in the Control-NRTOR group modeling results, humidity was positively correlated with moderate-risk interaction frequency. This could be explained by wet or damp road surfaces that reduce traction and increase stopping distances or high humidity leading to fog which reduces visibility. Meanwhile, both pushbutton actuated and fixed-time APS can significantly reduce the moderate-risk interaction frequency compared to no APS. APS is implemented in areas after the City of Toronto identifies a need (i.e.: due to the presence of people with visual impairments or people with other disabilities). Because of this, other enhanced accessibility features could have been implemented that also improve safety and awareness for pedestrians such as allocating more time for pedestrian crossing phases, curb bump outs, or tactile warning plates. For the results of low-risk interaction frequency modeling, compared with the control group, the frequency of low-risk interaction frequency in the LPI and NRTOR groups significantly increased by 43.6% and 55.9%.

Meanwhile, environmental factors such as humidity, wind speed, and visibility all negatively affect the frequency of low-risk interactions. This could be due to the existence of fewer pedestrians during inclement weather. As shown in the results of Control - BL group modeling, the frequency of low-risk interactions was lower in the APS-equipped intersections than in no-APS intersections.

Table 4-6 shows the modeling results of interaction frequency when only the PET value is used for interaction severity classification. Similar to the findings in Table 4-5, both right-turn traffic volume and crosswalk volume positively affect the frequency of various types of traffic interactions between vehicle and VRUs. Compared with the control group, the LPI and BL groups significantly reduced the frequency of very dangerous interactions by 37.3% and 52.9%, respectively. As shown in the results of Control - LPI group, the frequency of very dangerous interactions increased by 513.3% in rainy days compared to sunny days. For the results shown in dangerous interaction frequency modeling, compared with the control group, the NRTOR and BL groups significantly increased the frequency of dangerous interactions by 79.6% and 177.9%, respectively. In addition, the results of three comparison modeling showed that, compared with the daytime, the frequency of dangerous interaction was reduced by 32.4%, 35.5% and 27.3% at nighttime, respectively. This could be due to there being fewer pedestrians present at intersections in the evening or in the dark. Lastly, as shown in the results of mild interaction frequency modeling, compared with the control group, the frequency of mild interactions at the intersections with NRTOR and BL countermeasures significantly increased by 78.2% and 176.3%, respectively. At the same time, the three comparison modeling results showed that, compared with daytime, the frequency of mild interactions at nighttime was reduced by 32.2%, 35.1% and 26.4%, respectively. Meanwhile, as shown in Control - LPI group modeling results, intersections with fixed-time APS devices have a higher frequency of mild interactions than intersections without APS.

Table 4-5 Results of Traffic Conflict (Classified by PET Value and Conflict Speed) Frequency Modeling

Type of Countermeasure	Control - LPI		Control - NRTOR		Control - BL	
Explanatory Variables	β	IRR	β	IRR	β	IRR
High-risk Interaction Frequency Modeling						
Fixed Effects						
Constant	- 9.256***	0.000	- 7.256***	0.001	- 9.270***	0.000
Right turn volume	0.006**	1.006	0.006***	1.006	0.008***	1.008
Crosswalk volume	0.004***	1.004	0.003***	1.003	0.004***	1.004
Countermeasure (Baseline: Control)	--	--	-1.796**	0.166	--	--
Random Effects (Group name: Intersection)						
Variance (Constant)	0.765		--		--	
Log likelihood	-46.856		-68.467		-44.050	
LR test vs. negative binomial model	χ²=1.83, P=0.08		--		--	
Moderate-risk Interaction Frequency Modeling						
Fixed Effects						
Constant	- 2.454***	0.086	-1.348*	0.260	- 2.859***	0.057
Right turn volume	0.004***	1.004	0.003***	1.003	0.003***	1.003
Crosswalk volume	0.002***	1.002	0.001***	1.001	0.002***	1.002
Time-of-day - Nighttime (Baseline: Daytime)	- 0.557***	0.573	- 0.486***	0.615	- 0.690***	0.502
Humidity (%)	--	--	0.011**	1.011	0.010*	1.010
APS - Pushbutton Actuated (Baseline: No APS)	--	--	-1.214*	0.297	--	--
APS - Fixed time (Baseline: No APS)	--	--	-1.201*	0.301	--	--

Countermeasure (Baseline: Control)	- 0.611***	0.543	-0.731**	0.481	-0.931*	0.394
Random Effects (Group name: Intersection)						
Variance (Constant)	0.413		0.271		0.338	
Log likelihood	-583.031		-702.091		-443.303	
LR test vs. negative binomial model	$\chi^2=76.99$, P<0.01		$\chi^2=48.14$, P<0.01		$\chi^2=47.02$, P<0.01	
Low-risk Interaction Frequency Modeling						
Fixed Effects						
Constant	1.951***	7.033	1.597***	4.937	2.677***	14.535
Right turn volume	0.003***	1.003	0.004***	1.004	0.004***	1.004
Crosswalk volume	0.001***	1.001	0.001***	1.001	0.001***	1.001
Time-of-day - Nighttime (Baseline: Daytime)	- 0.578***	0.561	- 0.658***	0.518	- 0.477***	0.621
Humidity (%)	- 0.008***	0.992	- 0.005***	0.995	- 0.006***	0.994
Wind speed (km/h)	-0.004**	0.996	--	--	-0.003*	0.997
Visibility (km)	-0.006**	0.994	--	--	--	--
APS - Pushbutton Actuated (Baseline: No APS)	--	--	--	--	- 1.145***	0.318
APS - Fixed time (Baseline: No APS)	--	--	--	--	-0.684*	0.504
Countermeasure (Baseline: Control)	0.362***	1.436	0.444**	1.559	--	--
Random Effects (Group name: Intersection)						
Variance (Constant)	0.074		0.245		0.081	
Log likelihood	-2158.765		-2520.843		-1729.532	
LR test vs. negative binomial model	$\chi^2=152.44$, P<0.01		$\chi^2=375.86$, P<0.01		$\chi^2=125.51$, P<0.01	

Note: *, **, *** refer to significance at 10%, 5%, 1% level.

Table 4-6 Results of Traffic Conflict (Classified by PET Value) Frequency Modeling

Type of Countermeasure	Control - LPI		Control - NRTOR		Control - BL	
Explanatory Variables	β	IRR	β	IRR	β	IRR
Very dangerous Interaction Frequency Modeling						
Fixed Effects						
Constant	-3.577***	0.028	-3.501***	0.030	-3.791***	0.023
Right turn volume	0.004**	1.003	0.005***	1.005	0.005***	1.004
Crosswalk volume	0.002***	1.002	0.001***	1.001	0.002***	1.002
Weather - Rain (Baseline: Snow)	1.813***	6.133	--	--	--	--
Countermeasure (Baseline: Control)	-0.466**	0.627	--	--	-0.752***	0.471
Random Effects (Group name: Intersection)						
Variance (Constant)	--		0.065		--	
Log likelihood	-403.679		-523.400		-347.003	
LR test vs. negative binomial model	--		χ²=4.22, P=0.02		--	
Dangerous Interaction Frequency Modeling						
Fixed Effects						
Constant	0.173*	1.189	--	--	--	--
Right turn volume	0.003***	1.003	0.004***	1.004	0.003***	1.003
Crosswalk volume	0.002***	1.002	0.001***	1.001	0.002***	1.002
Time-of-day - Nighttime (Baseline: Daytime)	-0.391***	0.676	-0.439***	0.645	-0.318***	0.727
Countermeasure (Baseline: Control)	--	--	0.585*	1.796	1.022***	2.779
Random Effects (Group name: Intersection)						
Variance (Constant)	0.106		0.354		0.154	
Log likelihood	-1519.009		-1859.096		-1252.479	
LR test vs. negative binomial model	χ²=86.26, P<0.01		χ²=281.19, P<0.01		χ²=65.83, P<0.01	
Mild Interaction Frequency Modeling						
Fixed Effects						

<i>Constant</i>	0.107*	1.113	--	--	--	--
Right turn volume	0.003***	1.003	0.004***	1.004	0.003***	1.003
Crosswalk volume	0.002***	1.002	0.001***	1.001	0.002***	1.002
Time-of-day - Nighttime (Baseline: Daytime)	-0.388***	0.678	-0.432***	0.649	-0.307***	0.736
APS - Fixed time (Baseline: No APS)	0.466***	1.593	--	--	--	--
Countermeasure (Baseline: Control)	--	--	0.578*	1.782	1.016***	2.763
<i>Random Effects (Group name: Intersection)</i>						
Variance (Constant)	0.063		0.343		0.148	
Log likelihood	-1544.233		-1885.253		-1272.367	
LR test vs. negative binomial model	$\chi^2=59.81$, P<0.01		$\chi^2=275.49$, P<0.01		$\chi^2=60.97$, P<0.01	

Note: *, **, *** refer to significance at 10%, 5%, 1% level.

Chapter 5

Conclusions

5.1 Summary of Main Findings

Traffic safety is a growing issue that many cities and municipalities face. Conflict analysis continues to show great promise for diagnosing safety issues and determining the safety impacts of treatments.

The objective of this research was to present a framework for determining the minimum sample size required for conflict-based safety studies and to evaluate the effectiveness of Vision Zero safety countermeasures using traffic conflict data collected by video cameras in the City of Toronto.

Conflict analysis requires shorter observation periods than collision analysis, but currently lacks the same statistical rigor found in the fields of health sciences and medicine. Power analysis is robust, well accepted in the fields of medical sciences and reliability studies, and highly generalizable to many statistical tests, as summarized in the proposed framework, and demonstrated through three case studies. Following key conclusions can be drawn:

- This study outlines a comprehensive framework that encompasses common traffic conflict studies and their associated statistical tests. By following this framework and the conventional values of $\alpha = 0.05$ and $1-\beta = 0.8$, experiments are protected from Type I (false positive) and Type II (false negative) errors.
- The proposed framework can facilitate the researchers to conduct a priori analysis, that is, to determine the sample size required based on the effect size estimated before data collection begins; and a post hoc power analysis, that is, to determine the power of your study based on the sample size used and effect size calculated.
- The validity of the power analysis should be checked after the analysis is conducted using methods such as p-values and by constructing confidence intervals. The adoption of this framework before the project starts will protect investigators from Type I and Type II errors and save them the effort from futile endeavors or misinformation.

Video-based trajectory data from ten selected intersections in the City of Toronto were used to determine two surrogate traffic safety indicators including post-encroachment time (PET) and conflict speed. Three widely used intersection safety countermeasures targeting vulnerable road users (VRUs), namely, Leading Pedestrian Interval (LPI), No Right Turn On Red (NRTOR), and installation of dedicated Bicycle Lane (BL), were evaluated using case-control and cross-sectional approaches. A mixed-effect negative binomial regression model was applied to quantify the impact of different traffic

countermeasures on the traffic conflict frequency of the intersection. Three key conclusions can be drawn from our investigation as follows:

- Based on both PET-based or conflict speed-based conflict severity analyses, the frequency of the highest risk interactions between vehicles and VRUs at intersections with the three traffic countermeasures (i.e., LPI, NRTOR, and BL) was consistently lower than those at the intersections without these safety countermeasures.
- The mixed-effect negative binomial regression model worked well in quantifying the impact of the traffic safety countermeasures on intersection safety. Compared with the control group, the frequency of high-risk interactions at intersections with NRTOR countermeasures decreased by 83.4%. Meanwhile, the LPI group and the BL group significantly reduced the frequency of critically dangerous interactions by 36.3% and 52.9%, respectively. In addition, the three traffic safety countermeasures also reduced the frequency of moderate-risk interactions by 45.7%, 51.9% and 60.6%, respectively.
- Among different explanatory variables, traffic volume plays a key role in influencing the frequency of traffic conflicts at any severity levels. Furthermore, time-of-day and some meteorological variables, such as weather, humidity, wind speed, and visibility, also significantly affect the frequency of certain types of traffic conflicts. In addition, different types of accessible pedestrian signals (APS) also have a significant impact on the safety performance of intersections.

5.2 Recommendations and Future Research

The proposed sample size framework could become a valuable tool for planning various conflict-based safety studies and is recommended to be used in future conflict-based traffic safety analyses. Through this additional step, researchers will be cognizant of possible limitations and can consider collecting a larger sample to protect themselves from Type I and II errors. Future work is best aimed at employing the proposed framework and publishing validated results. The findings from this research could help road safety engineers and decision makers make better informed decisions on their road safety initiatives and projects.

The research on applying conflict-based analysis for evaluating the effectiveness of safety countermeasures has provided a theoretical basis for conducting a data-driven safety benefit assessment

of specific safety countermeasures and projects. However, several limitations still exist in this study. Due to the lack of a sufficient long history of traffic crash data at the selected intersections, the traffic safety indicators applied in this paper have not been validated. Due to the limitation of the data collection period, the sample of high-risk interactions of some treatment groups is still limited and the generalizability of the conclusions obtained in this study still needs to be further verified. In future work, real-world crash data should be collected and used to validate the effectiveness of the traffic safety indicators. In addition, safety diagnosis of different types of intersections and different conflict scenarios should also be carried out to form a more general knowledge base on the effectiveness of various traffic safety countermeasures. Lastly, this study mainly focuses on the traffic conflict between right-turning vehicles and pedestrians in crosswalks without considering the possible conflicts between left-turning vehicles and pedestrians. The impact of left-turning traffic volume on the efficiency of LPI treatment is also worthy of further exploration.

Bibliography

- AASHTO. (2010). *Highway Safety Manual* (1st ed.). Washington DC: AASHTO.
- Alhajyaseen, W. K. M. (2014). The Development of Conflict Index for the Safety Assessment of Intersections The Development of Conflict Index for the Safety Assessment of Intersections Considering Crash Probability and Severity. *Procedia - Procedia Computer Science*, 32(October), 364–371. <https://doi.org/10.1016/j.procs.2014.05.436>
- Allen, B. L., Shin, B. T., & Cooper, P. J. (1977). Analysis of Traffic Conflicts and Collisions. *Transportation Research Record*, 667, 67–74.
- Alluri, P., Saha, D., & Gan, A. (2016). Minimum sample sizes for estimating reliable Highway Safety Manual (HSM) calibration factors. *Journal of Transportation Safety & Security*, 8(1), 56-74.
- Alonso, W. (1967). *The quality of data and the choice and design of predictive models*.
- American Association of State Highway and Transportation Officials. (2011). *A Policy on Geometric Design of Highways and Streets* (6th ed.). Washington, DC: American Association of State Highway and Transportation Officials.
- Amundsen, F. H., & Hydén, C. (1977). Proceedings: First workshop on Traffic Conflicts. In *Proceedings: First Workshop on TRAFFIC CONFLICTS*. Oslo, Norway.
- Autey, J., Sayed, T., & Zaki, M. H. (2012). Safety evaluation of right-turn smart channels using automated traffic conflict analysis. *Accident Analysis and Prevention*, 45, 120–130. <https://doi.org/10.1016/j.aap.2011.11.015>
- Breene, J., Meleckidzedek, K., McNerney, R., Sukhai, A., Toroyan, T., Ward, D. (2018). *Global Status Report on Road Safety 2018*. Geneva.
- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., De Rosario, H. (2018). Package “pwr.” CRAN. Retrieved from <https://cran.r-project.org/web/packages/pwr/pwr.pdf>
- Chin, H. C., & Quek, S. T. (1997). Measurement of traffic conflicts. *Safety Science*, 26(3), 169–185.

[https://doi.org/10.1016/S0925-7535\(97\)00041-6](https://doi.org/10.1016/S0925-7535(97)00041-6)

- City of Toronto. (2017). *2017-2021 | Toronto's road safety plan*. Toronto.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second). Lawrence Erlbaum Associates, Publishers.
- Cohen, J. (1992). A Power Primer. *Quantitative Methods in Psychology*, 112(1), 155–159.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.
- El-Basyouny, K., & Sayed, T. (2013). Safety performance functions using traffic conflicts. *Safety Science*, 51 (1), 160-164.
- Ellis, R., & Van Houten, R. (2009). Reduction of Pedestrian Fatalities, Injuries, Conflicts, and Other Surrogate Measures in Miami–Dade, Florida: Results of Large-Scale FHWA Project. *Transportation Research Record*, 2140 (1), 55-62.
- European Transport Safety Council. (2007). European Road Safety Observatory (2006) Speeding.
- Fu, T. (2018). *A novel approach to investigate pedestrian safety in non-signalized crosswalk environments and related treatments using trajectory data*. McGill University.
- Fu, T., Miranda-Moreno, L., & Saunier, N. (2016). Pedestrian Crosswalk Safety at Nonsignalized Crossings During Nighttime: Use of Thermal Video Data and Surrogate measures of safety. *Transportation Research Record: Journal of the Transportation Research Board*, 2586(1), 90–99. <https://doi.org/10.3141/2586-10>
- Gettman, D., Pu, L., Sayed, T., & Shelby, S. (2008). *Surrogate Safety Assessment Model and Validation : Final Report*. Federal Highway Association.
- Glauz, W. D., Bauer, K. M., & Migletz, D. J. (1985). Expected traffic conflict rates and their use in predicting accidents. *Transportation Research Record*, 3(1026), 1–12. Retrieved from <https://trid.trb.org/view/271535>
- Guo, Y., Sayed, T., & Zheng, L. (2020). A hierarchical Bayesian peak over threshold approach for conflict-based before-after safety evaluation of leading pedestrian intervals. *Accident Analysis &*

Prevention, 147, 105772.

Hakkert, A. S., Glauz, W. D., Hauer, E., Lightburn, A., Kraay, J. H., & Older, J. (1987). Guidelines for the Conduct of Traffic Conflict Studies.

Hauer, E., & Garder, P. (1986). Research into the validity of the traffic conflicts technique. *Accident Analysis and Prevention*, 18(6), 471–481. [https://doi.org/10.1016/0001-4575\(86\)90020-5](https://doi.org/10.1016/0001-4575(86)90020-5)

Hauer, E. (1996). The role of time-to-collision in the analysis of traffic safety. *Traffic Engineering & Control*, 37(5), 224-231.

Hauer, E. (1997). *Observational before-after studies in road safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Pergamon.

Hoening, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *American Statistician*, 55(1), 19–24. <https://doi.org/10.1198/000313001300339897>

Hyden, C. (1996). Traffic conflict techniques - state of the art.

Ismail, K., Sayed, T., & Saunier, N. (2010). Automated Analysis of Pedestrian-Vehicle. *Transportation Research Record: Journal of the Transportation Research Board*, 2198(2198), 52–64. <https://doi.org/10.3141/2198-07>

Jackson, S., Miranda-Moreno, L., St-Aubin, P., & Saunier, N. (2013). Flexible, Mobile Video Camera System and Open Source Video Analysis Software for Road Safety and Behavioral Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2365(1), 90–98. <https://doi.org/10.3141/2365-12>

Kassim, A., Ismail, K., & Hassan, Y. (2014). Automated measuring of cyclist–motor vehicle post encroachment time at signalized intersections. *Canadian Journal of Civil Engineering*, 41(7), 605-614.

Kay, J., Gates, T. J., Savolainen, P. T., & Mahmud, M. S. (2022). Safety performance of unsignalized median U-turn intersections. *Transportation Research Record*.

Laureshyn, A., Jonsson, C., De Ceunynck, T., Svensson, A., & de Goede, M. (2016). *Review of*

current study methods for VRU safety. Warsaw.

- Laureshyn, A., & Varhelyi, A. (2018). The Swedish Traffic Conflict Technique-Observer's Manual. *Lund University*, (April).
- Lee, C., So, J., & Ma, J. (2018). Evaluation of countermeasures for red light running by traffic simulator-based surrogate safety measures. *Traffic Injury Prevention*, 19(1), 1-8.
- Lord, D. (1996). Analysis of Pedestrian Conflicts with Left-Turning Traffic. *Transportation Research Record: Journal of the Transportation Research Board*, 1538, 61–67.
<https://doi.org/10.3141/1538-08>
- Lord, D. (2002). Synthesis on the Safety of Right Turn on Red in the United States and Canada. *Transportation Research Board 82nd Annual Meeting*.
- Lu, P., Liu, J., & Koestler, D. (2017). Package 'pwr2.' CRAN. Retrieved from <https://cran.r-project.org/web/packages/pwr2/pwr2.pdf>
- Ma, Z., Zhang, H., Chien, S. I. J., Wang, J., & Dong, C. (2017). Predicting expressway crash frequency using a random effect negative binomial model: A case study in China. *Accident Analysis & Prevention*, 98, 214-222.
- McKight, P. E., & Najab, J. (2010). Kruskal-Wallis test. *The Corsini Encyclopedia of Psychology*, ' 1-1.
- Navarro, B., Miranda-Moreno, L., Saunier, N., Labbe, A., & Fu, T. (2022). Do stop-signs improve the safety for all road users? A before-after study of stop-controlled intersections using video-based trajectories and surrogate measures of safety. *Accident Analysis & Prevention*, 167, 106563.
- Paul, M., Ghosh, I., & Haque, M. M. (2022). The effects of green signal countdown timer and retiming of signal intervals on dilemma zone related crash risk at signalized intersections under heterogeneous traffic conditions. *Safety Science*, 154, 105862.
- Peesapati, L. N. (2014). *METHODS TO EVALUATE THE EFFECTIVENESS OF CERTAIN SURROGATE MEASURES TO ASSESS SAFETY OF OPPOSING LEFT-TURN INTERACTIONS METHODS TO EVALUATE THE EFFECTIVENESS OF CERTAIN LEFT-*

TURN INTERACTIONS. Georgia Institute of Technology.

- Rajeswaran, T., Persaud, B., & Anarkooli, A. J. (2023). Exploration of some key issues in developing and applying crash–conflict models for signalized intersections. *Transportation Research Record*, 03611981221148483.
- Saunier, N., & Sayed, T. (2006). A feature-based tracking algorithm for vehicles in intersections. *Third Canadian Conference on Computer and Robot Vision, CRV 2006*, 2006. <https://doi.org/10.1109/CRV.2006.3>
- Sasidharan, L., & Donnell, E. T. (2013). Application of propensity scores and potential outcomes to estimate effectiveness of traffic safety countermeasures: Exploratory analysis using intersection lighting data. *Accident Analysis & Prevention*, 50, 539-553.
- Sayed, T., Brown, G., & Navin, F. (1994). Simulation of traffic conflicts at unsignalized intersections with TSC-Sim. *Accident Analysis and Prevention*, 26(5), 593–607. [https://doi.org/10.1016/0001-4575\(94\)90021-3](https://doi.org/10.1016/0001-4575(94)90021-3)
- Sayed, T., Zaki, M. H., & Autey, J. (2013). Automated safety diagnosis of vehicle-bicycle interactions using computer vision analysis. *Safety Science*, 59, 163–172. <https://doi.org/10.1016/j.ssci.2013.05.009>
- Sayed, T., & Zein, S. (1999). Traffic conflict standards for intersections. *Transportation Planning and Technology*, 22(4), 309–323. <https://doi.org/10.1080/03081069908717634>
- Shahdah, U., Saccomanno, F., & Persaud, B. (2015). Application of traffic microsimulation for evaluating safety performance of urban signalized intersections. *Transportation Research Part C: Emerging Technologies*, 60, 96-104.
- Sheskin, D. J. (2000). *Handbook of parametric and non-parametric statistical procedures* (Second). Chapman & Hall / CRC. Retrieved from www.crcpress.com
- Songchitruksa, P., & Tarko, A. P. (2006). Practical Method for Estimating Frequency of Right-Angle Collisions at Traffic Signals, *Transportation Research Record*, 1953(1), 89–97.
- Souleyrette, R., & Hochstein, J. (2012). *Development of a Conflict Analysis Methodology Using*

- SSAM. Washington, DC: Federal Highway Administration.
- Stapleton, S. Y., Gates, T. J., Avelar, R., Geedipally, S. R., & Saedi, R. (2019). Safety performance functions for low-volume rural stop-controlled intersections. *Transportation Research Record*, 2673(5), 660-669.
- St-Aubin, P., Saunier, N., Miranda-Moreno, L.F., Ismail, K. (2013). Detailed Driver Behavior Analysis and Trajectory Interpretation at Roundabouts using Computer Vision Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2389(1), 65–77. <https://doi.org/10.3141/2389-07>
- St-Aubin, P., Miranda-Moreno, L., & Saunier, N. (2013). An automated surrogate safety analysis at protected highway ramps using cross-sectional and before-after video data. *Transportation Research Part C: Emerging Technologies*, 36, 284–295. <https://doi.org/10.1016/j.trc.2013.08.015>
- St-Aubin, P., Saunier, N., & Miranda-Moreno, L. (2015a). Comparison of Various Time-to-Collision Prediction and Aggregation Methods for Surrogate Safety Analysis. *Transportation Research Board (TRB) 94th Annual Meeting*, 1(514).
- St-Aubin, P., Saunier, N., & Miranda-Moreno, L. (2015b). Large-scale automated proactive road safety analysis using video data. *Transportation Research Part C: Emerging Technologies*, 58, 363–379. <https://doi.org/10.1016/j.trc.2015.04.007>
- Tarko, A., Davis, G., Saunier, N., Sayed, T., & Washington, S. (2009). White Paper Surrogate Measures of Safety ANB20(3) Subcommittee on Surrogate Measures of Safety. *ANB20(3) Subcommittee on Surrogate Measures of Safety ANB20 Committee on Safety Data Evaluation and Analysis*, 20(3), 1–13.
- Transoft Solutions. (n.d.). *TRAFXSafety*. Retrieved from [\[https://safety.transoftsolutions.com/trafxsafe/\]](https://safety.transoftsolutions.com/trafxsafe/)(<https://safety.transoftsolutions.com/trafxsafe/>).
- van Haperen, W. (HU). (2016). *Review of current study methods for VRU safety: Appendix 5 - Systematic literature review: Behavioral observations*. Warsaw.
- Walters, S. J. (2009). Consultants’ forum: should post hoc sample size calculations be done?

Pharmaceutical Statistics, 8, 163–169. <https://doi.org/10.1002/pst>

- Wu, C. & Thompson, M. (2020). 2.5 Sample Size Calculation. In *Sampling Theory and Practice* (pp. 25–28).
- Yuan, K.-H., & Maxwell, S. (2005). On the Post Hoc Power in Testing Mean Differences. *Journal of Educational and Behavioral Statistics*, 30(2), 141–167.
<https://doi.org/10.3102/10769986030002141>
- Zangenehpour, S. (2016). *A video-based methodology for extracting microscopic data and evaluating safety countermeasures at intersections using surrogate safety indicators*. Retrieved from http://digitool.library.mcgill.ca/webclient/DeliveryManager?pid=139122&custom_att_2=direct%0Ahttps://trid.trb.org/view/1416683
- Zangenehpour, Sohail, Miranda-Moreno, L. F., & Saunier, N. (2015). Automated classification based on video data at intersections with heavy pedestrian and bicycle traffic: Methodology and application. *Transportation Research Part C: Emerging Technologies*, 56, 161–176.
<https://doi.org/10.1016/j.trc.2015.04.003>
- Zangenehpour, S., Strauss, J., Miranda-Moreno, L. F., & Saunier, N. (2016). Are signalized intersections with cycle tracks safer? A case–control study based on automated surrogate safety analysis using video data. *Accident Analysis & Prevention*, 86, 161–172.

Appendix A: Automated Data Collection for Video-based Traffic Conflict Technique

Using human observers to conduct the traffic conflict technique requires a lot of effort, training, and time. Also, by using human observers, it introduces error due to variability, repeatability, and consistency between different observers. To address these limitations, researchers have used software and high-quality cameras to automate the traffic conflict technique.

Many researchers have developed frameworks for automated video-based traffic safety analysis, and many follow the same four steps. Jackson et al. (2013) developed a framework for a flexible, mobile video camera system for automated road safety analysis. It is not important that the video camera system is mobile, which may be useful in locations that do not have existing traffic cameras, but rather the steps taken to automatically collect and analyze traffic conflict systems from cameras. The requirements for a mobile video camera system can be applied to locations with existing traffic cameras.

Video-based traffic analysis consists of four steps: 1) data collection, 2) pre-processing, 3) processing, and 4) analysis of video data (Jackson et al., 2013).

1) Data Collection

For a video camera system to be considered mobile, 9 system requirements must be achieved: mobility, height, weatherproof, reliable, stable, vandal proof and safe, discreet, and easy to install.

Examples of the installation of mounting equipment and the cameras can be found in Figure A-1.



Figure A-1: Mobile video-based traffic conflict analysis system installed in Montreal, Quebec
(Jackson et al., 2013)

2) Pre-processing

Pre-processing converts the road users detected from video data to a real-world coordinate system to characteristics, such as speed and distance, so that computer software can be applied to conduct the conflict-based road safety analysis. It is important to know the intrinsic camera parameters, such as focal length and frame rate, to convert the camera footage. This is also the step which the video footage should be stabilized, and distortions should be clarified.

3) Processing

Software has been developed to track the movement of objects from the video footage. There are three methods that can be applied (Saunier & Sayed, 2006):

- 1) Tracking by detection
- 2) Feature-based tracking
- 3) Tracking by probability

Jackson et al. used feature-based tracking (Jackson et al., 2013). Each tracking method has different strengths and weaknesses, but this is outside the scope of work for this research.

The computer software uses feature-based tracking. The software identifies individual pixels and groups them together as a unique object based on the movement patterns of that object. These groups of pixels, which represent the object of interest, are tracked, and the trajectory is automatically extracted based on the object's unique motion patterns. The software can track the trajectories of every road user in the frame and remove erroneous trajectories as required.

4) Analysis of video

Based on the road user trajectories extracted from the video footage in the processing step, data can be collected, and different surrogate measures of safety can be determined. Basic traffic data such as traffic counts, road user classification, and turning movements can be collected. Road user behavior can also be tracked, such as lane change behavior. Most importantly, because of the capabilities of the software, trajectories can be tracked and can automatically extrapolate time-to-collisions (TTC), which is based on motion prediction. TTC requires more computing power, but post-encroachment time (PET), based on real life conditions, can also be extracted from the video data.