Impact of Mobile Photo Enforcement (*MPE*): An Analysis of the Duration Between Collisions at Enforced Sites

by

Samaa Agina

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Transportation Engineering

Department of Civil and Environmental Engineering

University of Alberta

ABSTRACT

Road collisions are one of the leading causes of death globally (*1*), with pedestrians, cyclists, and motorcyclists accounting for more than 50% of road traffic fatalities. Many countries work hard to provide safer roads for all users, especially vulnerable ones, not least because the cost to most countries is approximately 3% of their total domestic product. Vision Zero (*VZ*) aims to eliminate fatalities and severe injuries from road collisions. There is a need to determine the contributing factors leading to collisions to assess and apply appropriate solutions to accomplish that goal. For instance, speeding is a significant causal factor of road collisions, and an increase in the average speed increases the probability of a crash and its severity. To deter speed violations, Mobile Photo Enforcement (*MPE*) programs have been used as an effective countermeasure.

Past research has investigated the effectiveness of *MPE* in reducing the number of speed violators and providing safer roads. Following that, this thesis's primary goal is to study the potential impact of the *MPE* deployment efforts on the duration between two consequent collisions by examining 250, 175, 212, and 219 sites in the City of Edmonton, Canada, in 2019, 2018, 2017, and 2016, respectively. These sites had varying traffic volume levels, roadway categories, and conditions. The research methodology was performed in two main stages, namely, preparing the data for testing and applying a rigorous statistical analysis. The data was obtained from the City of Edmonton's Safe Mobility Section. Survival analysis was applied to investigate the relationship between the *MPE* variables and the duration between collisions.

The *KM* survival estimates were applied to the classified groups to determine those had higher survivability. The groups were classified into two clusters based on *MPE* hours, *MPE*

visits, *MPE HpVs,* and traffic count separately. A graph highlighted those clusters with a higher survival probability and a median survival probability over the study period (one year). Next, log-rank tests were carried out to emphasize and support the results of the *KM* survival graphs. The log-rank test established whether the two tested groups had the same survivability.

The outcomes of this analysis further support the positive effect of deployed *MPE* hours and visits on increasing the duration between consequent collisions, which correspondingly reduced the risk of collision occurrence. The results showed that the ratio between hours and visits (i.e., hours per visit) has the most impact on increasing the duration between collisions and reducing the risk of collision occurrence. The expected reduction in the collision hazard (i.e., collision occurrence) varied between 22% and 52%; the maximum reduction could be expected when the deployment occurs in high traffic volume locations, and the minimum reduction could be expected at "Arterial Only" and "All Sites" together regardless of any classifications. This conclusion is based on the results of the Cox proportional hazard models in 2019, 2018, 2017, and 2016. Moreover, *KM* graphs showed that the above-average *MPE* variables groups had a higher survival probability than those below-average. In addition, the log-rank tests confirmed the *KM* graphs inference.

Dedicated To My Parents, Ali Agina & Manal Shouman

My beloved husband, Amr Mohamed

My lovely daughter, Salma

ACKNOWLEDGEMENTS

Above all, praises are first to Allah, by whose grace enabled me to finish this work successfully.

I would like to extend my sincere gratitude and appreciation to my supervisor, Dr. Karim El-Basyouny, for giving me the opportunity to pursue this degree and for his valuable support and guidance throughout the completion of this research. His time, support, and confidence in my capabilities helped me develop and excel in many areas.

I also would like to extend my science thanks and appreciation to the examination committee members, Dr. Ahmed Hammad, Dr. Tae J. Kwon, Dr. Stephen Wong, and Dr. Karim El-Basyouny for their time and insightful comments.

I would also like to express my appreciation to my friends and colleagues for their help, encouragement, and moral support.

I would like to extend my sincere gratitude and thanks to my parents for their prayers, help, love, and presence in my life. I owe all my success in my life to my beloved father who was a role model of unlimited support, love, and care and I will always be grateful to him. I am also immensely grateful to my mother for her prayers and support. With all my love and gratitude, I would like to especially thank my amazing husband, Amr Mohamed, for his unconditional support, encouragement, and motivation. Also, I extend my thanks to my dear brother Mohamed and my sister Alyaa for their love and being in my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ASE: Automated Speed Enforcement

CoE: City of Edmonton

EPDO: Equivalent Property Damage Only

HpV: Hours per Visit

KM: Kaplan Meier

MPE: Mobile Photo Enforcement

MTO: Ministry of Transportation Ontario

PH: Proportional Hazard

PDO: Property Damage Only

VZ: Vision Zero

WHO: World Health Organization

1 INTRODUCTION

Traffic collisions are among the major leading causes of accidents, claiming over 1.3 million lives in addition to injuring twenty million people annually (*2*). Moreover, the World Health Organization classifies road collisions within the top ten causes of fatalities (*1*). In Canada, for example, at least five people die daily due to road traffic collisions. In 2001, 2,778 and 224,000 fatalities and injuries, respectively, resulted in an estimated social cost to Canadians of $25 billion (*3*). The situation is heightened in the US, where reports indicate that about 90 daily deaths occur (*2*). Aside from the tragic toll on the victims, these road collisions greatly impact those around them both psychologically and financially (*4*).

Two significant aspects contribute to road collisions: roadway features and drivers' behaviour. Roadway characteristics are considered as contributing factors to collisions that might be limited sight distance, weather conditions, pavement conditions, or traffic signals visibility. Drivers' behaviour exerts an even greater influence on road crashes, observed as speeding, impaired driving, or fatigue. Indeed, speed alone was identified as the major contributing factor in 30% of fatalities in high-income countries and 50% of all crashes in low-income and middle-income countries (*1*). Since speeding is defined as either exceeding the speed limit or driving within the speed limit but too fast for the road conditions (*6*), more stopping and braking distances are required. If those distances are not factored into drivers' decision-making, then speeding increases the probability of collision occurrence. Furthermore, collision severity is proven to be directly correlated to vehicle speed (*7*). As such, even small increases in vehicle speed could lead to remarkable increases in collision probability and severity (*8*).

As speeding and collision occurrence are significantly related, targeted strategies are regularly implemented to deter irresponsible driving behaviour, manage driving speed, and reduce violations. At their core, these strategies are based on three approaches: Engineering, Education, and Enforcement (3Es) (*9 - 11*). Research has shown that speed enforcement is one of the most effective countermeasures to stop speed violators and improve drivers' behaviour (*12*). For instance, the Automated Speed Enforcement programs (*ASE*) are proven interventions that reduce the collision severity and frequencies. Previous findings attribute an 8.9 % drop in collision numbers and a 12 % decrease in collision fatalities and injuries to *ASE* programs (*13*).

Many Canadian cities, such as Winnipeg, Calgary, and Edmonton, have effectively implemented such programs. In Edmonton, for instance, data indicate that *ASE* programs have reduced collisions from 20 to 14 % (*6*). The City of Edmonton's primary goal when employing these programs is to minimize traffic incidents and enhance road safety. To achieve this, authorities have to prioritize and map the collision risk sites. Accordingly, *ASE* measures must be implemented in tandem with camera use at high collision sites for effective enforcement. Findings in Alberta suggest that Mobile Photo Enforcement (*MPE*) deployment is the best alternative at these sites (*14*). However, it is unclear whether the *MPE* programs have a considerable impact on the survivability of collision occurrences at different road locations. Therefore, this study seeks to explore this research gap and investigate the influence of deploying different *MPE* variables on the duration between collisions. The work done here emphasizes the effectiveness of the *MPE* program in reducing the duration between two consequent collisions, demonstrated through hazard models. The collision severity, collision frequency, exposure to collision risks, traffic counts, and *MPE* data were analyzed for this research.

Several national and international reports have shown that the introduction of *ASE* programs has registered positive outcomes in traffic safety (*16*). However, the consensus leans towards speed enforcement because intersection safety devices, in most cases, produce varying results (*15*). Previous reviews of worldwide findings between the late 1990s and 2000s showed that the introduction of *MPE* effectively reduces speeding by 82 and 51 %, respectively (*17*). In particular, Li points out that the *MPE* program is reliable because the perceived risk of detection increases with each subsequent deployment, resulting in broader deterrence levels and lower numbers of speed violations (*15*). Although these studies are based on different methodologies and showed a positive effect of *ASE*, there is still a need to investigate and evaluate its influence on the duration between collisions and consequently on enhancing traffic safety.

1.1 Background

As highlighted in the introduction, red-light running and speeding constitute the leading causes of road accidents in the USA and Canada (*13*). Research suggests that both increase the risks of crashing, death, and injuries. Studies have shown varied consequences from these two infractions based on magnitude and severity. Indeed, speeding and the chances of being involved in a collision are directly proportional (*13*). According to Evans, an increase in the average speed by roughly 1 % raises the fatality risk from anywhere between 4 and 12 % (*18*). Doubling the speed doubles the risks of collisions, injuries, and death, while exceeding the average speed by 20 %

escalates the risks six times the norm (*18*). In sum, previous meta-analyses engaging more than ninety-eight studies concluded that the interrelations between road safety and speed are significant and meet all the causality categories typically used in predominant evaluation studies. Elvik (2005) argues that not only is speeding the single highest determinant in traffic fatality cases but also, there is a significant difference in fatality risk when a moving vehicle moves faster than the surrounding traffic (speed dispersion) (*19*).

Collisions resulting from red-light running vary in magnitude. Studies suggest that this type of infraction often results in right-angle collisions, which are believed to be more severe than other collision cases (*20*), costing more than US$14 billion (*21*). While excessive speeding contributes to more than 18 % of all crashes (*21*), translating to over 2,000 injuries and deaths attributable to the collision every year, red-light running accounts for more than one-quarter of the traffic injuries (*22*). Studies conducted by the Ministry of Transportation Ontario (*MTO*) in 2014 noted that disobeying traffic signals account for 42 % of fatal crashes alongside 29 % of injury crashes.

Traffic enforcement is one way the authorities mitigate the number and severity of these two serious traffic infractions. Therefore, this research aims to correlate the impact of different *MPE* variables (i.e., hours, visits, and hours/visits) and the survivability of collisions occurrence in different road cases. The subsequent sections shall discuss the implementation of *MPE* technologies in reducing the duration between two consequent collisions by employing a survival analysis.

1.2 *MPE* Operation

Today, many cities worldwide have embraced *MPE* technology, registering positive results by limiting speeding and speed-related collisions. For example, the technology has reduced non-fatal and fatal collisions in France by 26 and 21 %, respectively (*23*). In Charlotte, North Carolina, the introduction of *MPE* has reduced collision cases by an average of 10 %. In Washington D.C., the speeds were reduced by 14 % with a further reduction of 82 % based on the number of vehicles exceeding the posted speed by ten mph (*24*). In Canada's British Columbia, speed-related collisions were trimmed by 25 % in various enforcement locations (*23*). Moreover, in Australia, collision cases dropped by 22 %, which culminating resulted in a 38 % fall of collision-related injuries (*23*).

Moreover, automated enforcement overcomes the equity issues associated with policing enforcement. When conducting automated enforcement, officers do not interface directly with citizens, unlike traditionally policing which requires in-person exposure, which might lead to racial or personal bias.

However, despite these reported success cases, it is still unclear how we can assess the impact efficacy of this technology in relation to enhancing safety. To this point, the primary concern has been how the resources can be assigned and utilized to obtain maximum safety and positive impacts (*23*). However, there is a gap in understanding the relationship between the *MPE* deployment and collision frequency. Therefore, this research aims to incorporate the program performance information in the literature alongside the data obtained from its enforcement to produce a robust study on the *MPE*'s effectiveness in reducing the duration between consequent collisions, which subsequently would provide safer roads for all users.

1.3 Problem Statement

Current road safety countermeasures vary in their impacts on road users, with the *MPE* gaining greater traction for regulating speed and enhancing road safety. However, the research to date lacks data on its effectiveness in increasing the duration between consequent collisions in different road conditions. Previous work focused solely on assessing the *MPE* using the before and after Empirical Bayes method (*25*), applied on urban arterial roads. Since the *MPE* programs are vital resources and need to be appropriately utilized, more comprehensive assessments in greater detail are necessary.

Although previous studies focused on exploring the *MPE* effectiveness on speeding violations and road collisions, it is unknown whether the *MPE* variables (i.e., deployed *MPE* hours and visits) significantly influence the duration between two consequent collisions on different road categories. Therefore, there is a need to investigate the correlation between *MPE* deployment variables and the survival time for different locations if we are to evaluate the *MPE* program performance. Exploring this relationship would assist municipalities in understanding the consequences of deploying *MPE* hours, visits, and the ratio between them. As a result, they may consider adopting a proper *MPE* program and efficiently invest their resources.

1.4 Research Objective

The research's main objective is to correlate the deployed *MPE* variables (i.e., hours, visits, and hours/visit) and the interval between two consequent collisions. In order to achieve this goal, the *MPE* and collision data are analyzed using survival analysis, which illustrates the survival probabilities as a result of deploying *MPE* hours and visits within various road categories. Moreover, this research has two sub-objectives i) to investigate the impact of deployed *MPE* variables on different road categories (i.e., All sites, Arterials, Collectors, High/low traffic volume locations, & Speeding-related collisions) and ii) to determine the optimal *MPE* deployment variable that provides the longest duration between collisions and correspondingly reduces collision hazards. Also, the analysis predicts the expected reduction of collision occurrence by deploying different *MPE* variables. These outcomes result in a better understanding of the optimal *MPE* deployment strategies for different roadway categories.

1.5 Research Methodology & Organization

To achieve the research objective, the methodology employed here is shown in **Figure 1**. It was executed in two major stages: i) preparing the data and ii) applying the survival analysis. The analysis used available *MPE* and collision data from 2016 to 2019 and was applied to explore the relationship between the frequency of deployed *MPE* (i.e., number of visits and hours) and the duration between two consequent collisions.

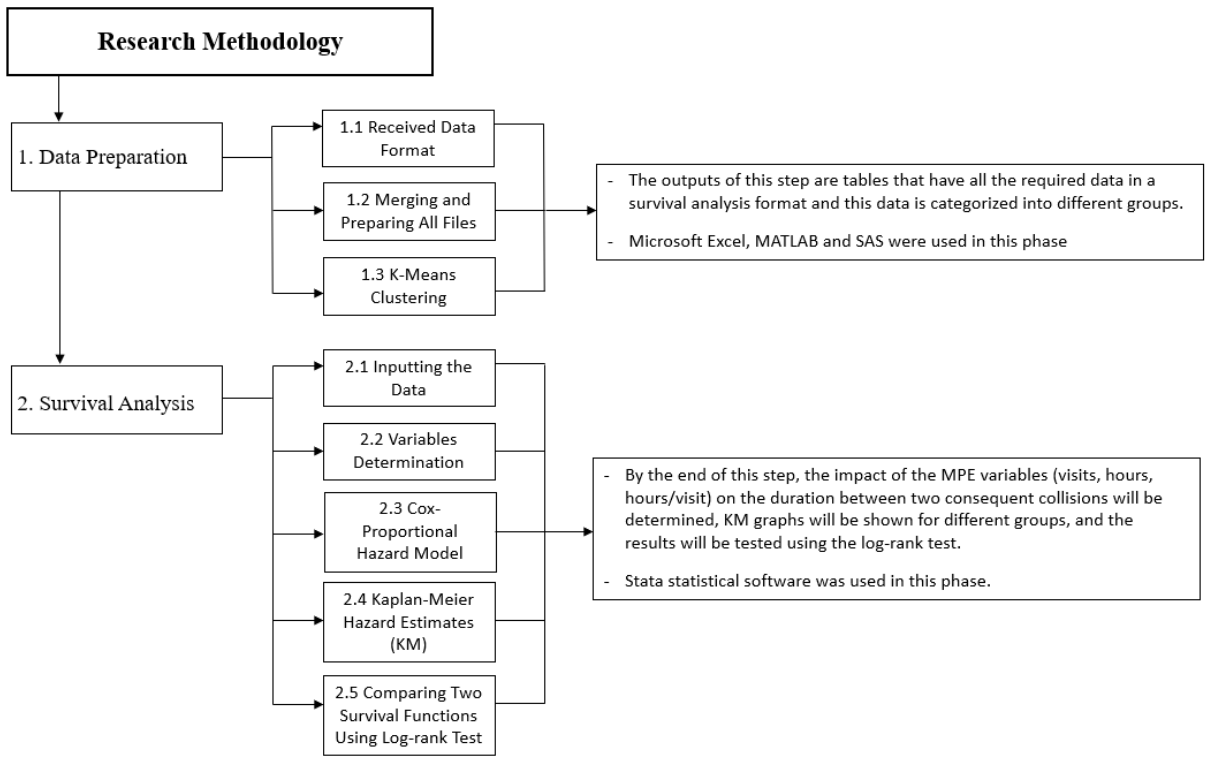


Figure 1: Summary of the Research Framework.

1.6 Thesis Structure

The remainder of the thesis is divided into five chapters as described below:

- *Chapter 2* reviews existing studies on the *MPE* program and its effectiveness. The chapter also outlines previous assessment methods used for analyzing deployed *MPE* hours and visits. Moreover, Survival Analysis and its use in transportation engineering are introduced.

- *Chapter 3* describes the data provided by the City of Edmonton and highlights the main variables considered in this study. It also provides details of the analysis period and how it is utilized in the analysis process.
- *Chapter 4* illustrates the methodology's two major stages: *i*) preparing the data and *ii*) applying the survival analysis. The details of each step are provided and supported by relevant sources.
- *Chapter 5* outlines the significant analysis results, as tested on different road categories, followed by an in-depth discussion of their implications.
- *Chapter 6* summarizes the research conclusions and discusses the study's contribution to our understanding of the effects of *MPE* on traffic safety. Moreover, it outlines the research limitations and potential future work.

2 LITERATURE REVIEW

This chapter provides an overview of the *MPE* and previous studies on the *MPE* allocations, applications, and impacts. It also presents background information on survival analysis and its usage in transportation engineering. Moreover, the Cox proportional hazard model is introduced to illustrate building a hazard regression model using survival analysis.

2.1 Mobile Photo Enforcement (*MPE*)

Studies on *MPE* generally cover its effectiveness on collisions and speeding by underscoring the particular deterrence effects, deployment strategies, and resource allocation. Most studies reviewed for this research concern the *MPE* programs' influence on collisions and vehicle speeds. Previous studies have shown that the practical application of *MPE* can minimize mean vehicle speeds by 2% (*23*). Also, related studies have pointed out that *MPE* reduces severe collision cases causing injuries and fatalities.

The effectiveness of the *MPE* program can be attributed to the immediacy, unavoidability, and severity of punishment, which influences driver behaviour and attitude based on the specific and general deterrence mechanisms. General deterrence is when potential violators adhere to the outlined standards when they become aware of others being punished or possible consequences for breaking road traffic rules. Research has linked general deterrence with dangerous driving education, *MPE,* and awareness campaigns. On the other hand, specific deterrence is where a driver receives a firsthand experience of detection and punishment (*25*). Some scholars posit that general deterrence can be enhanced by targeting high-risk periods and locations through nonvisible and visible enforcement strategies. Utilizing both can enhance unpredictability and enforcement publicity while also embracing long-term schemes in the enforcement program (*23*). In Edmonton, for instance, an increased number of issued tickets and enforcement sites simultaneously reduced speed-related collision cases (*16*). Significantly, the same City of Edmonton report indicated that the reduction in collisions was linked to the *MPE* program's reliability, notably when implemented with higher location coverage, increased issued tickets, and consistent checks (*25*).

Although different jurisdictions shape the guidance and regulations of these programs (e.g., Alberta's provincially-dictated Automated Enforcement Guidelines (*14*)), *MPE* units are typically located at sites known for speed limit violations, collisions, and public complaints regarding speeding (*26*). Additionally, officers deploy *MPE* units when they receive a special request from the local government.

There are some cases where the conventional speed control measures have failed or are infeasible. Studies show that public awareness is effective for speed management. Programs, such as the Speed Management programs in Alberta and Manitoba, aim to change the perceptions and attitudes of drivers by raising awareness of the risks linked to red-light running and speeding (*23*,*26*). In addition, reports have shown that public awareness can be effective when complemented with other enforcement techniques (*23*).

Speed cameras and enforcement programs are not without controversy, however. Carnis (2011) highlights persistent public debates over privacy, reliability of cameras, and fairness that can emerge in jurisdictions when speed cameras are installed. Additionally, there is some skepticism by segments of the population that moderate speeding heightens crash risks, although some evidence shows that public awareness programs could be instrumental in reducing such beliefs (*27*) by targeting drivers through visible enforcement programs, awareness, and education efforts.

Further research assessed the *MPE* using the before and after Empirical Bayes method (*28*). This method showed a reduction of 14% to 20% in collision severities by deploying speed enforcement with the highest effect of the *MPE* scored for severe collisions. However, the method was limited to only urban arterial roads, limiting our understanding of *MPE* programs' effectiveness in varying road conditions.

2.2 Survival Analysis & Hazard Models in Transportation Engineering

Survival analysis and hazard models have been widely used in medical applications. However, engineering is increasingly applying both to field-related problems. Broadly speaking, survival analysis is used to determine how long an event lasts, either in terms of "survival" or "failure," with failure often meaning structural failure in many engineering applications. More specifically, transportation engineering commonly uses survival analysis in relation to the duration that roadway incidents, such as car crashes, impact traffic (*29 - 32*).

There are three related functions in survival analysis: the failure function, the survival function, and the hazard function. The failure function illustrates the probability of a failure incident occurring before the specified point in time (*29*). The survival function is the inverse of the failure function, representing the likelihood that the recorded duration continues beyond the specified point in time. The hazard function is related to, but distinct, from the other two. It represents the potential that an individual will “fail” at a specific point in time, having survived until then (*29*). For instance, Nam and Mannering (1999) examined the duration of traffic incidents from a dataset of incidents in Washington State during 1994-1995, related to reporting, response, and clearance times by the Washington State Incident Response Teams (*31*). The researchers considered the “failure” to be when an Incident Response Team fully cleared the incident and “survival” to be the persistence of the incident. The hazard function represented the probability that the incident would be removed at any given point in time.

Previous studies show several variables that may impact transportation engineering outcomes (i.e., incidents). For instance, researchers using the Cox proportional hazard model found that both young and male drivers (i.e., gender and age variables) received speeding citations more often than other gender and age groups (*33*). Moreover, drivers who receive speeding citations (i.e., tickets) are more likely to get them more frequently, which means speeding citations have little influence on changing speeding behaviour compared to other speeding penalties (*33*).

Studies in transportation engineering have used survival analysis to examine the vehicles’ mandatory lane-changing behaviours and related variables using the Cox proportional hazard model (*34*). They show that the type of vehicle has no significant impact on the duration of mandatory lane changing and that the mandatory lane-changing survival rate during the peak period is higher than the off-peak time (i.e., time is the considered factor).

2.2.1 *Cox Proportional Hazard Models & Log-Rank Tests*

The Cox regression is applied to evaluate survival statistics and compute the regression constant of multi-impacting aspects represented by β_1. Unlike the conventional regression evaluation, the Cox regression does not use survival duration t as the dependent adjustable of the regression function (*30*). All the inclining aspects are undertaken as a covariate adjustable x. Moreover, the hazard function h (t, x) relative to standard hazard function h_0 (t) describes the measurable influence of all the covariates adjustable to the survival duration, where h_0 (t) is the intrinsic hazard equation under the state of no impacting aspects (*30*). Therefore, the Cox regression is also referred to as a proportional hazard prototype. The Cox proportional hazard sample can be applied in the co-evolution evaluation of multi-inclining aspects of traffic event time without any statement of survival time spreading.

A log-rank trial can be applied to analyze the importance of the inclining aspects. If an individual refutes the aspects that their importance gratifies a state, the inclining or influencing aspects are persistent, each of which has a distinctly substantial impact on traffic event duration (*30*). The factors include event type, event scene, night and daytime, traffic condition, count of lanes, event region, trailer condition, and bottleneck condition. The Cox proportional hazard can be used to analyze the co-evolution among the inclining aspects of the traffic event period.

This model can apply the reverse technique to choose a significant covariate adjustable and approximate its constant regression filter, the ultimate probability ratio. Even though one prompting factor can have enormous importance to the traffic event time discretely, it may have a negligible impact under the mixture of these multi-prompting aspects when considering other factors (*30*). Thus, the approximation outcomes for the Cox proportional hazard sample-based parameter may show only part of the prompting aspects that can endure in the prototype of traffic event time (*35*). The factors may include event location, vehicle number, night and daytime, trailer condition, event type, bottleneck condition, and intricate lane number.

2.3 Discussion & Research Gaps

As illustrated in the literature review, speeding is considered a primary contributing factor to collisions. Therefore, municipalities exert significant effort to deter speed violators using different tools and programs such as speed cameras, increasing the fines for speed violating, Mobile Photo Enforcement (*MPE*), and educational campaigns. Studies add to this knowledge by identifying relevant factors contributing to road safety. For instance, variables such as age, gender, time, and vehicle type significantly affect the likelihood of receiving a speeding citation. Moreover, previous research focused on assessing the effectiveness of deployed *MPE* programs by using the Before and After Empirical Bayes method. In addition, previous methods evaluated the *MPE* programs without insights into how to efficiently relate *MPE deployment parameters to improvements in safety*. Therefore, it is vital to move forward from studies that evaluate the *MPE* programs to studies that can explain how *MPE* strategies can improve road safety outcomes. Moreover, there is room for improvement and potential development to assess the impact of *MPE* variables on different roadway categories and conditions.

Survival analysis has been increasingly utilized in transportation engineering studies. Previously, survival analysis was used in examining the duration of traffic incidents. By using survival analysis, this research explores the effectiveness of *MPE* deployment variables in increasing the time between collisions and consequently reducing the risk of collisions occurring and improving the safety of the roads.

3 DATA DESCRIPTION

The data used in the analysis was accessed through the Safe Mobility Section in the City of Edmonton (*CoE*) and provided in various spreadsheet files: "All-collision data," "*MPE* data," "Site," and "Event." The all-collision data spreadsheet included historical information, and these files combined data about collision cause, time, location, and travel direction. The *MPE* data file established the *MPE* control types, and the start and the end of the *MPE* at each location. By using this spreadsheet, detailed information regarding the duration and number of *MPE* visits for each location was extracted. The control types covered all the feasible methods of the *MPE* in the city. The Site spreadsheet contained site IDs for all locations and detailed descriptions of each site. The "site ID" info was the core element that linked all the files together, clarified in the following sections. Finally, the Event dataset was a supporting file that provided an overall idea of the traffic count in each location.

The analysis period for this study is four years, from 2016 to 2019. Although data for the year 2020 was available, it was excluded from the study due to the documented impact of the *COVID-19* pandemic on the transportation system. The data were analyzed for each year separately and then integrated for further analysis.

This work was implemented on various sites with different geometric characteristics and traffic volumes. All the available *MPE* control sites with sufficient information were analyzed based on the provided data. Therefore, the number of sites and locations differed yearly based on the available information. The following sub-sections show the received data format and some data descriptive statistics.

3.1.1 *Data*

An extensive set of data was provided across several unlinked spreadsheets. The following bullets outline the file titles and their contents.

- <u>All collision data</u>: Data included collision code, attachment code, collision key, collision cause, collision classification, collision data, collision location name, collision month, collision time, collision report year, travel direction, on-street name, and at-the street name.
- <u>Control – *MPE*</u>: Data provided included control ID, site ID, violation category, control type, start date, end date, posted speed, and speed threshold. The data was filtered to contain only the *MPE* control type.
- <u>Site</u>: Data included site ID, location description, police division, direction, speed, double fine site, speed time site, speed posted, photo enforcement posted, and site type.
- <u>Jenoptik Event</u>: Data included deployment, site ID, watch date, and traffic. This data was integrated to get the traffic count for each site.

3.1.2 *Data Descriptive Statistics*

With respect to the number of *MPE* sites, it was found that they varied through the study period. The maximum number of *MPE* sites was recorded in 2016 as it might be related to the initiation of Vision Zero in the City of Edmonton. Moreover, as shown in **Figure 2** the number of *MPE* sites tended to decelerate from 2016 to hit the minimum in 2018, then increased in 2019. This fluctuation in the number of deployed *MPE* sites could be associated with the historical collision records for each site.

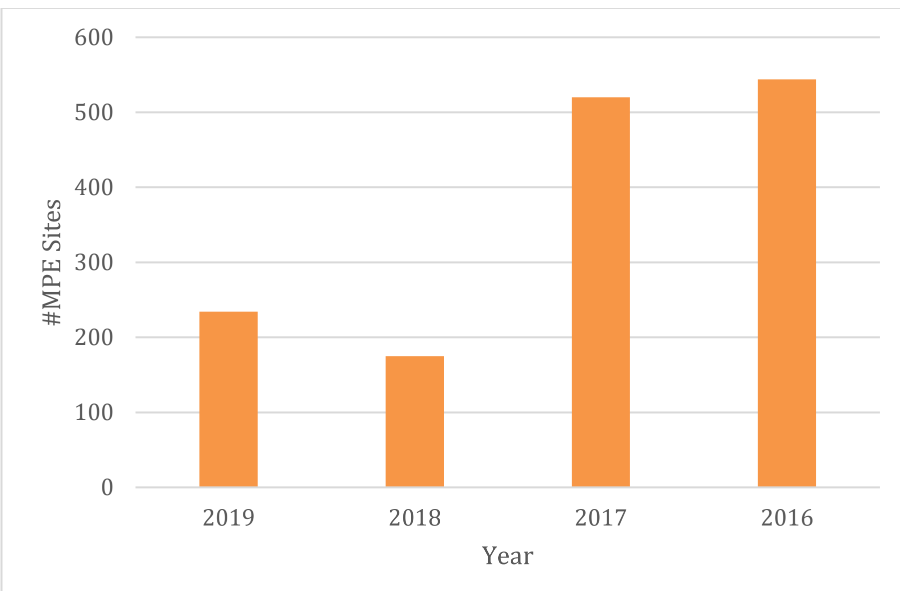


Figure 2: *The Recorded Number of MPE Sites.*

In addition, the *MPE* deployments were implemented on different road classifications. The majority of deployed *MPE* visits and hours were focused on arterials and collector roads as the speeding violations and potential collisions are more expected in these categories. The following chart shows the distribution of arterial and collector *MPE* sites.

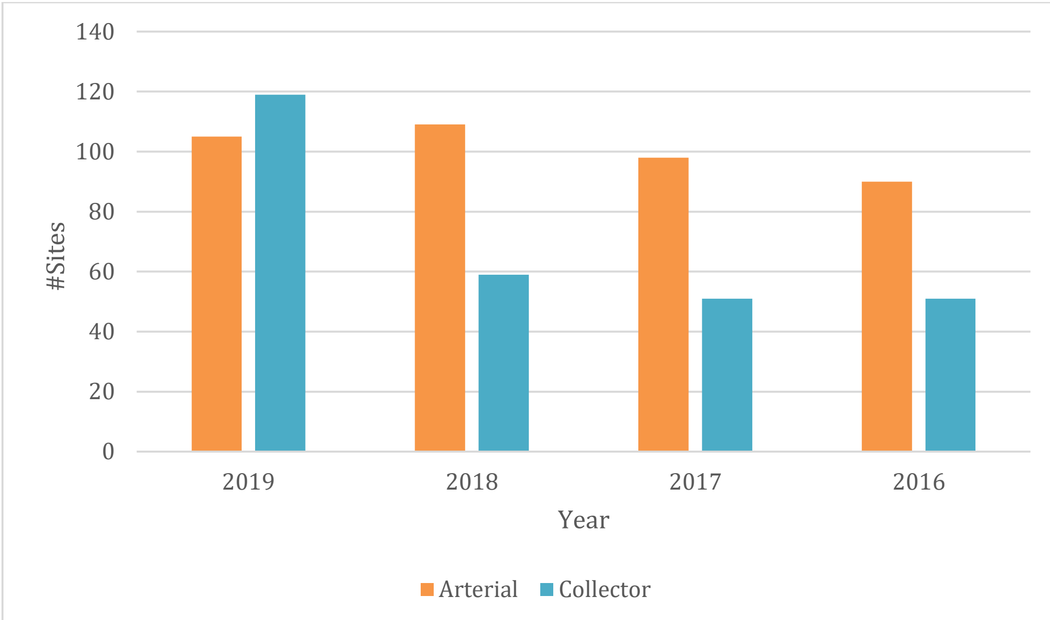


Figure 3: *Distribution of Number of Arterials and Collector Roads.*

Regarding the number of collisions at the enforced sites, it decreased dramatically from 2016 to 2019 which might reflect the influence of deployed *MPE* and other countermeasures. Moreover, the number of speeding-related collisions was recorded at its minimum value in 2019 at 217 collisions, compared to 625 collisions in 2016. **Figure 4** shows the distribution of the total number of collisions and speeding-related collisions.

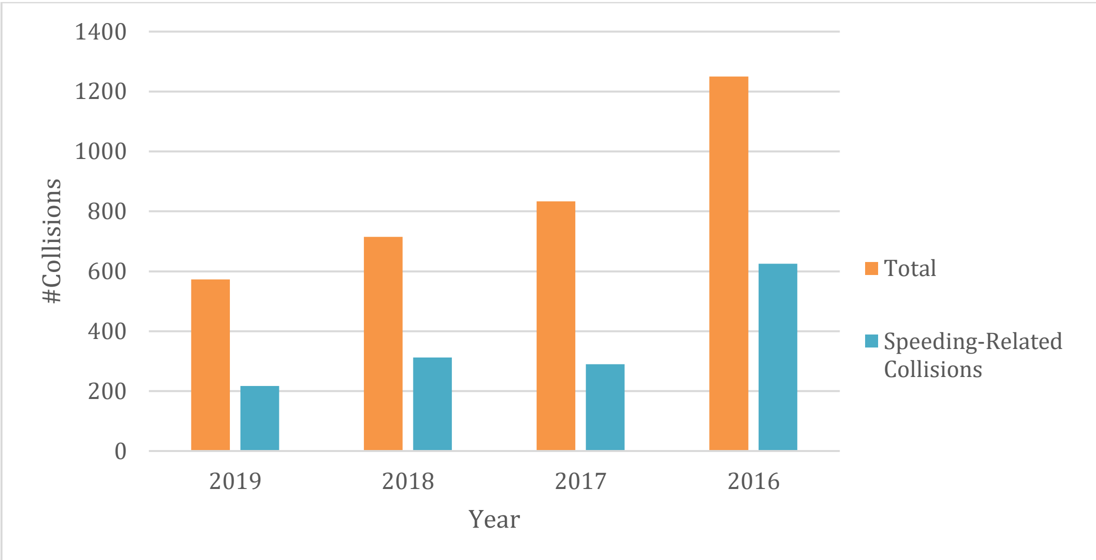


Figure 4: *Distribution of Total Number of Collisions and Speeding-Related Collisions.*

With respect to collision causes, following too closely was the most common collision cause (34%) throughout the study period. The following figure shows the distribution of the top 5 collision causes.

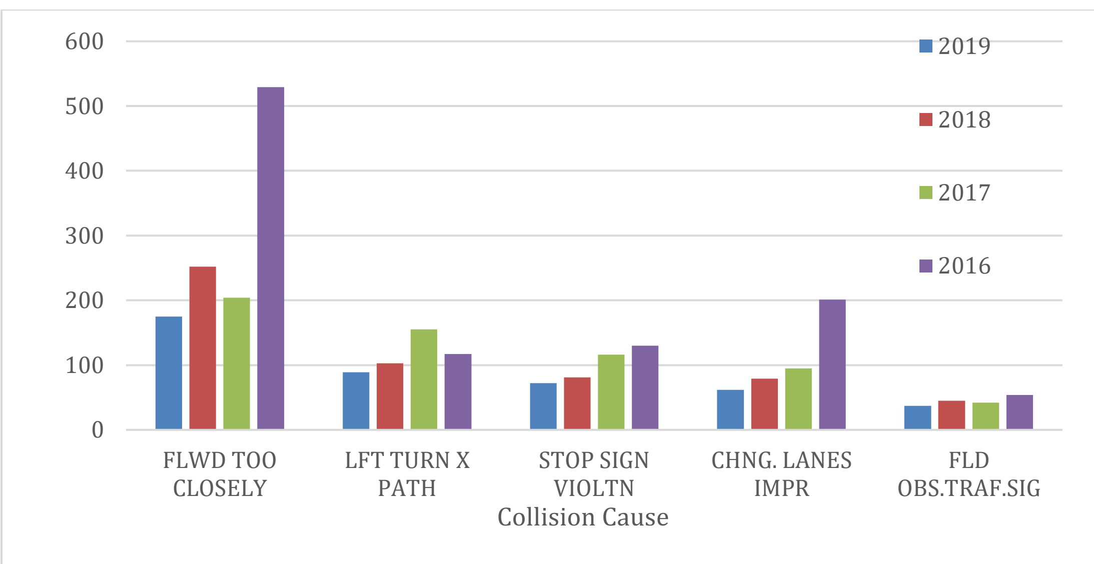


Figure 5: Distribution of Collision Causes.

4 METHODOLOGY

4.1 Data Preparation

The data was first organized to be used readily in the survival analysis (i.e., the second stage). This step aimed to merge all the data files into one, then change all the combined files into a survival analysis format. The output from this process was an excel sheet that contained the essential info for each location (i.e., Site ID). While this step could be done manually, the time required made it impractical. So, a MATLAB code (attached in the appendix) was scripted to speed up the processing time. The following subsections clarify this data preparation step thoroughly.

4.1.1 *Merging & Preparing All Files*

In order to overcome the difficulty of merging files manually, a MATLAB code was scripted and tested on the raw data files. The code read the input data files and stored them in four tables (All-collision data, Site IDs, *MPE* data, and Traffic data), setting up the excel files to store the output along with all the headers for the data. The code then looped through all the Site IDs to determine locations for each, i.e., the street and avenue number/name and the direction of travel (e.g., Site ID 104 represents the location of 142^{nd} street and 95^{th} avenue, Northbound direction). Thereafter, a filter was applied to extract all the collisions for particular locations within the given site description (i.e., street and avenue name and travel direction) along with information to determine collision dates, times, and seasons (i.e., winter, summer, etc.), accomplished by looking at each collision individually and checking its location.

Using another loop, the collisions were sorted, and the number of hours visited, the number of visits, traffic volume, and the time difference between collisions were all calculated. The number of hours visited was calculated by filtering all data from the *MPE* data table for those within a given year. The aim was to extract the time from the start to the end of each visit. The number of visits was calculated by counting the total visits from the *MPE* data table that fell within the time range. The traffic count was calculated by filtering all the data from the traffic data table to search only for the ones within the given year and extract the amount of traffic. These loops were executed on all site IDs for each year.

Finally, each year's calculated data was stored within a proper excel spreadsheet. The average processing time for this step was 90 minutes. The code outputs were visually inspected to verify accuracy and that they matched the expected results.

4.1.2 *K-Means Clustering*

K-means clustering was used to categorize the sites based on different variables; *MPE* number of hours, *MPE* number of visits, *MPE* hours/visit, and traffic count. This form of clustering groups observations with similar characteristics (*36*) (i.e., into clusters in which each observation belongs to the set with the nearest mean). First, the number of clusters (*K*) should be determined. The *K* was determined to be two clusters, therefore, each cluster would have enough observations to represent the behaviour of each cluster. It is worth noting that, the observations were tested and clustered into three clusters (i.e., high, intermediate, and low) but it was found that the high and intermediate observations had the same behaviour. Therefore, only two clusters were used in this study, namely, above-average and below-average variables. Then, two cluster seeds were randomly specified as the clusters' centroids. After that, each observation was assigned to one of the clusters based on its proximity that had the least squared Euclidean distance. Finally, the centroid of each cluster was calculated, and iterations were done until convergence was reached (i.e., the same points were assigned to the same cluster in repetitive cycles). In order to apply the *K*-means clustering process, SAS software was used. Though the *K*-means was conducted in SAS, the data was entered and clustered into two groups

Table **1** shows the clustered groups.

Table 1: The Clustered Groups.

Variable	Group 1	Group 2
#*MPE* hours	Above-average *MPE* hours	Below-average *MPE* hours
#*MPE* visits	Above-average *MPE* visits	Below-average *MPE* visits
#*MPE* (hours/visit)	Below-average *MPE HpV*	Above-average *MPE HpV*
Traffic volume	Below-average traffic volume	Above-average traffic volume

4.2 Survival Analysis

The second phase in the methodology was to apply the survival analysis process. Survival analysis was used to study the time to event occurrence. Thus, in this case, the failure event was the collision occurrence. The basic concepts of the survival analysis are to define the hazard and survival functions, create the Kalan-Meier survival curves for different variables and compare two survival curves using the log-rank test. In order to execute this step, Stata Statistical Software was used. There were five stages applied to perform the survival analysis. The relationship between the *MPE* variables (the number of hours, the number of visits, and the ratio of hours to visits) and the duration between two consequent collisions was determined by this process's end. The following subsections will illustrate the implemented phases.

4.2.1 ***Inputting the Data***

The data was first classified as survival-time data and was tabulated to include the time and failure variables. The time variable represented the duration between two consequent collisions in days, and the failure variable was a binary value (i.e., 0 or 1). The failure value was 0 when there was no collision and 1 when there was a collision during the specified period. Moreover, the data in this step included site ID, duration between every two consequent collisions, first and second collision dates used to estimate the time between collisions, number of deployed *MPE* visits, number of deployed *MPE* hours, the ratio between the number of *MPE* hours to the number of *MPE* visits (*HpV*), first and second collisions occurrence season, collision severity, collision cause, land-use, and traffic count.

4.2.2 ***Variables***

The concerned variables were determined to be the number of deployed *MPE* visits, the number of deployed *MPE* hours, the ratio between them (*HpV*), and the traffic count. The impact of these variables on the duration between two consequent collisions would be examined separately in the survival analysis phase to establish the variable with a higher impact on collision occurrence.

Moreover, these variables were clustered in the previous stage to explore whether the above-average or below-average variable provided more survivability to the exposed *MPE* locations.

4.2.3 *Cox-proportional Hazard Model*

The cox-proportional hazard model was used to investigate the effect of each variable on the time between collisions. This model is the most popular technique of the semi-parametric methods since it does not make a hazard baseline assumption. This is beneficial when choosing a predictive model (*37*) since it explores the rate of a specific event occurrence (i.e., hazard rate) influencing different factors.

The general Cox proportional hazard model is:

$$h\ (t, x) = h_0\ (t)\ \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_t)$$

The Cox regression prototype can be changed into another equation by the logarithmic transformation (*30*);

$$\ln \frac{h\ (t,x)}{h_0\ (t)} = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_t$$

The relative risk *RR* can be denoted as $\frac{h\ (t,x)}{h_0\ (t)}$, then, the *COX* regression is the linear model of the logarithm of the *RR*. Under other covariate variables remaining constant, βi shows the logarithm changes of the *RR* with the unit change of the i^{th} covariate variable. Based on the definition above, the *COX* regression has the following properties: 1) If $\beta i>0$, it means that the i^{th} variable is a risk factor, and its hazard may be higher with the increasing time. And this indicates that the incident may be disposed of quickly. 2) If $\beta<0$, it means this variable is a protective factor, and the duration of the traffic incident is longer, indicating that the incident takes longer to be cleared. 3) If $\beta=0$, it means this variable has nothing to do with the traffic incident duration.

(*30*);

$$\ln \frac{h\ (t,x)}{h_0\ (t)} = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_t$$

On the other hand, $S\ (t/Y = 1, x) = \exp\ (-\ \exp\ (\beta' x) \int_0^t h_0(u/Y = 1)du)$

$$\text{Significantly, } \frac{h_1(t \mid x_1)}{h_2(t \mid x_2)} = \frac{h_o(t)\exp(\sum_1^p \beta_j x_{1j})}{h_o(t)\exp(\sum_1^p \beta_j x_{2j})} = \exp[\sum_1^p \beta_j (x_{1j} - x_{2j})]$$

The outcome of this step is a regression model that relates the deployed *MPE* hours, visits, and *HpV* separately. These models provide the hazard ratio for each variable; the impact of the aforementioned variables will now be explored.

4.2.4 *Kaplan-Meier Hazard Estimates (KM)*

The *KM* hazard estimate is a univariate nonparametric analysis used to estimate the survival probability from observed survival times (*38*). In this study, the *KM* survival curve was established to differentiate the impact of *MPE* variables on different road categories. It was generated by taking the product of conditional probabilities sequences and obtaining the standard estimator (i.e., *KM* estimator). Carroll provides full details of the calculation (*37)*. As shown in the following figure, this curve consists of a series of steps, and each step represents an event occurrence (i.e., collision). The survival probability is on the *Y*-axis of the curve, and the time duration is on the *X*-axis. Thus, the cumulative survival probability can be extracted at any time point by obtaining the corresponding value on the *Y*-axis. The estimated cumulative survival at any time point is at 95% confidence intervals.

If T_i represents the event period of the traffic of the i^{th} term example, as the time-series established T gratifies the traffic event in the state where $T_1 < T_2 <\ldots T_n$, then the Kaplan-Meier based survival likelihood of traffic occurrence period represented by Š (t) is given by;

$$\check{S}\,(t) = \prod_{T_i^c \Sigma t} \frac{n-i}{n-i+1}$$

In this case, the traffic event time is represented by T_i^c of the ith term entire samples. However, for a sample to be complete, it has to attain certain conditions, including, T_i^c is less than t besides being a positive integer, where $T_i^c \leq t$ and$T_i^c \in Z$ (*39*).

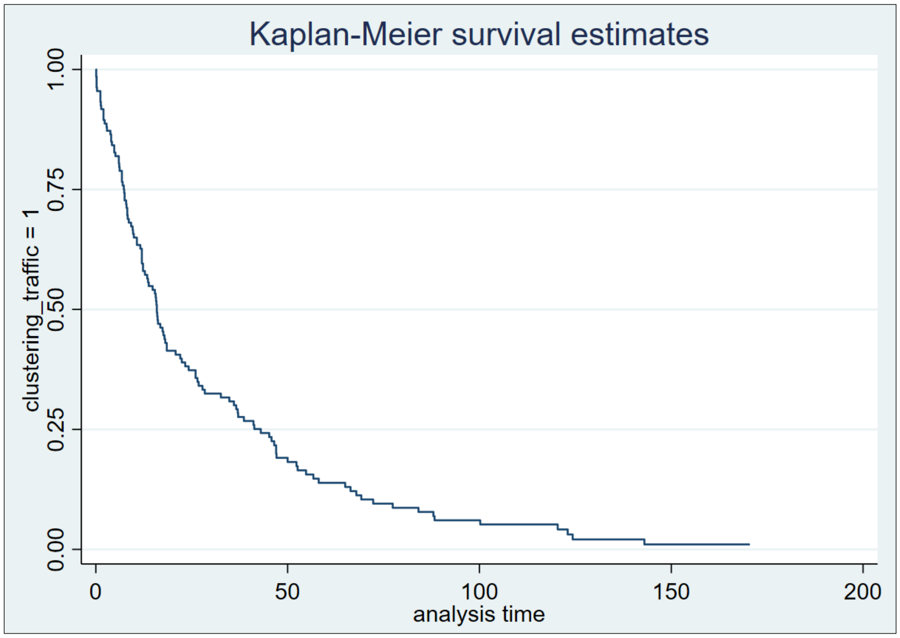


***Figure 6**: Example of KM Graph.*

The curve shown in **Figure 6** provides a beneficial data summary that can be used to estimate measures such as median survival time. It is used to plot the difference between the survival probability for two comparable groups (i.e., clustered variables). The data of these groups should be in categories as this method estimates the cumulative survival probability for each group separately. These groups are determined after being clustered based on the K-means Clustering method that is performed in Stage 1, previously outlined.

Therefore, this step's output is a graph for each variable that shows the survival probability at any time point during the analysis period. Also, it can be used to compare the survival probabilities for different categorical groups. The survival probability for each group is generally checked against the mean value (50%) to facilitate comparison between the groups.

4.2.5 *Comparing Two Survival Functions Using Log-rank Test*

There are two methods to compare survival functions; the first method uses a prespecified time point; the other compares the overall survival experience, called the log-rank test. The log-rank test is considered more reliable than the prespecified time point method for reasons outlined in (*36*). Based on these reasons, the log-rank test was applied and executed in this study across the entire survival time range. The test null hypothesis between the two groups is:

$H0\text{: } S1\,(.) \;=\; S2\,(.)$

where the dot represents the whole survival time range.

The alternative hypothesis is applicable when this null hypothesis is rejected. Moreover, the log-rank test compares the observed and expected collisions if the two groups have the same survival function. Thus, if the null hypothesis is true, the two groups would have the same survival probability, determined based on the P-value. This test estimates the Chi-square value for the compared groups and concludes the result. Then, the Chi-square value is compared for each group using the standard Chi-square test. By the end of this step, the equity test of the two groups will determine whether or not they have the same survival time probability. In this study, the log-rank test is used as a validator for the previous steps as it shows whether the different road categories have the same survivability or not.

5 DATA ANALYSIS & RESULTS

As discussed, the methodology involved two major stages, as shown in **Figure 1**. The following section discusses the results of the implemented procedure, shown in detail for the year of 2019. In addition, the analysis for each year was carried out for different groups (i.e., arterials, collectors, high-traffic volume locations, and low-traffic volume locations), as illustrated in the following subsections.

5.1 Survival Analysis Results (2019)

5.1.1 *All Sites*

The methodology was applied to all sites from Jan 1, 2019, to Dec 31, 2019. There were 250 sites in Edmonton that employed *MPE* during this period. These sites experienced 573 collisions, mainly classifiable as Property Damage Only (*PDO*). In addition, there were seven major collisions and fifty-two minor collisions. The prime collision cause was following too closely, often coupled with speeding, and considered to be a speeding-related collision cause. With respect to deployed *MPE* units between two collisions, the number of deployed *MPE* visits ranged from 0 to 195, the number of deployed *MPE* hours varied between 0 to 620, and the ratios between hours and visits were estimated as low as 0 and as high as 4 hours/visit.

For the all-sites group, the Cox *PH* models were conducted, the Kaplan-Meir graphs were then plotted for different variables groups, and, finally, the log-rank tests were carried out. **Table 2 a-c** summarise the results:

5.1.1.1 Cox-proportional hazard model:

***Table 2**: The Hazard Ratio Estimates for the MPE Variables (All Sites).*

	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Hours	0.9942	0.0008	0.000	0.9925	0.995

(a) The total number of deployed *MPE* hours.

	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Visits	0.9795	0.0026	0.000	0.9742	0.9848

(b) The total number of deployed *MPE* visits.

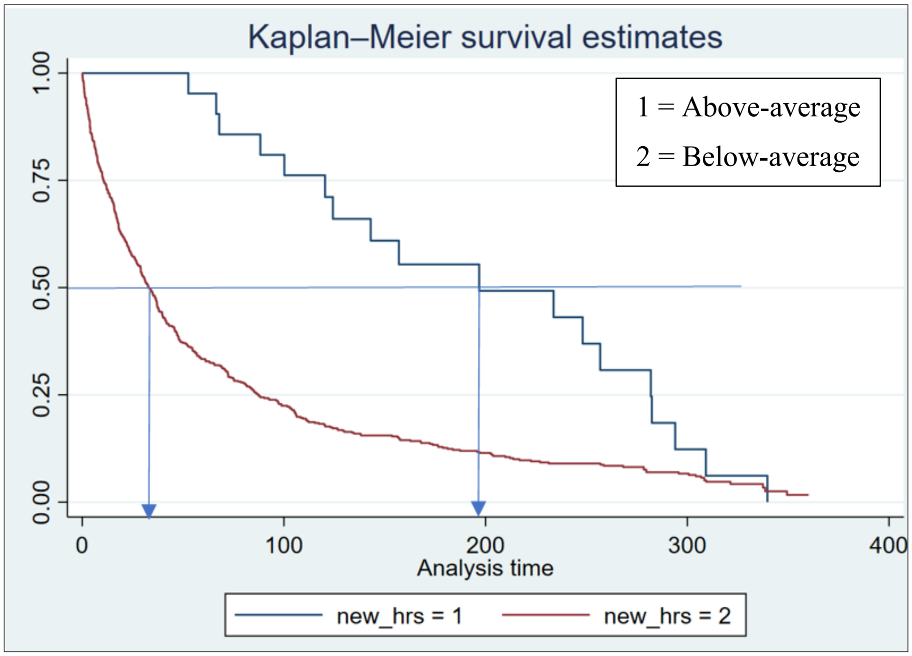
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Hours per Visits	0.7861	0.0235	0.000	0.7413	0.8336

(c) The ratio between the number of *MPE* visits and hours (*HpV*).

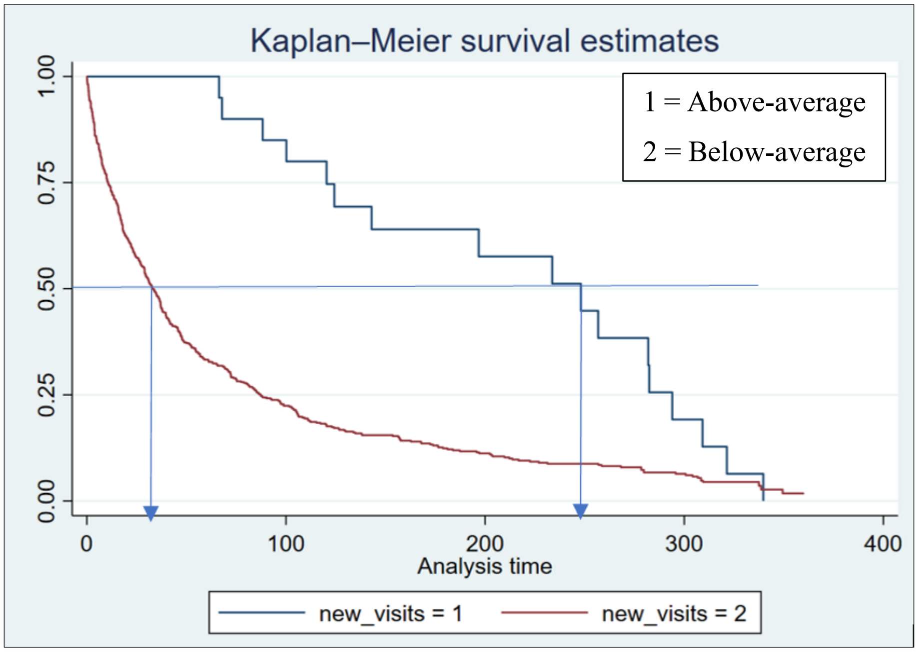
The results of this group show that there is significant evidence of *MPE* impact on reducing the duration between two consequent collisions. The deployed *MPE* hours and visits showed only small percentage reductions (1% and 3% respectively) in the collision hazards. Given the percentage difference between the two, one conclusion might be that investing in increased visits would produce better outcomes than increasing the number of hours. For instance, if a municipality invested four *MPE* hours, a more significant benefit would be to split these into distinct shorter visits. In addition, the ratio of deployed *MPE* hours to visits (*HpV*) has a hazard ratio (*HR*) of 0.78 with a reduction of 22% (i.e., (*1 – HR*) *%*) in collision occurrences for locations that had a high *HpV* compared to sites without *MPE*.

5.1.1.2 Kaplan-Meier Graphs:

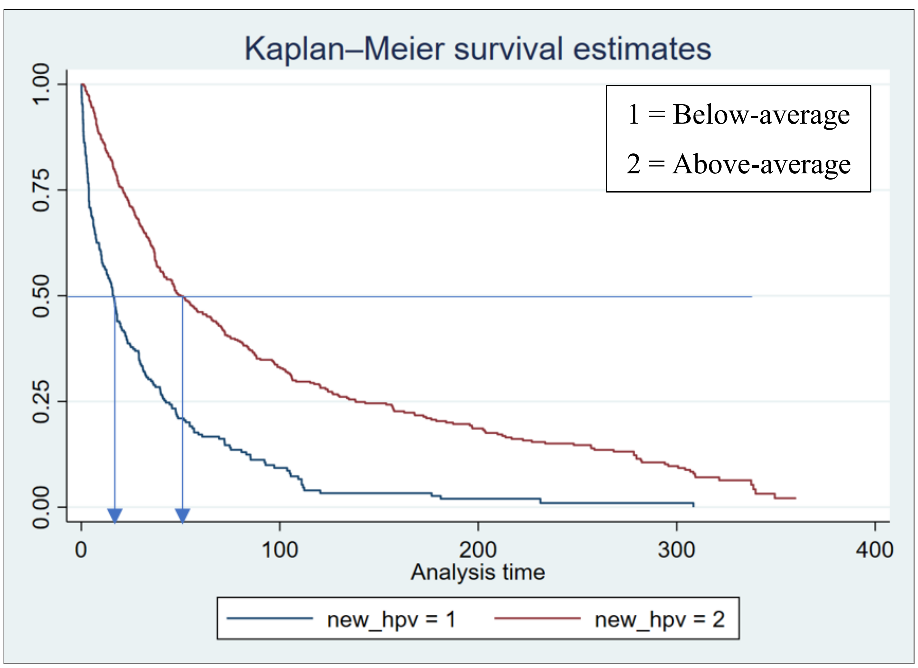
The following figures show the results of the Kaplan-Meier survival estimates for all sites.



(a) The clustered two groups of deployed *MPE* hours.



(b) The clustered two groups of deployed *MPE* visits.



(c) The clustered two groups of deployed *MPE* hours/visit.

Figure 7: *The KM Survival Estimates for the MPE Groups (All Sites).*

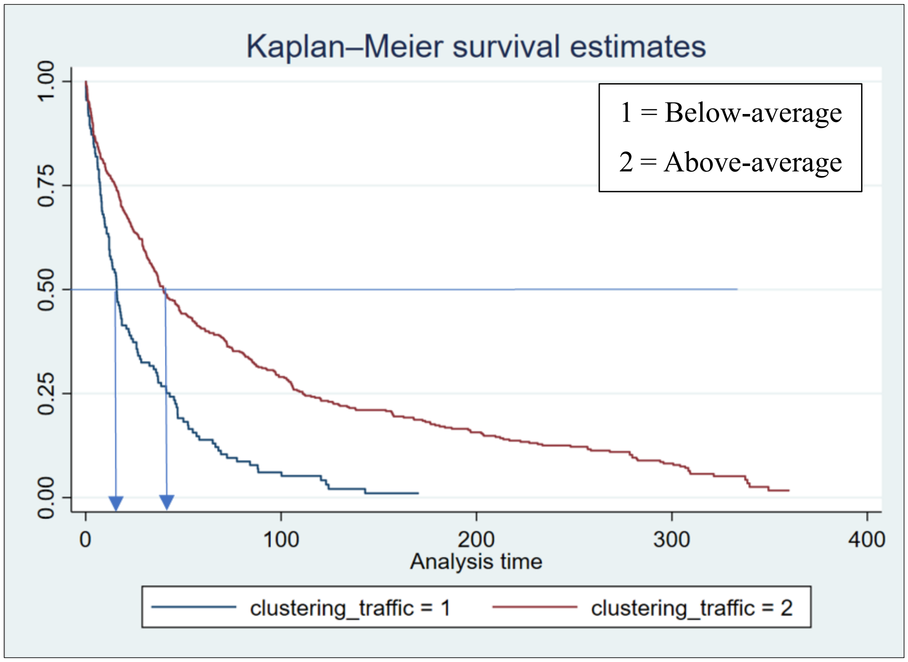


Figure 8: *The KM Survival Estimates for the Clustered Traffic Groups (All Sites).*

The *KM* graphs were plotted to compare the survival probability for different locations that experienced two levels of deployed *MPE* (i.e., above-average deployed *MPE* and below-average deployed *MPE*). The *KM* graphs yielded the same results regarding the deployed *MPE* hours and visits, showing that the locations that experienced high *MPE* hours or visits (above average) have higher survivability than those with lower *MPE* hours or visits (below average). As shown in **Figure 7 a**, the median survival probability (at 0.5 on *Y*-axis) for sites with above-average deployed *MPE* hours is 198 days. These days represent the cumulative survival time in days for the above-average deployed *MPE* hours at the median. In comparison, the below-average ones had a survival probability of 38 days. Similarly, in **Figure 7 b**, the survival probability at 0.5 on *Y*-axis for sites that had above-average *MPE* visits is 248 days; meanwhile, the below-average group is 37 days. Moreover, groups with below-average *MPE* hours and visits encountered more frequent steps in the *KM* graphs, indicating that these sites experienced more collisions over a shorter time than the above-average *MPE* groups.

Figure 7 c shows that, for the clustered *MPE* hours/visit groups, the sites that experienced a higher ratio of the deployed *MPE* hours to visits have more survival probability than the below-average locations. **Figure 8** illustrates that above-average traffic volume locations have higher survivability than the below-average sites. This difference may be because drivers tend to speed up, exceeding speed limits, when roadways have less or no congestion.

5.1.1.3 Log-rank Equality Test:

Table 3: *The Log-Rank Test for Different Groups (All Sites).*

Hours	Observed Events	Expected Events
Above-average	18	41.37
Below-average	555	531.63
Total	573	573.00
Chi2(1) = 14.61 Pr>chi2 = 0.0001		

a) Log-rank test for *MPE* hours groups

Visits	Observed Events	Expected Events
Above-average	17	41.74
Below-average	556	531.26
Total	573	573.00
Chi2(1) = 14.34 Pr>chi2 = 0.0001		

b) Log-rank test for *MPE* visits groups

HpV	Observed Events	Expected Events
Below-average	228	126.12
Above-average	345	446.88
Total	573	573.00
Chi2(1) = 111.63 Pr>chi2 = 0.0000		

c) Log-rank test for *MPE* hours/visit groups

Traffic Count	Observed Events	Expected Events
Below-average	123	66.91
Above-average	450	506.09
Total	573	573.00
Chi2(1) = 55.29 Pr>chi2 = 0.0000		

d) Log-rank test for traffic volume groups

The log-rank tests were to establish whether two different groups have the same survivability (i.e., null hypothesis) and whether those comply with the results from the *KM* graphs. Thus, if the *Pr* value is less than 0.05, the tested groups do not have the same survival probability.

As shown in **Table 3**, all of the tested groups had *Pr* < 0.05, proving the previous results of the *KM* graphs.

5.1.2 *Arterial & Collector Roads*

In this subsection, only arterial and collector sites are analyzed. This category was tested to examine the impact of *MPE* on the main roadway categories. There were 111 arterial and collector sites data available in 2019. These locations had 566 collisions, the main reasons for which were following too closely and left turn crossing path. These collisions resulted in seven major and fifty-two minor crashes.

5.1.2.1 Results Discussion

The outcomes of the *Cox PH* models provide almost the same results for all sites. The results for all the *MPE* variables are significant at a 95% confidence interval and had *HR* values less than 1, reflecting the positive impact of deployed *MPE* variables. The *HR* for the deployed *MPE* hours/visit surpassed other *MPE* variables as it resulted in a reduction in collisions occurrence of 22%.

The *KM* survival estimates for the arterial and collector locations yielded the same conclusion for all sites' results and followed similar trends. For instance, sites that experienced above-average *MPE* hours, visits, or *HpV* possessed a higher survival probability other than the below-average category. In addition, the below-average traffic volume locations were at higher risk of collision occurrence.

The log-rank tests examined the equity of the survival probability for the clustered groups. The *Pr* values for all tests were significant (i.e., *Pr* < 0.05). Thus, the log-rank test results compare favourably with the outcomes of *KM* graphs.

5.1.3 *Arterial Sites Only*

The arterial sites were tested separately to examine the impact of *MPE* variables. Sixty-four arterial sites recorded 436 collisions in 2019. These collisions' severity was mainly *PDO*, with four major and thirty-eight minor collisions. The leading causes for the arterial road collisions were following too closely, left turn cross-path, and improper lane change. Furthermore, the deployed *MPE* hours and visits between collisions ranged from 0 to 620 and 0 to 195, respectively.

5.1.3.1 Results Discussion

The results of the *Cox PH* models showed that the *MPE* variables had a proactive effect on collision occurrence in comparison to locations that were not exposed to *MPE* deployment. For example, the *deployed MPE hours/visit HR* was 0.78, which means an expected reduction in the risk of collision occurrence of 22%. These results are significant at a 95% confidence interval.

The *KM* survival graphs, emphasize the overall conclusion that the higher the level of deployed *MPE*, the less risk of collisions occurrence. To illustrate, locations with above-average *MPE* hours had an average survival probability of 231 days, while below-average sites survived for an average of 27 days. In other words, the below-average *MPE* locations experienced more frequent collisions. Moreover, the survival probability for above-average traffic volume locations was higher than for below-average traffic volume sites, possibly due to changes in drivers' propensity for speeding that otherwise may lead to a speeding-related collision.

Furthermore, the log-rank tests that examine the survival probability equity for two different groups were significant as the $Pr<0.05$ for all the test groups. Therefore, the survival probabilities for each of the two tested groups were not equal, both proving and matching the previous results of the *KM* graphs.

5.1.4 *Collector Sites Only*

In this section, the effectiveness of the deployed *MPE* variables is examined to explore whether the *MPE* is more efficient for collector roads. Forty-eight collector sites had 134 collisions in 2019, mainly classified as *PDO*. In addition, there were two major and fourteen minor collisions. The deployed *MPE* hours between two crashes varied between 0 and 280 hours, and the number of deployed *MPE* visits ranged from 0 to 86. In addition, the ratio between the number of deployed *MPE* hours to visits reached four hours/visit.

5.1.4.1 Results Discussion

The analysis of the collector sites data showed that the *MPE* variables' impacts on the duration between collisions are more beneficial than other categories (i.e., all sites, arterials and collectors, and arterials only groups). For instance, the deployed *MPE* hours/visit variable had $HR = 0.65$. Therefore, the consequence of accounting for the *MPE* hours/visit is a reduction in the risk of a collision occurrence by 35%. Also, the deployed *MPE* visits trigger a decrease of 5% in collisions.

The *KM* graphs results provided the same summary as previously explained for different categories. For example, the median survival probability (at 0.5 on *Y*-axis) for above-average *MPE* hours/visit locations is 265 days, compared to 34 days for the sites in the below-average group. Moreover, it is noticeable that the *KM* graph for the traffic volume clusters intersects at many points and yields almost the same survival probability over the year. This indicates that the traffic volume does not impact the survival probability for the collectors' sites. In other words, the traffic volume is not a factor that affects the survival probability of collector roadways.

The log-rank tests were not significant for all groups (i.e., *Pr* value> 0.05 in one case). For instance, the $Pr < 0.05$ for the *MPE* hours, visits, and hours/visit groups indicate that these groups have different survival probabilities. As expected, the *Pr* value was greater than 0.05 for the traffic volume groups, yielding the same results as the *KM* graphs. This indicates that the null hypothesis was not rejected, and both traffic volume clusters have similar survival probability.

5.1.5 *High Traffic Volume Locations Only*

All sites were classified based on roadway type to study the impact of *MPE* variables on different roadway categories. They were then classified based on the traffic volume to compare the effectiveness of deployed *MPE* variables on two different traffic volume categories (i.e., above-average and below-average traffic volume sites). Ten high-traffic volume sites experienced 133 collisions in 2019. These sites were all arterial roadways.

5.1.5.1 Results Discussion

The procedure's outputs for the high traffic volume sites indicated that the different *MPE* variables impacted the duration between two consequent collisions. All *MPE* variables significantly correlated to the period between collisions (95% confidence interval). The ratio of deployed *MPE* hours to visits (*HpV*) showed a remarkable influence on the collision occurrence. The hazard ratio for this variable is 0.48, indicating a 52% reduction in the risk of a collision. Moreover, the deployed *MPE* hours and visits had an *HR* of 0.97 and 0.91, respectively. This reflects a decrease in collisions by 3% and 9% by implementing *MPE* hours and visits, respectively. These results comply with the previous conclusion that deploying *MPE* visits is more effective than deploying *MPE* hours.

The *KM* charts' results showed that above-average *MPE* sites had a higher survival probability than below-average *MPE* sites. Notably, when the survival probability was at 0.5 on *Y*-axis, the above-average *MPE* hours group was 91 days while the below-average group was 15 days. Similarly, the survivability for the above-average *MPE* hours/visit cluster is at an average of 21 days, and the below-average is 1 day. This result continues to demonstrate that higher *MPE* hours, visits, or hours/visit are linked to higher survivability and lower risk of collision occurrence, along with correspondingly the longer durations between two consequent collisions.

With respect to the log-rank test results, the equity tests were significant for all considered groups. The *Pr* value was less than 0.05, which means the null hypothesis was rejected, and the clusters of each group have different survival probabilities. These results support the *KM* graphs by providing the same conclusion.

5.1.6 *Low Traffic Volume Locations Only*

One hundred four low-traffic volume sites had 450 collisions in 2019. The categories of these sites varied between arterials, collectors, and locals, with a balanced presence of both arterial and collectors' categories. The leading cause of these collisions was following too closely, followed by left turn cross path, and stop sign violations.

5.1.6.1 Results Discussion

The Cox-proportional hazard models for the below-average traffic volume sites showed that the deployed *MPE* hours, visits, and *HpV* significantly reduced collision risk. For instance, the ratio of deployed *MPE* hours to visits had the highest impact on collision rates with an *HR* = 0.7, which means a 30% reduction in the risk of collisions when considering the *MPE HpV*. Moreover, the deployed *MPE* visits have a higher positive impact than *MPE* hours since it reduces the risk of collision by 3%.

The clustered groups of *MPE* hours, visits, and *HpV* have different survivability for the *KM* graphs. The survival probability for the above-average *MPE* hours is 257 days, compared to forty-five days for the below-average group, in addition, similar survivability for the deployed *MPE* visits groups. Furthermore, for the *HpV MPE* clusters, the above-average sites survived for 138 days on average. On the other hand, the below-average sites survived for twenty-four days.

The log-rank tests were significant for all variables' groups. The *Pr* values for *MPE* hours, visits, and *HpV* clusters are zero, emphasizing these groups' different survivability. These results match the *KM* graphs' outcome.

5.1.7 *Speeding Related Collisions Only*

The speeding-related collisions are defined as the collisions that happened due to following too closely, run-off-road, or striking a parked vehicle. Two hundred seventeen speeding-related collisions took place in 2019. This study considered the impact of deployed *MPE* variables on these collisions.

5.1.7.1 Results Discussion

The results of the speeding-related collisions proved that deploying a higher rate of hours to visits increases the period between two consequent collisions, hence, decreasing the probability of collision occurrence. Comparing the results of all site analyses and the speeding-related collisions only shows that the deployed *MPE* is more effective on these specific collision causes. For instance, the *MPE HpV*'s *HR* for the speeding-related collisions is 0.67, which means a 33% reduction in the risk of collision occurrence by accounting for the *MPE HpV* ratio. On the other hand, the *HR* for all collisions is generally 22%.

The *KM* graphs showed that the average survival probability for above-average *MPE* hours or visits sites is 220 days, while it is seventeen days for below-average *MPE* locations. This difference means that the survivability for the above-average *MPE* hours or visits sites is more than ten times higher than that of the below-average *MPE* sites. For the deployed *MPE* hours/visit, the survival probability for the below-average site group is ten days, compared to forty-five days for the above-average *MPE* hours/visit sites.

Moreover, the log-rank tests are proof of the *KM* graphs results. The *Pr* values for all clustered groups are less than 0.05, reflecting that these different clusters do not have the same survivability. For instance, the *Pr* equals 0.0026 for the clusters of *MPE* hours per visit. Thus, these results comply with the *KM* graphs.

5.2 Survival Analysis Results (2018, 2017, 2016, and All years)

The methodology was applied to each year (2018, 2017, 2016) separately, then integrated into one study. The results of each year and the combined years yield the same conclusion that the *MPE* variables have a considerable impact on collision occurrence, whereby there is an increase in the time between two collisions and a corresponding decrease in the risk of collision. To sum up the results, the positive effect of the deployed number of *MPE* visits is greater than that of the number of deployed *MPE* hours. Moreover, the ratio between *MPE* hours to visits has the most influence on reducing the hazard of collision occurrence. Through applying the analysis procedure to different road categories and traffic volume classifications, the outcomes of these processes provided similar results to those for the 2019 analysis. The following table summarizes the hazard ratio for the optimal *MPE* variable (i.e., hours/visit) for all years.

Table 4: *Hazard Ratios for Deployed MPE Hours per Visit from 2016 to 2019 and All Years.*

Year	2019		2018		2017		2016		All years	
Site Classification	Hazard Ratio	Reduction in Collision Hazard (%)	Hazard Ratio	Reduction in Collision Hazard (%)	Hazard Ratio	Reduction in Collision Hazard (%)	Hazard Ratio	Reduction in Collision Hazard (%)	Hazard Ratio	Reduction in Collision Hazard (%)
All Sites	0.78	22	0.90	10	0.85	15	0.72	28	0.82	18
Arterial Only	0.78	22	0.74	26	0.74	26	0.71	29	0.75	25
Collectors Only	0.65	35	0.65	34	0.68	32	0.64	36	0.66	36
High Traffic Volume	0.48	52	0.52	48	0.63	37	0.45	55	0.52	48
Low Traffic Volume	0.70	30	0.78	22	0.72	28	0.68	32	0.72	28
Speeding-Related Collisions	0.67	33	0.72	28	0.71	29	0.66	34	0.69	31

6 CONCLUSIONS

This chapter summarizes the research conclusions and includes a discussion of the contribution. It also outlines research limitations and potential areas for future research.

6.1 Research Summary

The main goal of this research has been to develop survival analysis models that investigate the impact of deployed *MPE* hours, visits, and *HpV* on the duration between two consequent collisions, thereby helping authorities effectively mitigate the issue of speed violators who cause different types of collisions. *MPE* programs are considered an effective solution to restrict irresponsible drivers' behaviour. Therefore, this research explored the survival probability of various locations exposed to different *MPE* hours and visits. Moreover, it detected the potential reduction of the risk of collision occurrence by understanding the effect of deploying *MPE* hours and visits.

The proposed methodology consisted of two phases: preparing the provided data and applying the survival analysis. The output of the first step was a Microsoft Excel spreadsheet that contained the necessary survival data. Thus, the outcome table combined the duration between two consequent collisions and the corresponding number of *MPE* hours and visits in each time period. Moreover, the outcome contained collision causes, traffic count, and the first and second collision dates. Next, the data were categorized into separate groups based on the *MPE* variables and traffic counts. These groups were classified using *K*-means clusters, and MATLAB and SAS software were utilized to execute this step. The second step was to apply the survival analysis, including Cox proportional hazard models, *KM* survival estimates, and log-rank tests. The Cox proportional hazard models investigated the impact of *MPE* hours, visits, and *HpV* on the duration between two consequent collisions. In other words, it explored the expected reduction in collision occurrence by deploying *MPE* hours and visits. The *KM* graphs were plotted on different *MPE* clusters to emphasize the influence of the deployed number of *MPE* hours, visits, and *HpV*. These clusters were classified into two groups (i.e., above average and below average) based on *MPE* hours, *MPE* visits, *MPE HpV*, and traffic count. Finally, the log-rank tests were conducted to examine the clusters' survival probability, which was subsequently expected to comply with the *KM* graphs' conclusion. All steps were executed on different locations' groups.

The results showed that accounting for the ratio between hours and visits had the most impact on increasing the period between collisions and reducing the risk of collision. The expected reduction in the collision hazard varied between 52% and 22%, where the maximum reduction could be expected in high traffic volume locations, and the minimum reduction could be expected for all locations. In addition, the deployed *MPE HpV* had a better effect on collector roads compared to arterial roads. Also, it was noted that the number of deployed *MPE* visits had a higher impact on increasing the duration between collisions than the number of deployed *MPE* hours. Those planning for future use of the *MPE* program should consider this by providing more timing options for critical deployments.

The *KM* survival estimates were plotted for different groups of *MPE* hours, visits, *HpV,* and traffic volume. These graphs concluded that the groups of above-average *MPE* hours, *MPE* visits, and *MPE HpV* have higher survivability than the below-average ones. These results emphasize the importance of *MPE* deployment to reduce the risk of collisions occurring. The *KM* survival graphs were carried out for traffic count clusters. The results showed that the above-average traffic volume locations had higher survivability than those with below-average traffic volume, possibly because drivers tend to exceed the speed limit when the traffic flow is light. In addition, the log-rank tests were done, and the outputs comply with the conclusions drawn from *KM* survival estimates. Finally, the same methodology was applied to the data from 2016 to 2019.

6.2 Research Contributions

This thesis presented multiple contributions to the study of Mobile Photo Enforcement (*MPE*) use for controlling collision occurrence. The contributions are provided below.

1. Create a better understanding of the impact of using the *MPE* on the duration between two consequent collisions for different road types. The impact was explored by applying survival analysis to have a general idea of the influence of *MPE* variables (i.e., *MPE* hours, *MPE* visits, and *MPE HpV*) on road collision frequency. This investigation concluded that the MPE variables positively affect the duration between two consequent collisions, thus improving road safety. However, although all the *MPE* variables have acceptable probabilities of reducing the collision occurrence, the deployed *MPE HpV* demonstrates the most impact. Therefore, knowledge can be practically applied when allocating *MPE* deployment.
2. Recognize the relationship between the deployed *MPE* variables and the duration between collisions on different road hierarchies, i.e., arterial and collector roads. The most

significant finding is that *MPE* deployment was more impactful on collector locations compared to arterial roads. This conclusion helps in deploying the *MPE* units more efficiently. Therefore, better, more efficient utilization of the available *MPE* resources could be deployed.

3. Identify the impact of deploying *MPE* variables on the duration between collisions on high and low traffic volume sites. *MPE* variables were shown to be more potent in low traffic volume sites. As a result, the *MPE* programs should be allocated accordingly to benefit from it.
4. Predict the influence of *MPE* deployment on speeding-related collisions only. This finding emphasizes the importance of the *MPE* program in improving road safety by increasing the duration between two consequent collisions.
5. Estimate the expected reduction in road collisions as a consequence of deploying *MPE* variables for each aforementioned contribution. In other words, the expected survivability percent for different road categories and conditions is estimated and illustrated.

In conclusion, this thesis employed statistical analysis to comprehend the relationship between the deployed *MPE* variables; hours, visits, and hours/visits, and the duration between two consequent collisions. This analysis accounted for different road types and conditions. Moreover, the outcomes of this analysis fill a literature gap and can be practically adopted by transportation planners and authorities to improve road safety and allocate the *MPE* resources properly.

6.3 Limitations and Future Research

This research covered the previously discussed gaps in the literature. However, further investigation and improvements would be expected. Moreover, this research has limitations and assumptions. For instance, this research was limited to exploring the impact of one *MPE* variable (i.e., hours, visits, or *HpV*) at a time using the Cox proportional hazard model. The results, then, only show an understanding of the influence of a single variable without considering the presence of other variables. This limitation can be addressed by exploring the impact of multi-variables in the Cox proportional hazard model.

Moreover, this thesis analyzes the impact of deployed *MPE* variables on road collisions for different road types and conditions. However, there is a need to investigate the influence of deploying *MPE* on other road characteristics. For example, sites can be categorized based on the posted speed limit, which might provide significant results in understanding the drivers' behaviour and the influence of *MPE* variables on reducing the risk of getting involved in road crashes by increasing the duration between consequent collisions. In addition, locations could be categorized to account for roads' number of lanes. This classification might assist in exploring the effect of *MPE* variables on road collisions in sites where drivers have more lanes in which to maneuver.

Furthermore, this research can be expanded to include studying the impact of deployed *MPE* variables on the duration between collisions during different seasons (i.e., Fall, Winter, Spring, and Summer). This analysis can be implemented by categorizing the collisions into four groups based on the occurrence season and applying the same methodology to each group. The outcomes of this proposed analysis might lead to a better understanding of the optimal timing for deploying the *MPE*.

In summary, the limitations of this thesis can act as the starting point for further research. This thesis studied the correlation between deployed *MPE* variables and the duration between consequent collisions using a single-variant model. Thus, it is recommended to analyze this relationship using multi-variants Cox proportional models. Moreover, this thesis used different road types and conditions, and it is encouraged to explore the impact of *MPE* variables on road collisions based on other classifications.

REFERENCES

1. World Health Organization, Global status report on road safety: supporting a decade of action, Geneva, 2015.
2. Mohan, D., Khayesi, M., Tiwari, G. and Nafukho, F.M., 2006. *Road traffic injury prevention training manual*. World Health Organization.
3. Tay, R. (2010). Speed Cameras, Improving Safety or Raising Revenue?. Journal of Transport Economics and Policy, 44, pp.247–257
4. Sharma, B. (2008). Road traffic injuries: A major global public health crisis. *Public Health*, 122, 1399–1406.
5. RICHARDS, S. H., WUNDERLINCH, R. C., and DUDEK, C. L., Field Evaluation of Speed Control Techniques. *Transportation Research Record 1035. Transportation Research Board*, National 30 Research Council, Washington, DC, pp. 66-78, 1985.
6. Li, R., El-Basyouny, K. and Kim, A. (2014). *A City-Wide Safety Analysis of Mobile Speed Enforcement*. [online] City of Edmonton. Available at: https://www.edmonton.ca/transportation/Neighbourhoods/City_Wide_Safety_Analysis_of_Mobile_Speed_Enforcement.pdf [Accessed Jul 18. 2018].
7. Islam, M., El-Basyouny, K. and Ibrahim, S. (2014). The impact of lowered residential speed limits on vehicle speed behavior. *Safety Science*, 62, pp.483-494.
8. Kloeden, C. N., McLean, A. J., & Glonek, G. (2002). *Reanalysis of travelling*

9. Organisation for Economic Cooperation and Development, Speed Management, France, 2006.

10. NHTSA, N. H. T. S. A. (2015). Traffic safety facts 2013 data.

11. Transport Canada. (2015). Canadian Motor Vehicle Traffic Collision Statistics 2013.

12. Zaal, D. (1994). Traffic law enforcement: *A review of the literature* (Report No.53). Melbourne: MUARC.

13. Kim, A.M., Wang, X., El-Basyouny, K. and Fu, Q., 2016. Operating a mobile photo radar enforcement program: A framework for site selection, resource allocation, scheduling, and evaluation. *Case Studies on Transport Policy*, *4*(3), pp.218-229.

14. Alberta Justice and Solicitor General. (2014). Automated Traffic Enforcement Technology Guidelines. Edmonton: Province of Alberta Government.

15. Li, Y., 2019. Planning a Mobile Photo Enforcement Program by Mapping Program Goals to Deployment Decisions.

16. Li, R., El-Basyouny, K., Kim, A. and Gargoum, S., 2017. Relationship between road safety and mobile photo enforcement performance indicators: A case study of the city of Edmonton. *Journal of Transportation Safety & Security*, *9*(2), pp.195-215.

17. Rodier, C.J., Shaheen, S.A. and Cavanagh, E., 2007, July. Automated speed enforcement in the US: a review of the literature on benefits and barriers to implementation. In *Transportation Research Board 87th Annual Meeting. CD-ROM. Washington, DC*.

18. Evans, L., 2004. Traffic safety science serving society. *Bloomfield Hills, MI*, *179*.

19. Elvik, R., 2005. Speed and road safety: synthesis of evidence from evaluation studies. *Transportation Research Record*, *1908*(1), pp.59-69.

20. Helai, H., Chor, C.H., Haque, M.M. (2008). Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis. Accident Analysis and Prevention 40: 45–54

21. Vanlaar, W., Robertson, R. and Marcoux, K., 2011. Evaluation of the Photo Enforcement Safety Program of the City of Winnipeg Final Report. Pp. 11-17

22. Brault, M., Auger, A. and Montégiani, M., 2007, June. La sécurité aux intersections: Analyse des comportements des conducteurs au feu rouge. In *Présentation à la XVIIe Conférence Canadienne Multidisciplinaire sur la Sécurité Routière à Montréal* (pp. 3-6).

23. Carnis, L. and Blais, E., 2013. An assessment of the safety effects of the French speed camera program. *Accident Analysis & Prevention*, *51*, pp.301-309.

24. Vanlaar, W., Robertson, R. and Marcoux, K., 2011. Evaluation of the Photo Enforcement Safety Program of the City of Winnipeg Final Report. Pp. 11-17.

25. Li, R. and Guo, M., 2015. Competing risks analysis on traffic accident duration time. *Journal of advanced transportation*, *49*(3), pp.402-415.

26. Carnis L. Automated speed enforcement: What the French experience can teach us. Journal of Transportation Safety & Security. 2011 Mar 8;3(1):15-26.

27. Bonneson JA, Zimmerman KH. Effect of yellow-interval timing on the frequency of red-light violations at urban intersections. Transportation Research Record. 2004;1865(1):20-7.

28. Li R, El-Basyouny K, Kim A. Before-and-after empirical Bayes evaluation of automated mobile speed enforcement on urban arterial roads. Transportation research record. 2015 Jan;2516(1):44-52.

29. Hojati, A.T., Ferreira, L., Washington, S. and Charles, P., 2013. Hazard based models for freeway traffic incident duration. *Accident Analysis & Prevention*, *52*, pp.171-181.

30. Shi, Y., Zhang, L., and Liu, P. 2015. Survival Analysis of Urban Traffic Incident Duration: a Case Study at Shanghai Expressways. Journal of Computers, 26(1).

31. Nam, D., and Mannering, F. 1999. An exploratory hazard-based analysis of highway incident duration. Transportation Research Part A: Policy and Practice, 34(2): 85-102. doi: 10.1016/S0965-8564(98)00065-2.

32. Yang, X., Gao, Z., Guo, H., and Huan, M. 2012. Survival analysis of car travel time near a bus stop in developing countries. Science China Technological Science, 55(8). DOI: https://doi.org/10.1007/s11431-012-4900-z.

33. Lawpoolsri, S., Li, J., and Braver, E.R. 2007. Do Speeding Tickets Reduce the Likelihood of Receiving Subsequent Speeding Tickets? A Longitudinal Study of Speeding Violators in Maryland. Traffic Injury Prevention, 8(1): 26-34.

34. Wu J, Zhang S, Singh AK, Qin S. Hazard-Based Model of Mandatory Lane Change Duration. InCICTP 2017: Transportation Reform and Change—Equity, Inclusiveness, Sharing, and Innovation 2018 Jan 18 (pp. 805-811). Reston, VA: American Society of Civil Engineers.

35. Kang, G. and Fang, S.E., 2011. Applying survival analysis approach to traffic incident duration prediction. In *ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation* (pp. 1523-1531).

36. MacQueen, James. "Some methods for classification and analysis of multivariate observations." Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Vol. 1. No. 14. 1967.

37. Carroll R, Lin X. Non-parametric and semi-parametric regression methods. Chapman & Hall/CRC Handbooks of Modern Statistical Methods Longitudinal Data Analysis. London: UK; 2008. p. 191-7.

38. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53: 457–481.

39. Schober, P. and Vetter, T.R., 2018. Survival analysis and interpretation of time-to-event data: the tortoise and the hare. *Anesthesia and analgesia*, *127*(3), p.792.

APPENDIX

➢ **Arterial and Collector Roads in 2019**

- Cox Proportional Hazard Model:

	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Hours	0.99414	0.00085	0.000	0.99247	0.9958

(a) The total number of deployed *MPE* hours.

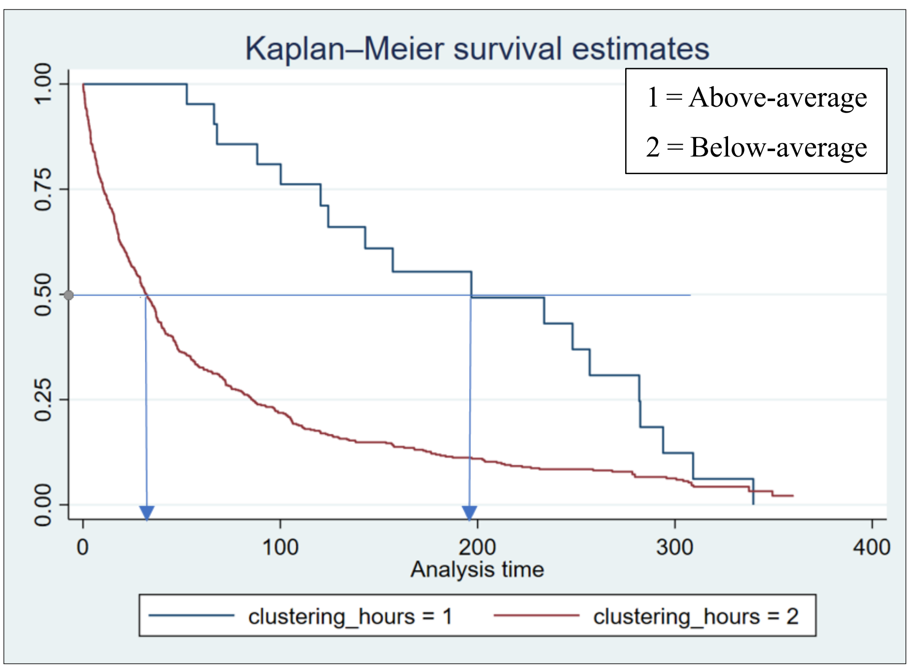
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Visits	0.9791	0.00275	0.000	0.9737	0.9845

(b) The total number of deployed *MPE* visits.

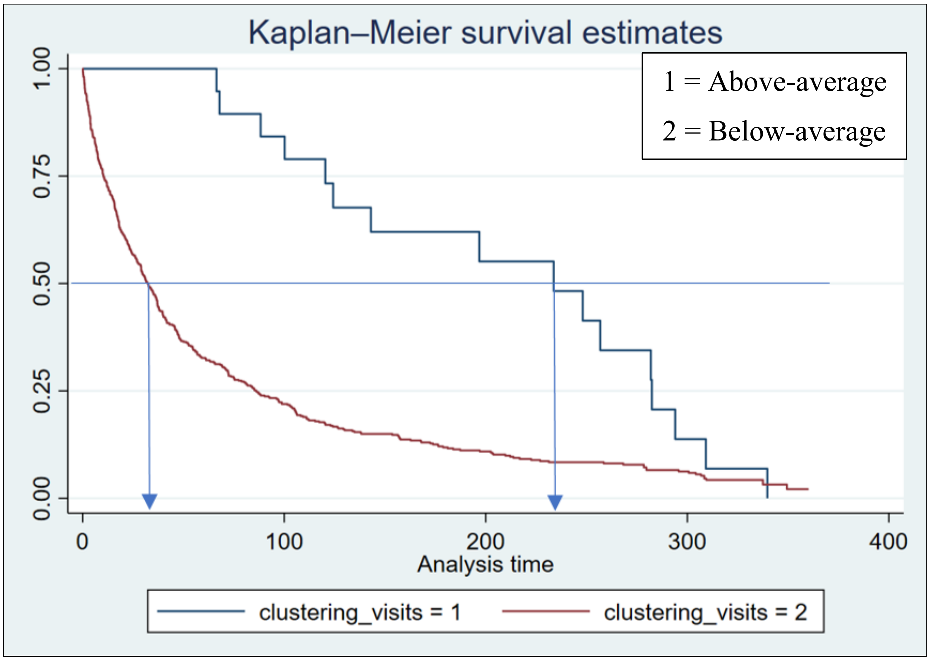
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Hours per Visits	0.78519	0.0234	0.000	0.7405	0.8324

(c) The ratio between the number *MPE* of visits and hours (*HpV*).

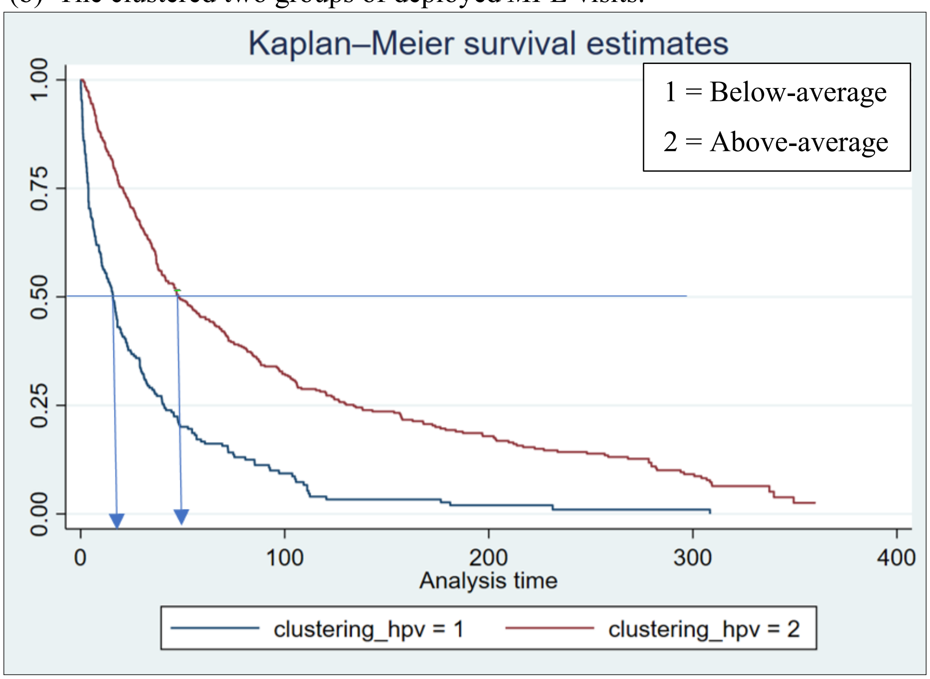
- KM Graphs



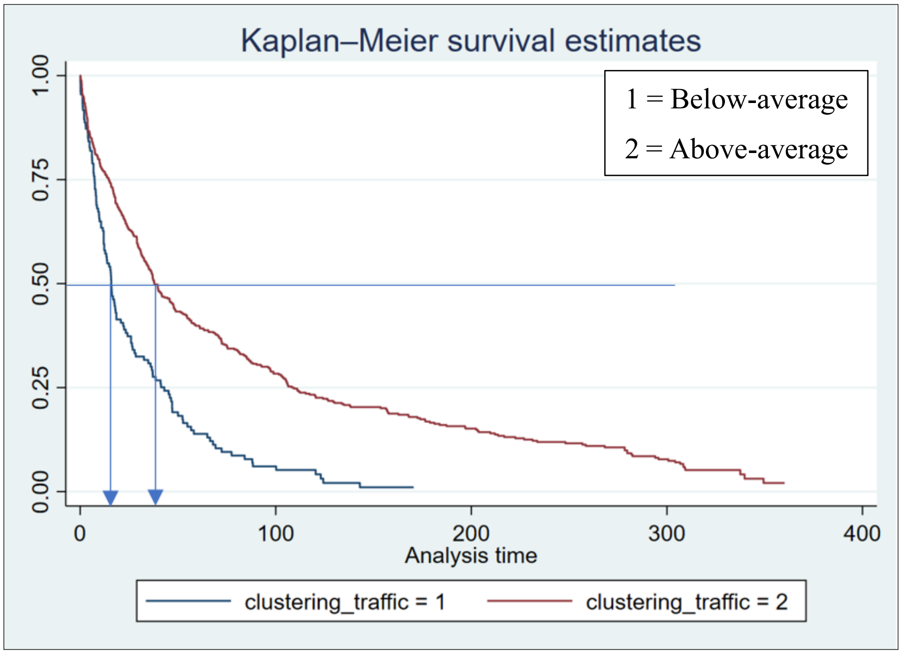
(a) The clustered two groups of deployed *MPE* hours.



(b) The clustered two groups of deployed *MPE* visits.



(c) The clustered two groups of deployed *MPE* hours/visit.



(d) The clustered two groups of traffic volume

- Log-Rank Equality Test

Hours	Observed Events	Expected Events
1	18	41.97
2	548	524.03
Total	566	566.00
		Chi2(1) = 15.21
		Pr>chi2 = 0.0001

a) Log-rank test for *MPE* hours groups.

Visits	Observed Events	Expected Events
1	16	39.15
2	550	526.85
Total	566	566.00
		Chi2(1) = 15.14
		Pr>chi2 = 0.0001

b) Log-rank test for *MPE* visits groups.

HpV	Observed Events	Expected Events
1	225	124.59
2	341	441.41
Total	566	566.00
		Chi2(1) = 109.52
		Pr>chi2 = 0.0000

c) Log-rank test for *MPE* hours/visit groups.

Traffic Count	Observed Events	Expected Events
1	123	68.26
2	443	497.74
Total	566	566.00
		Chi2(1) = 51.82
		Pr>chi2 = 0.0000

d) Log-rank test for traffic volume groups.

➢ **<u>Arterial Sites Only in 2019</u>**

- Cox Proportional Hazard Model

	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Hours	0.9939	0.00085	0.000	0.99230	0.99567

(a) The total number of deployed *MPE* hours.

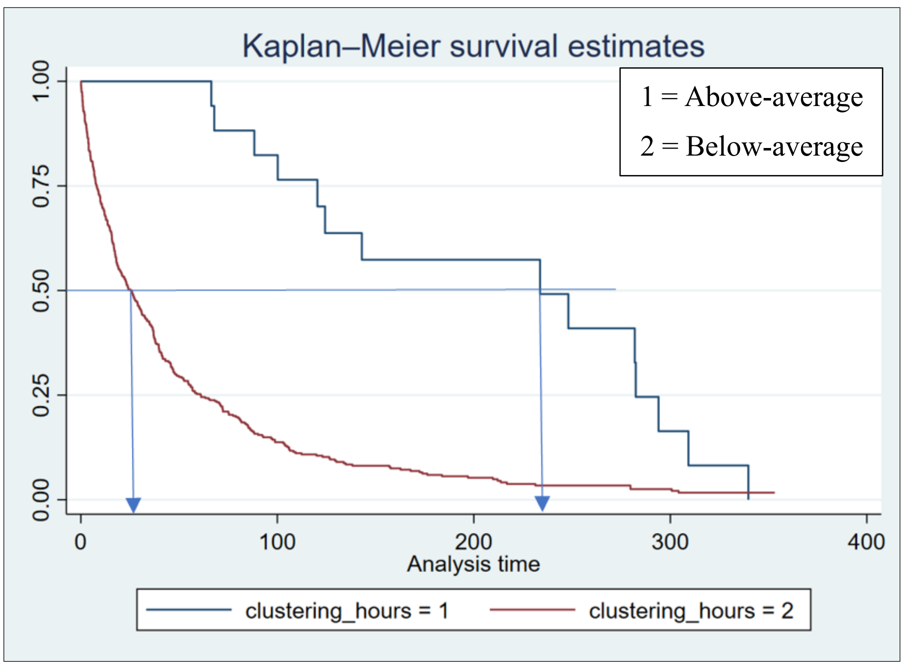
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Visits	0.9798	0.00278	0.000	0.9743	0.9852

(b) The total number of deployed *MPE* visits.

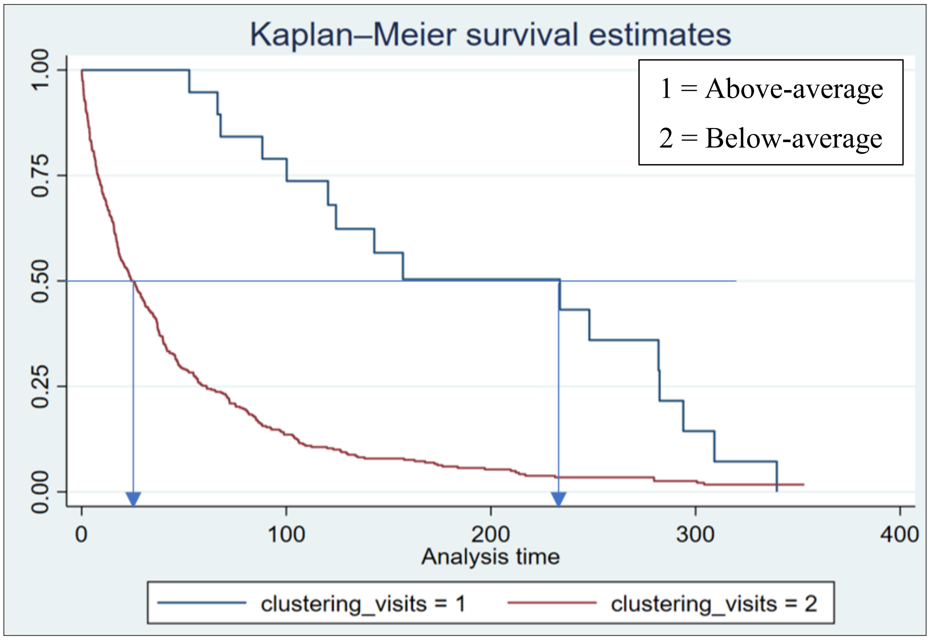
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Hours per Visits	0.78457	0.0255	0.000	0.73609	0.8362

(c) The ratio between the number of visits and hours (*HpV*).

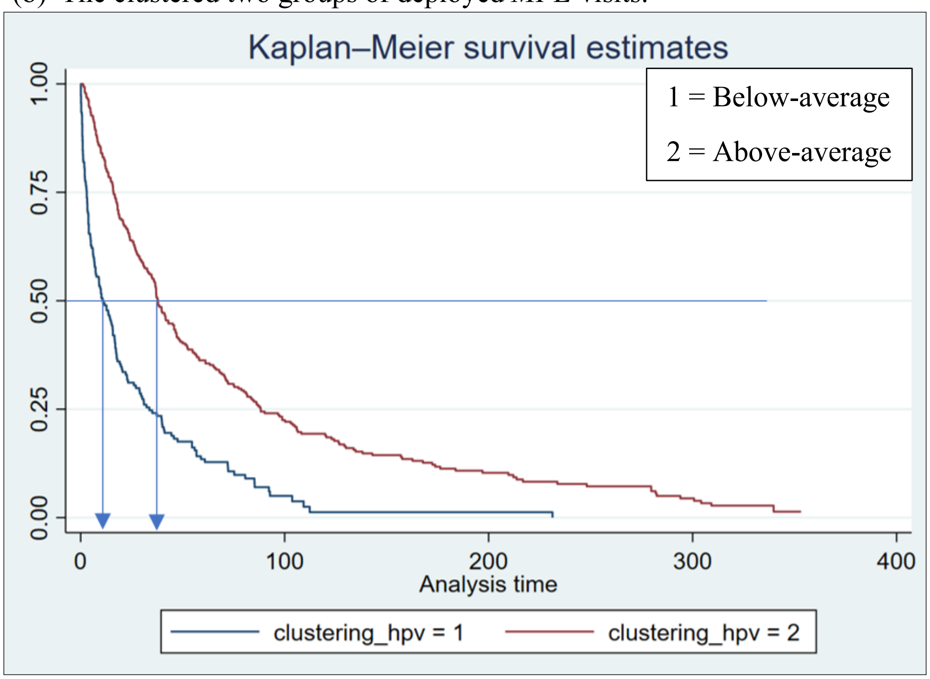
- KM Graphs



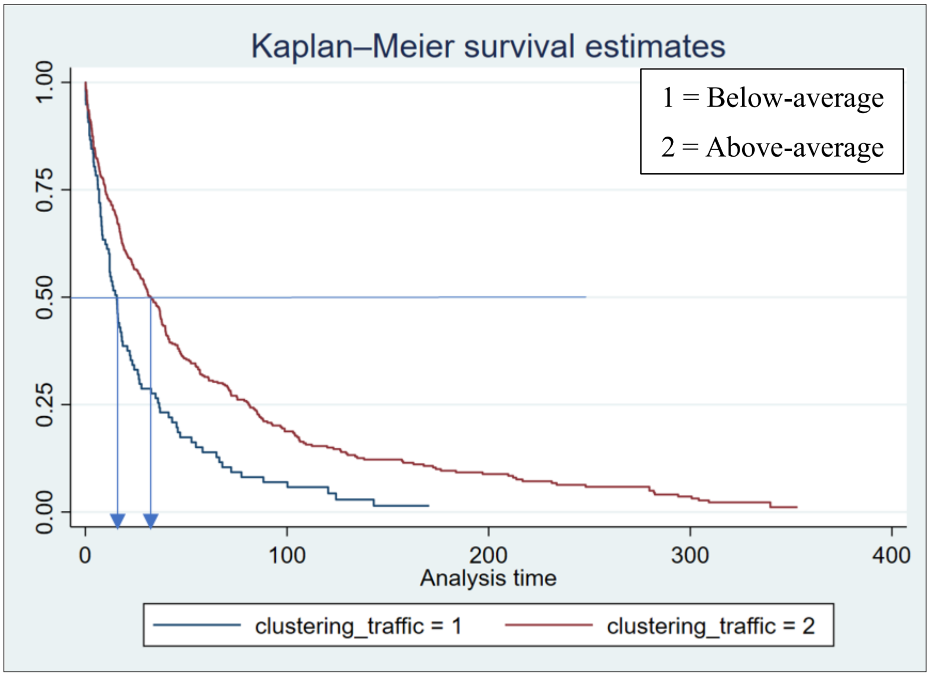
(a) The clustered two groups of deployed *MPE* hours.



(b) The clustered two groups of deployed *MPE* visits.



(c) The clustered two groups of deployed *MPE HpV*.



(d) The clustered two groups of traffic volume

- Log-Rank Equality Test

Hours	Observed Events	Expected Events
1	14	43.37
2	422	392.63
Total	436	436.00
		Chi2(1) = 23.98
		Pr>chi2 = 0.0000

a) Log-rank test for *MPE* hours groups.

Visits	Observed Events	Expected Events
1	16	46.85
2	420	389.15
Total	436	436.00
		Chi2(1) = 24.63
		Pr>chi2 = 0.0000

b) Log-rank test for *MPE* visits groups.

HpV	Observed Events	Expected Events
1	164	92.37
2	272	343.63
Total	436	436.00
		Chi2(1) = 74.16
		Pr>chi2 = 0.0000

c) Log-rank test for *MPE* hours/visit groups.

Traffic Count	Observed Events	Expected Events
1	90	57.40
2	346	378.60
Total	436	436.00
		Chi2(1) = 21.85
		Pr>chi2 = 0.0000

d) Log-rank test for traffic volume groups.

➢ **<u>Collectors Sites Only in 2019</u>**

- Cox Proportional Hazard Model

	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Hours	0.9851	0.00310	0.000	0.9790	0.99119

a) The total number of deployed *MPE* hours.

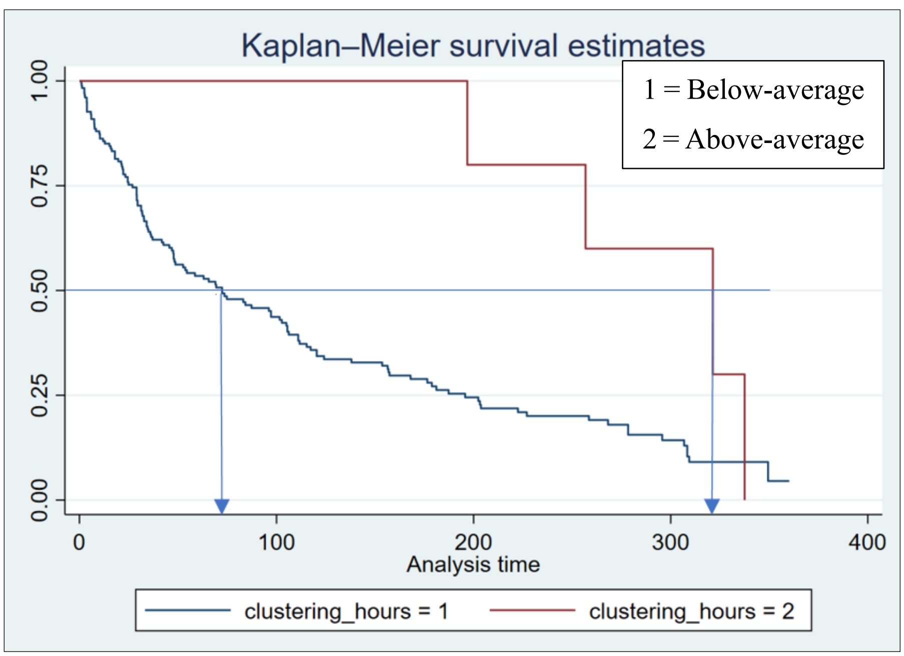
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Visits	0.9586	0.00809	0.000	0.9429	0.9746

b) The total number of deployed *MPE* visits.

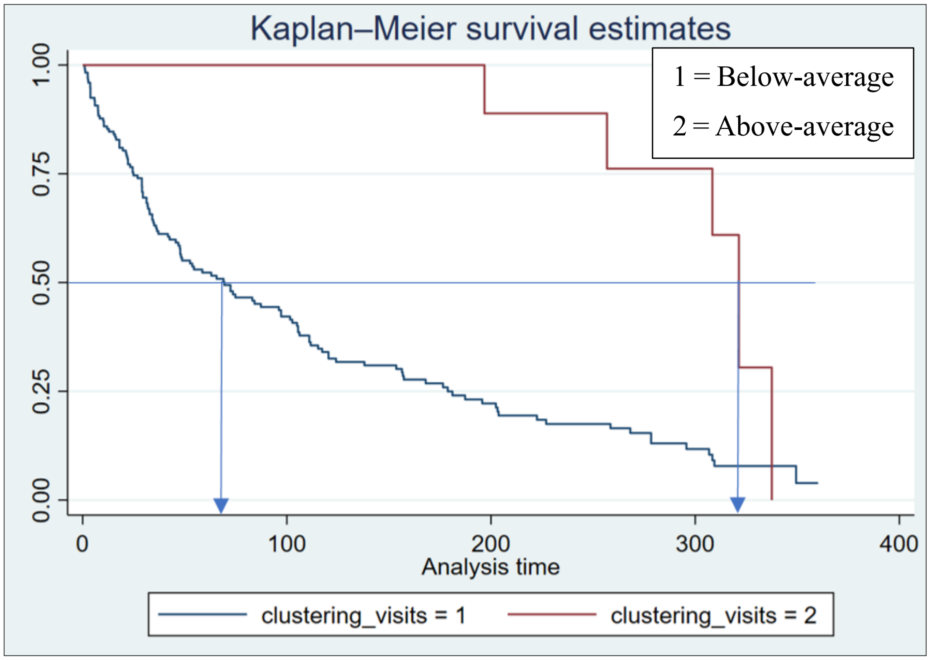
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Hours per Visits	0.65281	0.04603	0.000	0.5685	0.7495

c) The ratio between the number of *MPE* visits and hours (*HpV*).

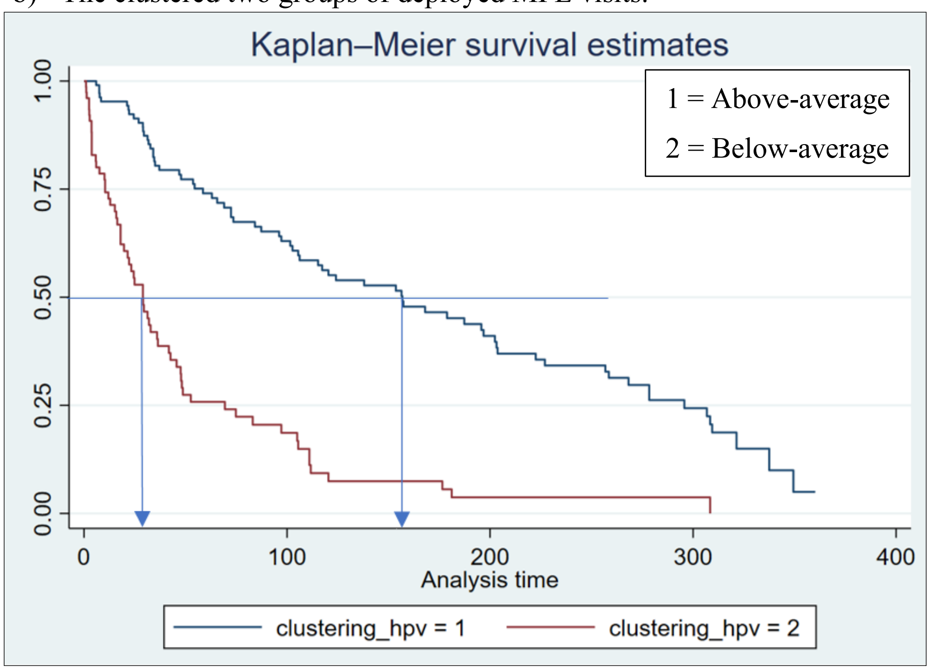
- KM Graphs



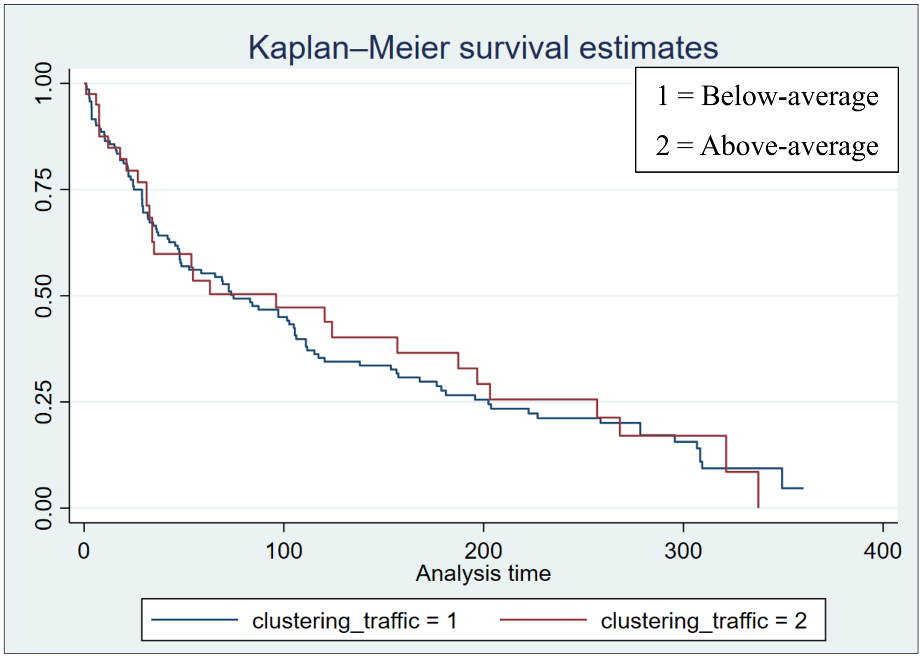
a) The clustered two groups of deployed *MPE* hours.



b) The clustered two groups of deployed MPE visits.



c) The clustered two groups of deployed *MPE* hours per visit.



d) The clustered two groups of traffic volume

- Log-rank Equality Test

Hours	Observed Events	Expected Events
1	130	124.06
2	4	9.94
Total	134	134.00
		Chi2(1) = 4.08
		Pr>chi2 = 0.0433

a) Log-rank test for *MPE* hours groups.

Visits	Observed Events	Expected Events
1	129	116.85
2	5	17.15
Total	134	134.00
		Chi2(1) = 10.68
		Pr>chi2 = 0.0011

b) Log-rank test for *MPE* visits groups.

HpV	Observed Events	Expected Events
1	71	105.70
2	63	28.30
Total	134	134.00
		Chi2(1) = 58.78
		Pr>chi2 = 0.0000

c) Log-rank test for *MPE* hours/visit groups.

Traffic Count	Observed Events	Expected Events
1	105	103.29
2	29	30.71
Total	134	134.00
		Chi2(1) = 0.13
		Pr>chi2 = 0.7233

d) Log-rank test for traffic volume groups.

➢ **High Traffic Volume Locations Only in 2019**

- Cox Proportional Hazard Model

	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Hours	0.97506	0.00286	0.000	0.96945	0.9806

(a) The total number of deployed *MPE* hours.

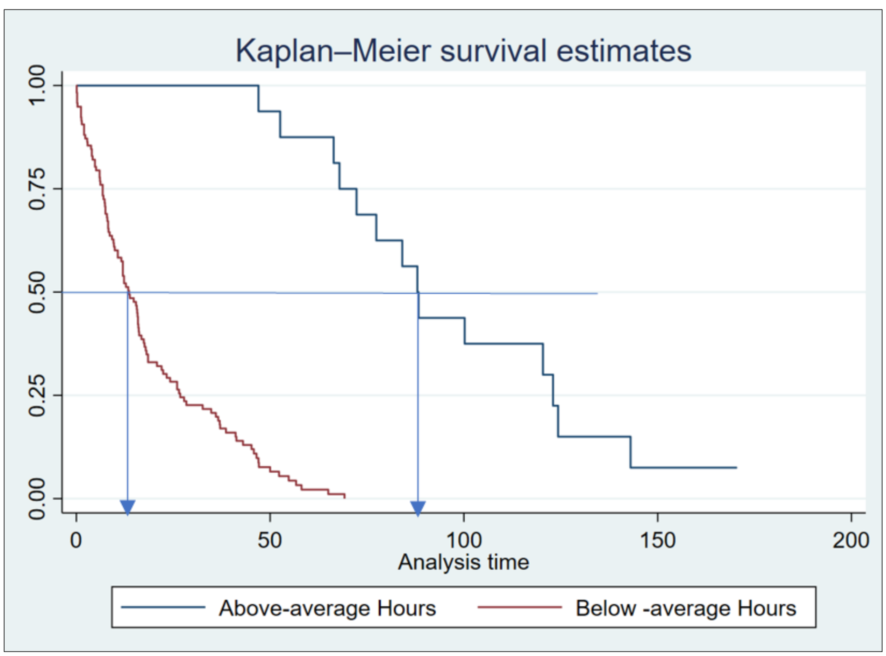
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Visits	0.9168	0.00908	0.000	0.8991	0.9347

(b) The total number of deployed *MPE* visits.

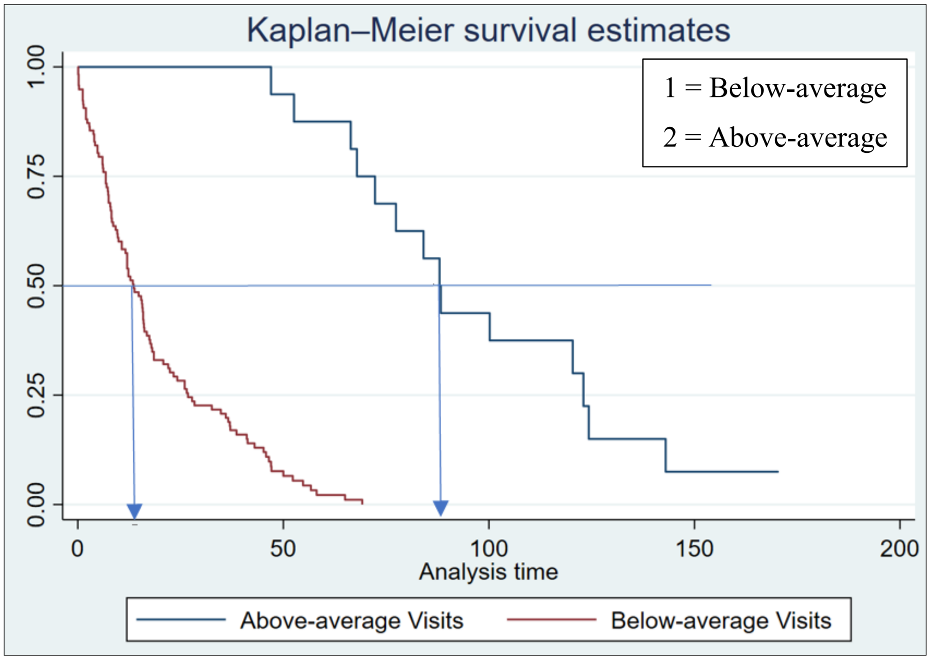
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Hours per Visits	0.4884	0.0545	0.000	0.3924	0.6078

(c) The ratio between the number of visits and hours (*HpV*).

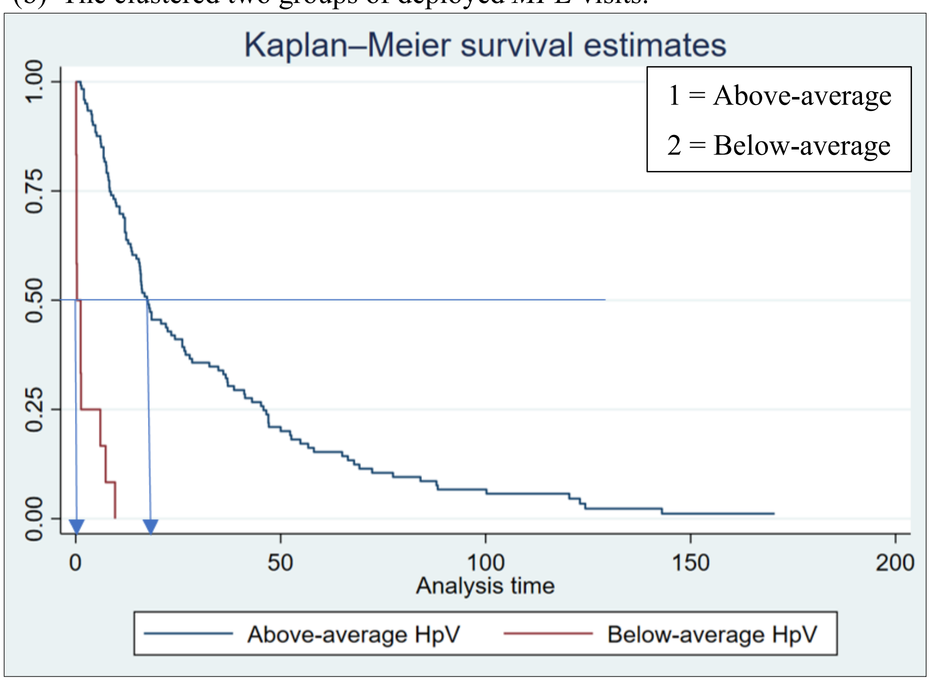
- KM Graphs



(a) The clustered two groups of deployed *MPE* hours.



(b) The clustered two groups of deployed *MPE* visits.



(c) The clustered two groups of deployed *MPE* visits.

- Log-rank Equality Test

Hours	Observed Events	Expected Events
Above average Below-average	14 109	44.25 78.75
Total	123	123.00
		Chi2(1) = 48.42 Pr>chi2 = 0.0000

a) Log-rank test for *MPE* hours groups.

Visits	Observed Events	Expected Events
Above average Below-average	14 109	44.25 78.75
Total	123	123.00
		Chi2(1) = 48.42 Pr>chi2 = 0.0000

b) Log-rank test for *MPE* visits groups.

HpV	Observed Events	Expected Events
Above average Below-average	111 12	121.69 1.31
Total	123	123.00
		Chi2(1) = 92.06 Pr>chi2 = 0.0000

c) Log-rank test for *MPE* hours/visit groups.

➢ **Low Traffic Volume Locations Only in 2019**

- Cox Proportional Hazard Model

	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Hours	0.9918	0.00124	0.000	0.9894	0.9943

(a) The total number of deployed *MPE* hours.

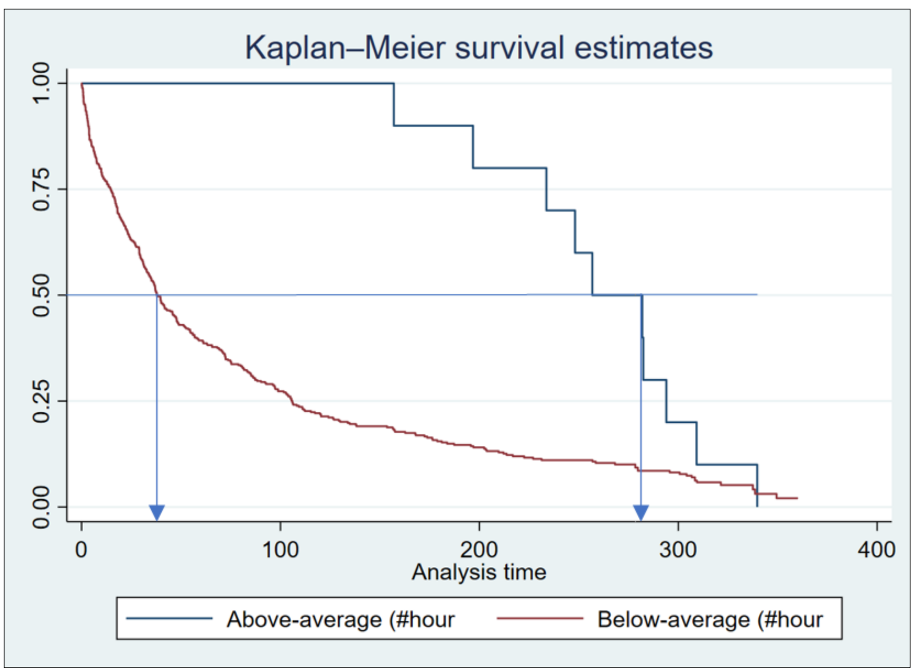
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Visits	0.9726	0.0037	0.000	0.9652	0.9800

(b) The total number of deployed *MPE* visits.

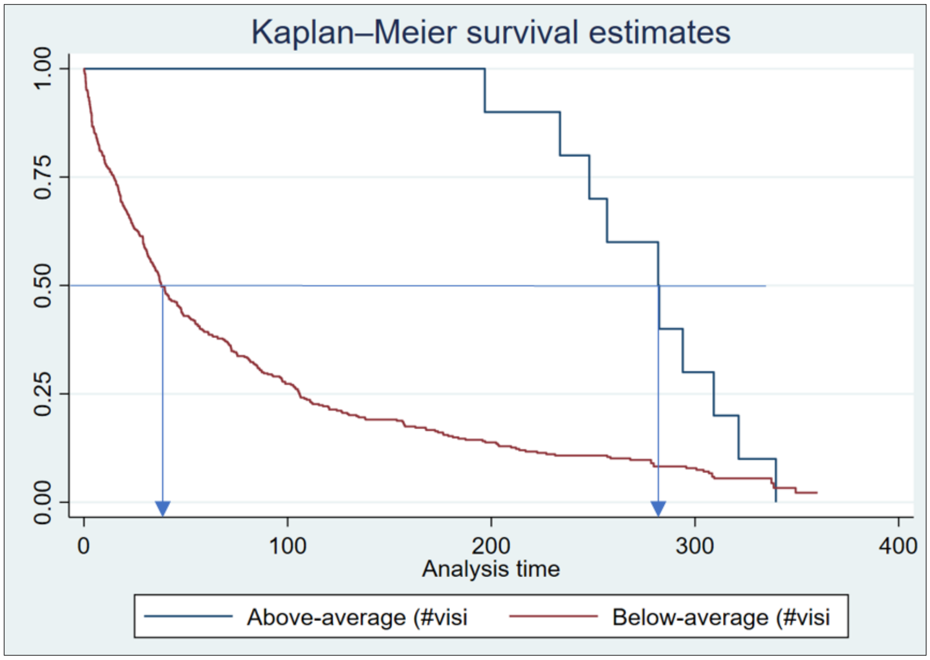
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Hours per Visits	0.7072	0.0244	0.000	0.6608	0.7569

(c) The ratio between the number of *MPE* visits and hours (*HpV*).

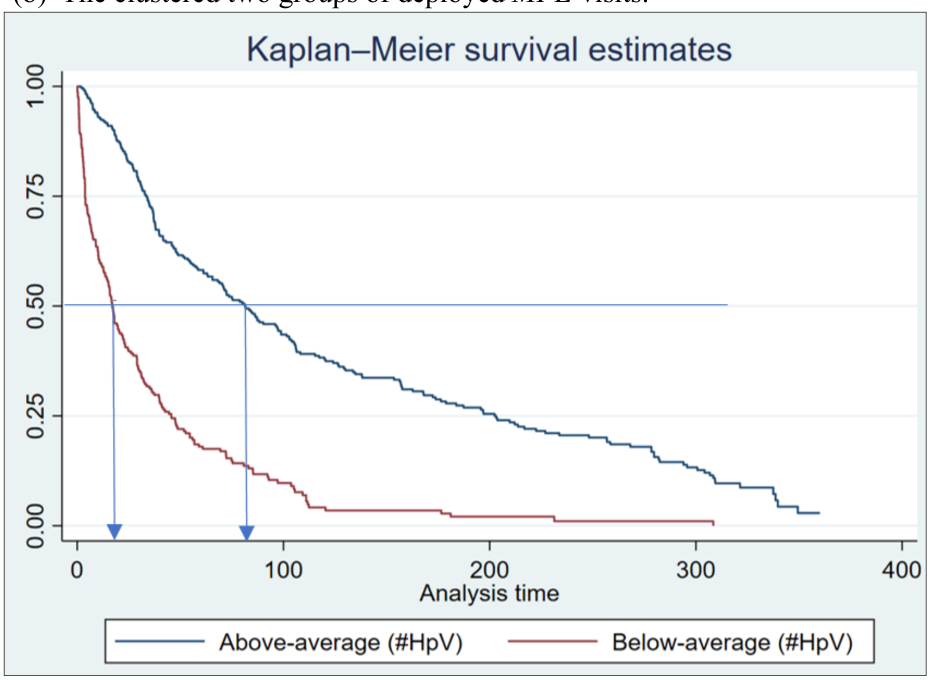
- KM Graph



(a) The clustered two groups of deployed *MPE* hours.



(b) The clustered two groups of deployed *MPE* visits.



(c) The clustered two groups of deployed *MPE* hours/visit.

- Log-rank Equality Test

Hours	Observed Events	Expected Events
Above average	10	24.78
Below-average	440	425.22
Total	450	450.00
		Chi2(1) = 9.67
		Pr>chi2 = 0.0019

a) Log-rank test for *MPE* hours groups.

Visits	Observed Events	Expected Events
Above average	10	26.11
Below-average	440	423.89
Total	450	450.00
		Chi2(1) = 11.03
		Pr>chi2 = 0.0009

b) Log-rank test for *MPE* visits groups.

HpV	Observed Events	Expected Events
Above average	234	340.67
Below-average	216	109.33
Total	450	450.00
		Chi2(1) = 149.95
		Pr>chi2 = 0.0000

c) Log-rank test for *MPE* hours/visit groups.

➢ **Speeding Related Collisions Only in 2019**

- Cox Proportional Hazard Model

	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Hours	0.9929	0.0015	0.000	0.9900	0.9959

(a) The total number of deployed *MPE* hours.

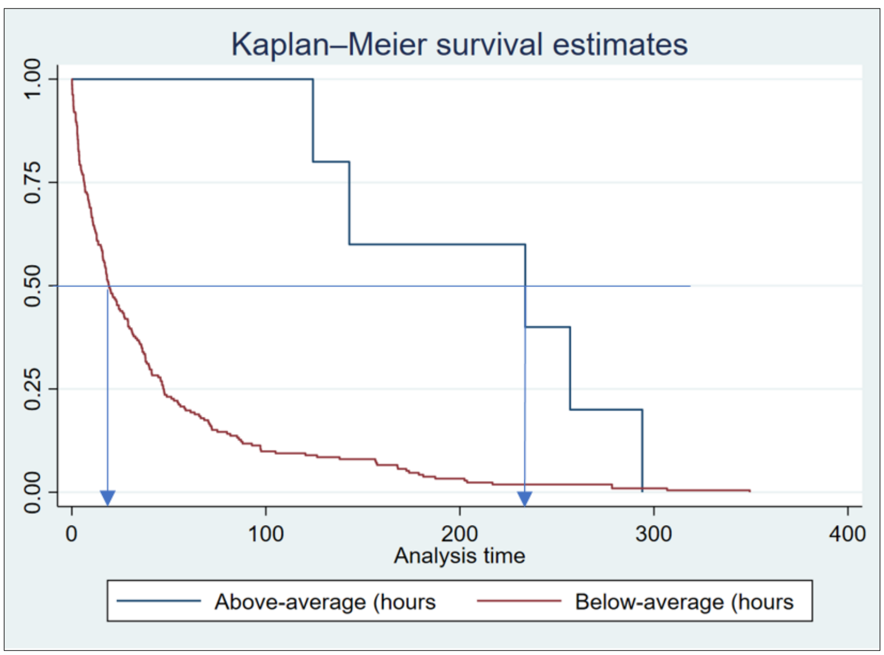
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Deployment Visits	0.9742	0.0050	0.000	0.9644	0.9841

(b) The total number of deployed *MPE* visits.

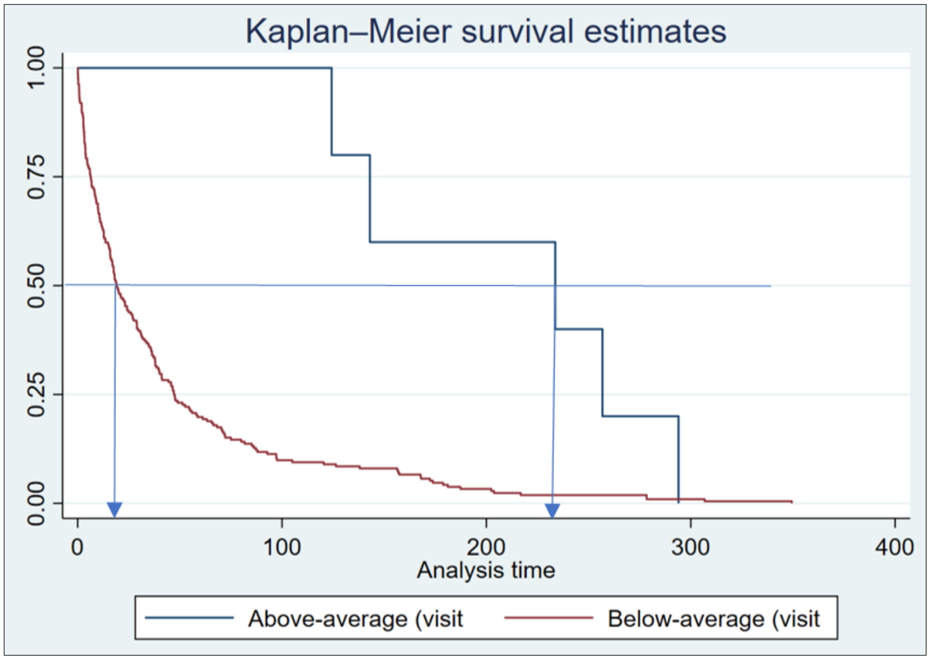
	Haz. Ratio	Std. Err.	P>|z|	[95% Conf. Interval]	
Hours per Visits	0.6789	0.0352	0.000	0.6131	0.7517

(c) The ratio between the number of visits and hours (*HpV*).

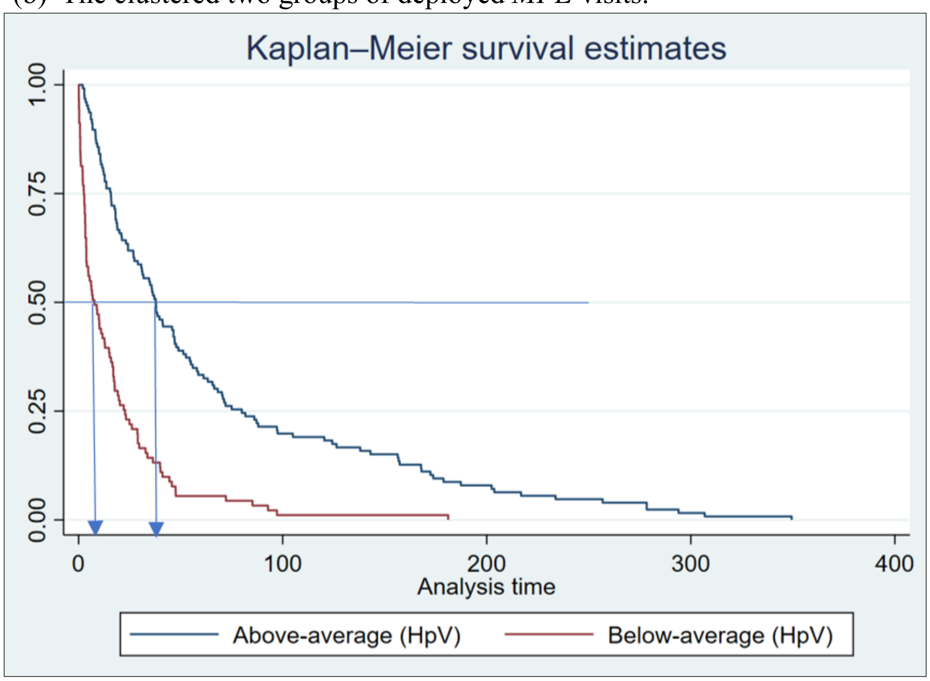
- KM Graph



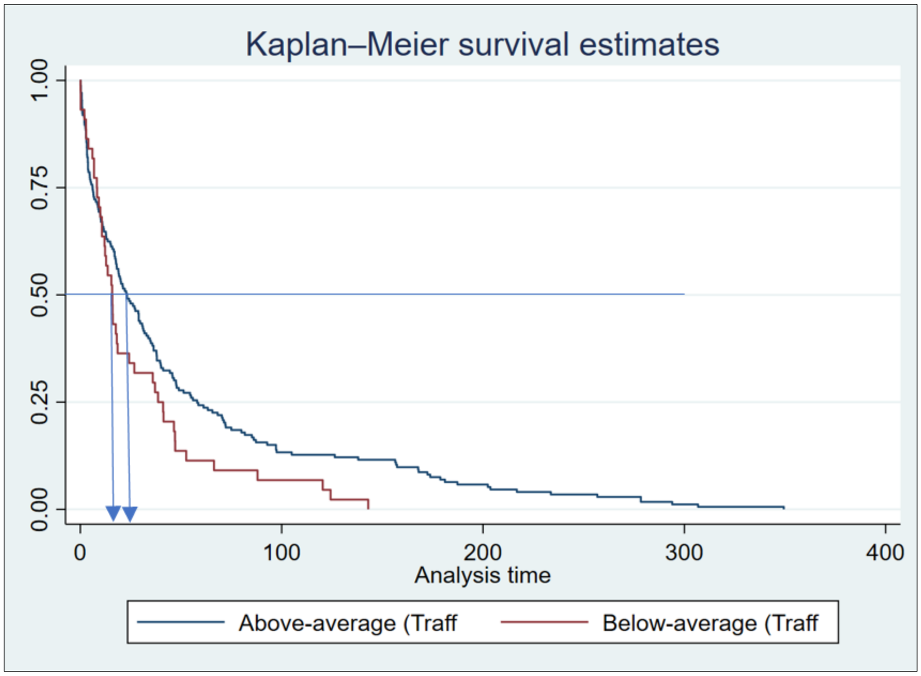
(a) The clustered two groups of deployed *MPE* hours.



(b) The clustered two groups of deployed *MPE* visits.



(c) The clustered two groups of deployed *MPE* hours/visit.



(d) The clustered two groups of traffic volume

- Log-rank Equality Test

Hours	Observed Events	Expected Events
Above average	5	16.16
Below-average	212	200.84
Total	217	217.00
		Chi2(1) = 9.07 Pr>chi2 = 0.0026

a) Log-rank test for *MPE* hours groups.

Visits	Observed Events	Expected Events
Above average	5	16.16
Below-average	12	200.84
Total	217	217.00
		Chi2(1) = 9.07 Pr>chi2 = 0.0026

b) Log-rank test for *MPE* visits groups.

HpV	Observed Events	Expected Events
Above average	126	170.44
Below-average	91	46.56
Total	217	217.00
		Chi2(1) = 58.60
		Pr>chi2 = 0.0000

c) Log-rank test for *MPE* hours/visit groups.

Traffic Count	Observed Events	Expected Events
Above average	173	184.06
Below-average	44	32.94
Total	217	217.00
		Chi2(1) = 4.53
		Pr>chi2 = 0.0333

d) Log-rank test for traffic volume groups.

The used MATLAB Code in Phase One (Data Preparation):

```
% -- Average Collision per day -- %
%prints out the average collision per day for each year for a given location
%{
avgCollision(colData)
input:
colData - Data set filled with the collision reports (eg. "Collsion
Data.xlsx")
Return:
a table containing the average duration between collision in each year
%}

function avgCollisionV5(colData)
    %Read tables from excel sheets
    disp("Start reading data");
    format long;
    [num,txt,allData] = xlsread(colData,1,'A1:BL587577');
    tempallData = allData;
    clear txt;
    clear num;
    allData = cell2table(allData(2:end,:));
    allData.Properties.VariableNames = tempallData(1,:);
    locationsID = readtable(colData,'ReadVariableName',true,'Sheet',2);
    mpeData = readtable(colData,'ReadVariableName',true,'Sheet',3);
    trafficData = readtable(colData,'ReadVariableName',true,'Sheet',4);
    siteIds = unique(locationsID.SiteId)'; % get an array of all the unique
siteIDs

    % Generates a table filled with 0 for the average Collision Duration
    avgCollDur = array2table(zeros(length(siteIds),10));
    avgCollDur.Properties.VariableNames =
{'SiteID','2013','2014','2015','2016','2017','2018','2019',...
        '2013-2015','2016-2019'};

    % Use this if the output needs strings like "N/A" being display on excel
sheet
    % these are the field names
    strOutput = ["SiteID" "2013" "2014" "2015" "2016" "2017" "2018" "2019"
"2013-2015" "2016-2019"];

    % Generates tables filled with
    collDur09 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur09.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
```

```
    collDur10 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur10.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
    collDur11 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur11.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
    collDur12 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur12.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
    collDur13 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur13.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
    collDur14 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur14.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
    collDur15 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur15.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
    collDur16 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur16.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
```

```
    collDur17 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur17.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
    collDur18 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur18.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
    collDur19 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur19.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
    collDur20 =
table(zeros(0,1),zeros(0,1),strings(0,1),strings(0,1),zeros(0,1),zeros(0,1),s
trings(0,1),strings(0,1),strings(0,1));
    collDur20.Properties.VariableNames = {'SiteID','Days b/w
Collision','First Collision Date','Second Collision Date',...
        '# of time visited','Deployment Hours','Season','Severity','Collision
Cause'};
    collDurID16 =
table(zeros(2017,1),zeros(2017,1),zeros(2017,1),zeros(2017,1));
    collDurID16.Properties.VariableNames = {'SiteId','# of time
visited','Deployment Hours','Total Traffic'};
    collDurID17 =
table(zeros(2017,1),zeros(2017,1),zeros(2017,1),zeros(2017,1));
    collDurID17.Properties.VariableNames = {'SiteId','# of time
visited','Deployment Hours','Total Traffic'};
    collDurID18 =
table(zeros(2017,1),zeros(2017,1),zeros(2017,1),zeros(2017,1));
    collDurID18.Properties.VariableNames = {'SiteId','# of time
visited','Deployment Hours','Total Traffic'};
    collDurID19 =
table(zeros(2017,1),zeros(2017,1),zeros(2017,1),zeros(2017,1));
    collDurID19.Properties.VariableNames = {'SiteId','# of time
visited','Deployment Hours','Total Traffic'};

    tic
    % Iterates through the MPE Comb sheet locations
    x = 0;
    disp("Start looping through each SiteId");
    tableLocationName = cell2table(allData.COLLISION_LOCATION_NAME);
    tableLocationName = string(tableLocationName{:,:});
```

```
    north = (string(allData.TRAVEL_DIRECTION) == "NORTH")';
    south = (string(allData.TRAVEL_DIRECTION) == "SOUTH")';
    west = (string(allData.TRAVEL_DIRECTION) == "WEST")';
    east = (string(allData.TRAVEL_DIRECTION) == "EAST")';
    for siteId = 1:length(siteIds)
        avgCollDur{siteId,1} = siteIds(siteId);
        disp(siteId);
        myStr = locationsID.LocationDescription(siteId);
        myStr = convertCharsToStrings(myStr);
        colIndex = zeros(1,587576);
        if contains(locationsID.LocationDescription(siteId),'-')
            commaLocation = strfind(myStr,',');
            hyphenLocation = strfind(myStr,'-');
            firstLocation = upper(extractBetween(myStr,1,commaLocation-
1,'Boundaries','inclusive'));
            secondNumber =
str2double(extractBetween(myStr,commaLocation+2,hyphenLocation-
2,'Boundaries','inclusive'));
            thirdNumber =
str2double(extractBetween(myStr,hyphenLocation+2,strfind(extractBetween(myStr
,hyphenLocation+2,hyphenLocation+6,'Boundaries','inclusive'),'
')+hyphenLocation+1,'Boundaries','inclusive'));
            roadName =
upper(extractBetween(myStr,hyphenLocation+strfind(extractBetween(myStr,hyphen
Location+2,hyphenLocation+6,'Boundaries','inclusive'),' ')+2,...
                strlength(myStr),'Boundaries','inclusive'));
            if secondNumber < thirdNumber
                smallerNumber = secondNumber;
            else
                smallerNumber = thirdNumber;
            end
            tableFirstLocation = table(firstLocation);
            tableFirstLocation.Properties.VariableNames = ["Var1"];
            for i = 1:abs(thirdNumber-secondNumber)+1
                secondLocation = upper(string(smallerNumber+i-1) + ' ' +
roadName);
                tableSecondLocation = table(secondLocation);
                tableSecondLocation.Properties.VariableNames = ["Var1"];
                tempcolIndex =
((startsWith(tableLocationName,tableFirstLocation{1,1}) |
contains(tableLocationName," "+tableFirstLocation{1,1})) &...
                    (startsWith(tableLocationName,tableSecondLocation{1,1}) |
contains(tableLocationName," "+tableSecondLocation{1,1})))';
                colIndex = colIndex | tempcolIndex;
            end
        else
            commaLocation = strfind(myStr,',');
            firstLocation = upper(extractBetween(myStr,1,commaLocation-
1,'Boundaries','inclusive'));
```

```
            secondLocation =
upper(extractBetween(myStr,commaLocation+1,strlength(myStr),'Boundaries','inc
lusive'));
            tableFirstLocation = table(firstLocation);
            tableFirstLocation.Properties.VariableNames = ["Var1"];
            tableSecondLocation = table(secondLocation);
            tableSecondLocation.Properties.VariableNames = ["Var1"];
            tempcolIndex =
((startsWith(tableLocationName,tableFirstLocation{1,1}) |
contains(tableLocationName," "+tableFirstLocation{1,1})) &...
                (startsWith(tableLocationName,tableSecondLocation{1,1}) |
contains(tableLocationName," "+tableSecondLocation{1,1})))';
            colIndex = colIndex | tempcolIndex;
        end
        switch string(locationsID.Direction(siteId))
            case "NB"
                colIndex = colIndex & north;
            case "SB"
                colIndex = colIndex & south;
            case "WB"
                colIndex = colIndex & west;
            case "EB"
                colIndex = colIndex & east;
            otherwise
                colIndex = colIndex & zeros(1,587576);
        end
        % Creating a cell array to orgainize the reports in terms of year
2013-2019
        reportArray = cell(1,12);

        for q = 1:12
            reportArray{q}{1,end+1} = datestr(datetime(q + 2008,1,1,0,0,1));
            reportArray{q}{2,end} = "Winter";
            reportArray{q}{3,end} = "N/A";
            reportArray{q}{4,end} = "N/A";
        end

        % Check which reports has the combCode
        collDurID16{siteId,1} = siteIds(siteId);
        collDurID17{siteId,1} = siteIds(siteId);
        collDurID18{siteId,1} = siteIds(siteId);
        collDurID19{siteId,1} = siteIds(siteId);
        colReports = allData(colIndex,:);
        % Going through each report to organize it by year
        for j = 1:size(colReports,1)
            report = colReports(j,:);

            %2012 use to get index because the oldest year recorded is 2013
            cellIndex = report.COLLISION_REPORT_YEAR - 2008;

            % Formating datetime
```

```
            if (report.HOUR_NAME{1,1}/100 == 24)
                reportDate =
datetime(report.COLLISION_REPORT_YEAR,report.COLLISION_MONTH,...
                    report.DAY_OF_MONTH,report.HOUR_NAME{1,1}/100-1,59,59);
            elseif (isa(report.HOUR_NAME{1,1},'char'))
                reportDate =
datetime(report.COLLISION_REPORT_YEAR,report.COLLISION_MONTH,...
                    report.DAY_OF_MONTH,0,0,1);
            else
                reportDate =
datetime(report.COLLISION_REPORT_YEAR,report.COLLISION_MONTH,...
                    report.DAY_OF_MONTH,report.HOUR_NAME{1,1}/100,59,59);
            end
            reportDate.Format = 'yyyy-MM-dd HH:mm:SS';
            reportDate = datestr(reportDate);

            % Find Season
            switch (report.COLLISION_MONTH)
                case {3,4,5}
                    reportSeason = "Spring";
                case {6,7,8}
                    reportSeason = "Summer";
                case {9,10,11}
                    reportSeason = "Fall";
                case {1,2,12}
                    reportSeason = "Winter";
            end

            % Adding the report to the correct year where cellIndex 1-7 is
2013-2016 respectively
            reportArray{cellIndex}{1,end+1} = reportDate;
            reportArray{cellIndex}{2,end} = reportSeason;
            reportArray{cellIndex}{3,end} = report.COLLISION_CLASSIFICATION;
            reportArray{cellIndex}{4,end} = report.COLLISION_CAUSE_NAME;
        end
        for q = 1:12
            reportArray{q}{1,end+1} = datestr(datetime(q +
2008,12,31,23,59,59));
            reportArray{q}{2,end} = "Winter";
            reportArray{q}{3,end} = "N/A";
            reportArray{q}{4,end} = "N/A";
        end

        avDCArray = zeros(1,12);
        % Calculating the Avg duration between collisions
        for cellIndex = 1:size(reportArray,2)
            totalHours = 0;
            numVisits = 0;
            whichArray = (mpeData.StartDate > datetime(cellIndex +
2007,12,31)) & (mpeData.EndDate < datetime(cellIndex + 2009,1,1)) &
(mpeData.SiteId == siteIds(siteId));
```

```
            whichSite = mpeData(whichArray,:);
            for k = 1:height(whichSite)
                if ~isempty(whichSite.Duration_dep{k,1})
                    [Y,M,D,H,MN,S] =
datevec(datetime(string(whichSite.Duration_dep(k)),'InputFormat','H:mm'));
                    totalHours = totalHours + H + MN/60;
                end
            end
            numVisits = height(whichSite);
            trafficArray = (trafficData.WatchDate > datetime(cellIndex +
2007,12,31)) & (trafficData.WatchDate < datetime(cellIndex + 2009,1,1)) &
(trafficData.Site == siteIds(siteId));
            trafficSite = trafficData(trafficArray,:);
            traffic = 0;
            for k = 1:height(trafficSite)
                if ~isempty(trafficSite.Traffic(k))
                    traffic = traffic + trafficSite.Traffic(k);
                end
            end
            switch cellIndex
                case 8
                    collDurID16{siteId,2} = numVisits;
                    collDurID16{siteId,3} = totalHours;
                    collDurID16{siteId,4} = traffic;
                case 9
                    collDurID17{siteId,2} = numVisits;
                    collDurID17{siteId,3} = totalHours;
                    collDurID17{siteId,4} = traffic;
                case 10
                    collDurID18{siteId,2} = numVisits;
                    collDurID18{siteId,3} = totalHours;
                    collDurID18{siteId,4} = traffic;
                case 11
                    collDurID19{siteId,2} = numVisits;
                    collDurID19{siteId,3} = totalHours;
                    collDurID19{siteId,4} = traffic;
            end
            b = size(reportArray{cellIndex});
            if b(2) > 2

                % Getting the list of dates and sorting them
                dates = reportArray{cellIndex};
                dates = dates';
                % disp(x);
                [~,idx] = sort(datenum(dates(:,1),'dd-mm-yyyy HH:MM:SS'), 1,
'ascend');
                x = x+1;
                dates = dates(idx,:,:,:)';
                numCol = size(dates);
                dates{1,1} = datetime(cellIndex + 2008,1,1);
                dates{1,end} = datetime(cellIndex + 2008,12,31);
```

```
                % If there is 1 collision assume 0 for now
                if numCol(2) == 1
                    % Assigned as "N/A" but since table is double will appear
as "NaN"
                    avDCArray(cellIndex) = "N/A";
                else
                    % Calculating the time difference in days between each
dates
                    clear onlyTimes;
                    for a = 1:numCol(2)
                        dates{1,a} = datetime(dates{1,a});
                        onlyTimes(a) = dates{1,a};
                    end
                    timediff = caldiff(onlyTimes,'time');
                    timediff = time(timediff);
                    timediff = days(timediff); % Doing this way makes it more
precise

                    % Calculating the Average Duration between Collisions
                    avDCArray(cellIndex) = round(mean(timediff),1);
                    mpeDurations = zeros(length(timediff),2);
                    for z = 1:length(timediff)
                        timeVisit = 0;
                        truthArray = (mpeData.StartDate > dates{1,z}) &
(mpeData.EndDate < dates{1,z+1}) & (mpeData.SiteId == siteIds(siteId));
                        truthMpe = mpeData(truthArray,:);
                        for y = 1:height(truthMpe)
                            if ~isempty(truthMpe.Duration_dep{y,1})
                                [Y,M,D,H,MN,S] =
datevec(datetime(string(truthMpe.Duration_dep(y)),'InputFormat','H:mm'));
                                mpeDurations(z,2) = mpeDurations(z,2) + H +
MN/60;
                            end
                        end
                        mpeDurations(z,1) = height(truthMpe);
                    end
                    % Assigning the duration between collision to the correct
year to be printed.
                    if cellIndex == 1
                        for duration = 1:length(timediff)
                            collDur09{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
                        end
                    elseif cellIndex == 2
                        for duration = 1:length(timediff)
```

```
                            collDur10{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
                        end
                    elseif cellIndex == 3
                        for duration = 1:length(timediff)
                            collDur11{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
                        end
                    elseif cellIndex == 4
                        for duration = 1:length(timediff)
                            collDur12{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
                        end
                    elseif cellIndex == 5
                        for duration = 1:length(timediff)
                            collDur13{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
                        end
                    elseif cellIndex == 6
                        for duration = 1:length(timediff)
                            collDur14{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
                        end
                    elseif cellIndex == 7
                        for duration = 1:length(timediff)
                            collDur15{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
```

```
                    end
                elseif cellIndex == 8
                    for duration = 1:length(timediff)
                        collDur16{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
                    end
                elseif cellIndex == 9
                    for duration = 1:length(timediff)
                        collDur17{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
                    end
                elseif cellIndex == 10
                    for duration = 1:length(timediff)
                        collDur18{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
                    end
                elseif cellIndex == 11
                    for duration = 1:length(timediff)
                        collDur19{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
                    end
                else
                    for duration = 1:length(timediff)
                        collDur20{end+1,:} = [ [siteIds(siteId)],
[round(timediff(duration),1)], [datestr(dates{1,duration})],
[datestr(dates{1,duration+1})],...

[mpeDurations(duration,1)],[mpeDurations(duration,2)],[dates{2,duration+1}],
[dates{3,duration+1}], [dates{4,duration+1}]];
                    end
                end
            end
        end
    end
```

```
        % Splitting avDCArray so it faster to go through and take value that
are not "NaN"
        firstHalf = avDCArray(1:3);
        secondHalf = avDCArray(4:7);
        % Getting the avg between 2013-2015 and 2016-2019
        avDCArray(end+1) = round(mean(firstHalf(~isnan(firstHalf))),1);
        avDCArray(end+1) = round(mean(secondHalf(~isnan(secondHalf))),1);

        strOutput(siteId+1,1) = siteIds(siteId);
        strOutput(siteId+1,2:15) = string(avDCArray);
        replace = find(isnan(avDCArray)); % Finding the indices with "NaN" to
replace with "N/A"
        strOutput(siteId+1, (replace+1)) = "N/A";
        %avgCollDur{siteId,2:10} = avDCArray;

    end
    toc
    %disp(avgCollDur); % has NaN
    % Saving Datasheet with averages in case the datasheet is large

    %writetable(avgCollDur,'avgCollDur.xlsx'); % Comment line if table can't
use NaN in table
    writematrix(strOutput,'avgCollDurV3.xlsx'); % Comment line if table can't
use strings like "N/A" in table

    % Saving duration between collision of each year at each location
    writetable(collDur09,'Duration btwn Collision2009.xlsx');
    writetable(collDur10,'Duration btwn Collision2010.xlsx');
    writetable(collDur11,'Duration btwn Collision2011.xlsx');
    writetable(collDur12,'Duration btwn Collision2012.xlsx');
    writetable(collDur13,'Duration btwn Collision2013.xlsx');
    writetable(collDur14,'Duration btwn Collision2014.xlsx');
    writetable(collDur15,'Duration btwn Collision2015.xlsx');
    writetable(collDur16,'Duration btwn Collision2016.xlsx');
    writetable(collDur17,'Duration btwn Collision2017.xlsx');
    writetable(collDur18,'Duration btwn Collision2018.xlsx');
    writetable(collDur19,'Duration btwn Collision2019.xlsx');
    writetable(collDur20,'Duration btwn Collision2020.xlsx');
    writetable(collDurID16,'siteID16.xlsx');
    writetable(collDurID17,'siteID17.xlsx');
    writetable(collDurID18,'siteID18.xlsx');
    writetable(collDurID19,'siteID19.xlsx');
end
```