THE UNIVERSITY OF MEMPHIS. Dreamers. Thinkers. Doers.

The Intermodal Freight Transportation Institute (IFTI)
Herff College of Engineering

Civil Engineering
Herff College of Engineering

EFFECT OF PRIMARY AND SECONDARY CRASHES: IDENTIFICATION, VISUALIZATION AND PREDICTION

DRAFT FINAL REPORT

Prepared By:

Sabya Mishra, Ph.D., P.E.
Assistant Professor, Dept. of Civil Engineering, University of Memphis, 112D Engr. Sc. Bldg., 3815 Central Avenue, Memphis, TN 38152, Tel: 901-678-5043, Fax: 901-678-3026, Email: smishra3@memphis.edu

Mihalis M. Golias, Ph.D. A.M. ASCE
Associate Professor, Dept. of Civil Engineering, Associate Director for Research Intermodal Freight Transportation Institute, University of Memphis, 104C Engr. Sc. Bldg., 3815 Central Avenue, Memphis, TN 38152, Tel: 901-678-3048, Fax: 901-678-3026, Email: mgkolias@memphis.edu

Afrid A. Sarker
Research Assistant, Dept. of Civil Engineering, University of Memphis and Research Assistant, Intermodal Freight Transportation Institute, 302A Engineering Admin. Bldg., 3815 Central Ave Memphis, TN 38152, USA, Phone: 260-797-1886, email: aasarker@memphis.edu

Alireza Naimi
Research Assistant, Dept. of Civil Engineering, University of Memphis and Research Assistant, Intermodal Freight Transportation Institute, 302A Engineering Admin. Bldg., 3815 Central Ave Memphis, TN 38152, USA, Phone: 201-855-6606, email: anaimi@memphis.edu

DISCLAIMER

This research was funded by the Wisconsin Department of Transportation (WisDOT) and the United States Department of Transportation (USDOT) Pooled Fund Program: TPF-5(274) in the interest of information exchange. Details on the Pooled Fund Program for this project can be found http://www.pooledfund.org/Details/Study/507. The material or information presented/published/reported is the result of research done under the auspices of the Department. The content of this presentation / publication / report reflects the views of the authors, who are responsible for the correct use of brand names, and for the accuracy, analysis and any inferences drawn from the information or material presented/published/reported. WisDOT and FHWA (USDOT) assume no liability for its contents or use thereof. This presentation/publication/report does not endorse or approve any commercial product, even though trade names may be cited, does not reflect official views or policies of the Department or FHWA (USDOT), and does not constitute a standard specification or regulation of the Department or FHWA.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Urban areas are experiencing increasing traffic congestion and significant portion is no-recurring in nature. Traffic incidents often result in occurrence of a crash, termed as "Primary Crash (PC)". PC causes reduction in roadway capacity, which in turn results in another crash, referred as a "Secondary Crash (SC)". Though a relatively small proportion of all the crashes are secondary, it is important to identify contributing factors as well as their characteristics because SCs increase congestion, delays, fuel consumption and emissions. According to past studies, up to 15% of reported crashes have occurred partly or entirely as the result of a PC. United States Department of Transportation (USDOT) estimates that 18% of freeway traffic related fatalities are attributed to SCs and can contribute up to 50% of congestion in urban areas. A number of states have proposed various programs to reduce SCs and estimate their benefit from wider economic impact of enhanced traffic operation and safety. Identifying the SCs and their primary contributing factors can help the traffic incident management (TIM) agencies in congestion management and safety improvement by preventing SC occurrence.

There are two major objectives of this study. First, to develop a procedure to identify SCs in a relatively large transportation network with multiple roadway facility types using a static and a dynamic approach. Second, to develop prediction models to determine primary contributing factors and PC characteristics that may induce a SC. These crash prediction models would allow the TIM agencies to respond to an incident with necessary actions to reduce the likelihood of secondary crash occurrence.

Two types of models were developed for identification of SCs: (1) static approach, and (2) dynamic approach. In static approach, pre-specified spatial and temporal thresholds were applied. In contrast, dynamic approach considered no such assumptions but rather SC identification was obtained dynamically with given traffic flow conditions. For prediction of SC occurrence, various discrete choice models were developed. Both identification and prediction models were validated for their robustness. In this research, Shelby County, TN is considered as the study area. Located in the western part of TN, Shelby County is the most populous and heavily travelled county in the state. Safety, traffic exposure, highway geometry, environmental and other data were obtained from Tennessee Roadway Information Management System (TRIMS) database for the years 2010 to 2012 and used for identification and prediction of SCs.

The study found that SCs account for approximately 1.5% of all the recorded crashes using 2010-2012 data and they are more prevalent on major arterials compared to freeways. The results from crash prediction models revealed that some of the factors that may induce secondary crashes are: increased number of vehicle involved in a primary crash, crash occurring on a roadway with relatively high Annual Average Daily Traffic

(AADT), bad weather condition (e.g. rain fog, snow, sleet etc.) and the type of primary incident type such as “rear-end”.

Most of the previous research was conducted on short segments of freeways in a small regional scale because detailed network and traffic data on arterials are often not available to capture the dynamic variation in traffic flow characteristics caused by a PC. This study has several major contributions: a) development of a procedure to identify secondary crashes in large networks using minimal available data for planning agencies, b) development of a dynamic queue length based approach to identify secondary crashes in a large scale multi-facility highway networks (freeway and arterials) using crash, traffic, incident management, and roadway network data), (c) development of econometric models to determine the factors that are more likely to induce secondary crash on multi-facility network. The proposed methodologies and developed models can be used by state and local transportation agencies for development of strategies and planning preparedness to enhance operation and safety.

CHAPTER 1: INTRODUCTION

Traffic crashes are a major source of congestion on freeway and arterial systems. A "*primary crash (PC)*" leads to reduction of roadway capacity and may result in what is known as a "*secondary crash (SC)*". In this report, the terms '*crashes'* and '*incidents'* are used interchangeably. SCs are defined as crashes that occur in close proximity of the primary incident's location as a result of either queuing (in the same direction) or driver distraction (in the opposite direction) (Margiotta et al., 2012). Earlier studies suggest that up to 15% of reported crashes have occurred partly or entirely as the result of a PC (Raub, 1997a). Though a relatively small percentage of all crashes are secondary, it is important to identify contributing factors and characteristics, and mitigate their effects on congestion, delay, fuel consumption and emission. SCs are non-recurring in nature and contribute up to 50% of congestion in urban areas (Kwon et al., 2006; Ozbay and Kachroo, 1999; Skabardonis et al., 1998). Reducing the occurrence of SCs is a major concern for traffic incident management (TIM) agencies, especially when dispatching rescue vehicles to clear the affected traffic lanes[1] (Dunn and Latoski, 2003; Owens et al., 2010). United States Department of Transportation (USDOT) estimates that 18% of freeway traffic related fatalities are attributed to SCs (Chimba et al., 2014). Limiting the impact of nonrecurring events, such as SCs and disabled vehicles, through effective incident management is one of the objectives of emergency response professionals (Raub and Schofer, 1997). Understanding the characteristics of primary and secondary crashes can help decision-makers select better traffic operation practices and safety programs. The first step towards achieving these goals is to identify SCs and their contributing factors such as crash severity, clearance time, and facility type. It is extremely important that SCs are identified with great accuracy otherwise any steps taken towards mitigation might prove inefficient.

Past research on SCs considered short segments of freeways in small regional scales for easier delineation of direction, and spatiotemporal thresholds. The most challenging task was identification of SCs in terms of these thresholds, and directional criteria (Zheng et al., 2014). The latter, often a complex process, is the task of attaching the precise location of a crash to a specific lane. Precise lane and direction identification may be relatively easier for freeways, but poses a challenge for undivided medians. Therefore, arterials were excluded in most of the published research to date even though they encounter a significant number of SCs and their identification warrants further research.

The first objective of this research is to develop a procedure to identify SCs in a relatively large transportation network with multiple roadway facility types using a static and a dynamic approach. The former approach assumes pre-specified temporal and spatial

[1] Recently, one of the performance measures used by TIM agencies is reduction of SCs.

thresholds, based on past experience or engineering judgment, while the latter determines these thresholds based on real-time traffic conditions. The contributions of this study in the area of identification of SCs are: a) development of a procedure to identify SCs in large-scale networks without using high resolution data and within acceptable computational times, and b) development of a dynamic queue length based approach to identify SCs in a multi-facility highway networks using crash, traffic, incident management, and roadway network data. Once the SCs are identified, analysis of their characteristics is imperative because to prevent SC occurrence, primary contributing factors need to be determined. Hence, the second objective of this study is to develop discrete choice models to identify the factors that are most likely to induce SCs. The factors considered include roadway and traffic characteristics, primary crash features, and time of day (TOD) factors.

A stand-alone tool is also developed incorporating the methodology. This tool identifies primary and secondary crashes using an archived incident database based on spatio-temporal criteria, facility types and traffic flow characteristics. The tool has options to select various features by the user such as time period of analysis, type of identification (static vs. dynamic), facility type, time-of-day etc. The tool will be useful to state and local planning agencies for identification of SC occurrence and their corresponding pattern.

The rest of the report is organized as follows. The next chapter discusses practices and published research on identification and prediction of SCs. The third chapter presents the proposed methodology for identification of SC, followed by a case study. This chapter also compares SC identification accuracy and consistency of both approaches along with validation. The fourth chapter introduces the prediction models and provides a discussion on the result. Description on a standalone tool is presented in the fifth chapter. The final chapter concludes the report summarizing findings and presenting future research directions.

CHAPTER 2: LITERATURE REVIEW

This chapter discusses a comprehensive literature review on (1) SC identification from the relevant literatures along with different criteria for spatiotemporal thresholds, (2) Recent techniques used for SC identification, and (3) crash prediction models for analyzing SC. At the end of this chapter strengths and weaknesses of past studies are discussed.

2.1 Temporal and spatial threshold

The first step in defining a SC is selection of temporal and spatial thresholds (relative to a PC). Two types of thresholds have been prominent in the literature: static (predefined) and dynamic (varies based on incident characteristics and queuing of vehicles). Several studies (Chang and Rochon, 2003; Hagen, 2005; Hirunyanitiwattana and Mattingly, 2006; Karlaftis et al., 1999; Moore et al., 2004; Pigman et al., 2011; Raub, 1997b; Zhan et al., 2009, 2008) illustrate the use of static thresholds in SCs classification (reaching up to 2 miles and 2 hours after the occurrence of a PC) with some studies only considering crashes in the same direction as the primary incident (Hirunyanitiwattana and Mattingly, 2006; Karlaftis et al., 1999).

The dynamic approach, on the other hand, has been used to identify SCs based on the influence area of the primary incident that depends on vehicle queue length, and other incident and traffic data (Khattak et al., 2011, 2010; Zhang and Khattak, 2010). An Incident Progression Curve (IPC) was proposed in 2007 and 2010 by Sun and Chilukuri (Sun and Chilukuri, 2010, 2007), to identify the dynamic impact area of a PC. Dynamic thresholds were modeled as a multivariate function of various parameters (e.g. primary incident duration, number of blocked lanes etc.). The use of IPC reduced SC misclassification (false positive and negative) significantly. Another study developed queuing models to determine the impact area of a primary incident using estimated queue length and incident duration (Zhang and Khattak, 2011).

The likelihood of SC occurrence is commonly associated with primary incident duration. Modeling incident duration is crucial in the process of developing prediction models for SC occurrence. One of the effective techniques used in the past to estimate incident durations has been hazard-based models (Chung, 2010; Jones et al., 1991) and recently Chung (2010) utilized accelerate failure time metric model to account for the influence of the explanatory variables. One particular advantage of hazard-based duration modeling is that it allows the explicit study of the relationship between incident duration and the explanatory variables. Most studies developed a correlation between incident duration and SC likelihood, considering the influence area to be independent of prevailing traffic conditions and incident characteristics. However, recently published research (Imprialou et al., 2014; Vlahogianni et al., 2010) identified real time traffic conditions as critical component in accurate estimation of the influence areas.

2.2 Recent SCs identification techniques

Yang et al. (2014) identified SCs using speed contour plots with approximately 75% and 50% of SCs occurring within two hours after and two miles upstream of the PC respectively (Yang et al., 2014b). Overall, 42% of SCs were found to occur within two hours of the onset of a PC and within a distance of two miles upstream. 58% of SCs occurred beyond these frequently used spatiotemporal thresholds. In addition, more than half of SCs occurred from PC-induced queues lasting more than two hours. Results also revealed that rear-end crashes were the dominant SC type and that the major contributing factor was "*following too closely*". Other significant contributing factors included improper lane change, distracted driving and unsafe speeds (Yang et al., 2014a). Speed contour plot analysis limits the scope of SC identification to urban freeways as real time network speeds are needed. Obtaining such data is challenging for arterials and, even more so, for suburban freeways.

Hirunyanitiwattana and Mattingly (2006) compared differences in the characteristics of secondary and primary crashes with respect to time-of-day, roadway classification, primary collision factors, severity level and type of crash. The study revealed a higher SC rate (expectation) in regions with high traffic volumes during morning and evening peak hours. The study concluded that a PC occurring in an urban area on a high speed facility is likely to have a high probability of inducing SCs. Sensitivity analysis measuring the impact of queue length and clearance time on the estimated number of SCs revealed that reduction in queue clearance time from 60 to 15 minutes reduced the number of SCs by approximately 43%.

The literature review reveals that in the very early stages, when the concept of "secondary crash" was introduced, studies proposed spatiotemporal thresholds, independent of the facility type, crash severity, clearance time, and flow characteristics; all of which are crucial determinants of SCs. While implementing static thresholds is relatively simpler and not computation-intensive, it comes with the risk of identifying SCs with significantly high numbers of false positive and negative (type I and II errors respectively). The proposed dynamic approach to identify SCs potentially eliminates such assumptions and errors.

2.3 Crash prediction models

Several past studies have focused on identifying contributing factors for SCs. One of the literature found that the peak hour during weekdays along with clearance time are associated with secondary incidents occurrence (Raub, 1997b). In the study by Karlaftis et al. (1999), the author developed a logit model to identify the relation between clearance time of primary incident and SC occurrence in which season, day of week, vehicle type (car, tractor-trailer) and vehicle location are found to be the most significant factors for higher secondary incident likelihood. Zhan et al. (2009) developed a binary logit model to estimate the likelihood of SC occurrence. It was observed that longer the freeway lane blockage duration, higher the likelihood of SCs because of increased congestion and queue length. The authors also concluded that SCs are more likely to occur during weekday morning, afternoon peaks and mid-day hours. Another past study found similar results concluding that incident type, lane blockage duration, number of lanes, time of day and number of vehicles involved are some of the key factors associated with SC occurrence (Zhan et al., 2008).

Khattak et al. (2011) developed several model for SC occurrence using logistic regression. All the models were probit models with certain variations in each of them. Because of the presence of endogeneity, the authors used two stage least square (2SLS) method where SC occurrence is estimated using duration as endogenous variable. The study also found that if the primary incident is a crash involving multiple vehicles, if it is occurring during peak hours on a roadway with high AADT, this primary incident is highly likely to induce a SC. In another study the authors focused primarily on the interdependence between SC occurrence and duration of primary incident and concluded that they are interdependent (Khattak et al., 2009). It means that secondary incidents are more likely to occur if the primary incident lasts long and simultaneously durations of primary incidents are expected to be longer if secondary incidents take place. A detailed list of contributing factors for SC occurrence is shown in Appendix A.

2.4 Discussions

The review of past literatures revealed that most of the past studies were conducted on short segments of freeways in a small regional scale. One of the major challenges encountered in the process of identification of SCs was availability of detailed dataset which led to such scope constraints. The studies were conducted on segments, using dynamic approach, only where high resolution traffic data were available. Arterials experience significant number of SCs and hence it is imperative that they are included in the scope. But, rarely arterials have such detailed data to capture the dynamic variation in traffic flow characteristics as a result of a primary incident. So, there has been extremely limited application of dynamic approaches to identify SCs on arterials. In this research, a methodology is proposed to identify SCs on freeways and arterials in a large-scale network.

Also, very few literatures considered SCs in the opposite direction. When drivers in the opposite direction slow down to observe the PC (known as “rubbernecking effect”), it causes congestion, reduction in capacity, and associated delays and hence has a significant potential of inducing SCs in the opposite direction of a PC. In this study, multiple scenarios are considered (including opposite direction) when identifying SCs.

CHAPTER 3: IDENTIFICATION OF SEDCONDARY CRASHES

In this chapter a methodology for identification of SC is presented. A pictorial representation of the proposed methodology and a step-by-step workflow is shown in Fig. 3-1 and then described in the following subsections. Before proceeding with the methodology, we present the notations used throughout the report.

Notation	Description
$a_{bf,s}$	Backward-forming shockwave speed in the same direction
$a_{fr,s}$	Forward-recovery shockwave speed in the same direction
$a_{bf,o}$	Backward-forming shockwave speed in the opposite direction
$a_{fr,o}$	Forward-recovery shockwave speed in the opposite direction
BLM	Beginning log mile
D	Impact area
dS	Distance between two paired crashes
dT	Time interval between two paired crashes
I	Set of all the crashes
i	A primary crash
j	A potential secondary crash
$(k_{ini})_s$, $(q_{ini})_s$ $(u_{ini})_s$	Density, flow, and speed of lane in the same direction prior to primary crash
$(k_{jam})_s$,$(q_{jam})_s$,$(u_{jam})_s$	Density, flow, and speed of lane in the same direction after primary crash but prior to clearance (jam condition)
$(k_{sat})_s$,$(q_{sat})_s$, $(u_{sat})_s$	Density, flow, and speed of lane in the same direction representing optimal (saturated) condition
$(k_{ini})_o$,$(q_{ini})_o$,$(u_{ini})_o$	Density, flow, and speed of lane in the opposite direction prior to primary crash
$(k_{jam})_o$,$(q_{jam})_o$,$(u_{jam})_o$	Density, flow, and speed of lane in the opposite direction after primary crash but prior to clearance (jam condition)
$(k_{sat})_o$,$(q_{sat})_o$, $(u_{sat})_o$	Density, flow, and speed of lane in the opposite direction representing optimal (saturated) condition
Pri_j	Primary crash for the identified secondary crash *j*
ql_1	End of impact area at the time of crash *j*
ql_2	Start of impact are, at the time of crash *j*
t	Duration between primary and secondary crash occurrence
t_1	Time of occurrence of primary crash
t_2	Time of occurrence of secondary crash
T_c	Primary crash clearance duration

3.1 Static Approach

Identification of SCs using a static approach requires selection of pre-specified temporal and spatial threshold values. In addition, directionality and location (impact region) of a PC play a crucial role and needs to be predefined. Directionality refers to the direction of the PC as compared to the SC (i.e. same or opposite direction). Location refers to the upstream or downstream location of the SC with respect to the direction of flow and location of PC. For the static approach, five possible combinations of directionality and location were considered in this study (graphically depicted in Fig. 3-2), capturing all possible types of SCs. These five cases are defined as follows:

- Case-1: Same Direction-Upstream: SC occurs in the upstream same direction of the PC
- Case-2: Opposite Direction-Upstream: SC occurs in the upstream opposite direction of the PC
- Case-3: Opposite Direction-Downstream: SC occurs in the downstream opposite direction of the PC
- Case-4: (Combination of cases 1 and 2): SC occurs either in the downstream or upstream opposite direction of the PC
- Case-5: Cases 1, 2, and 3 combined

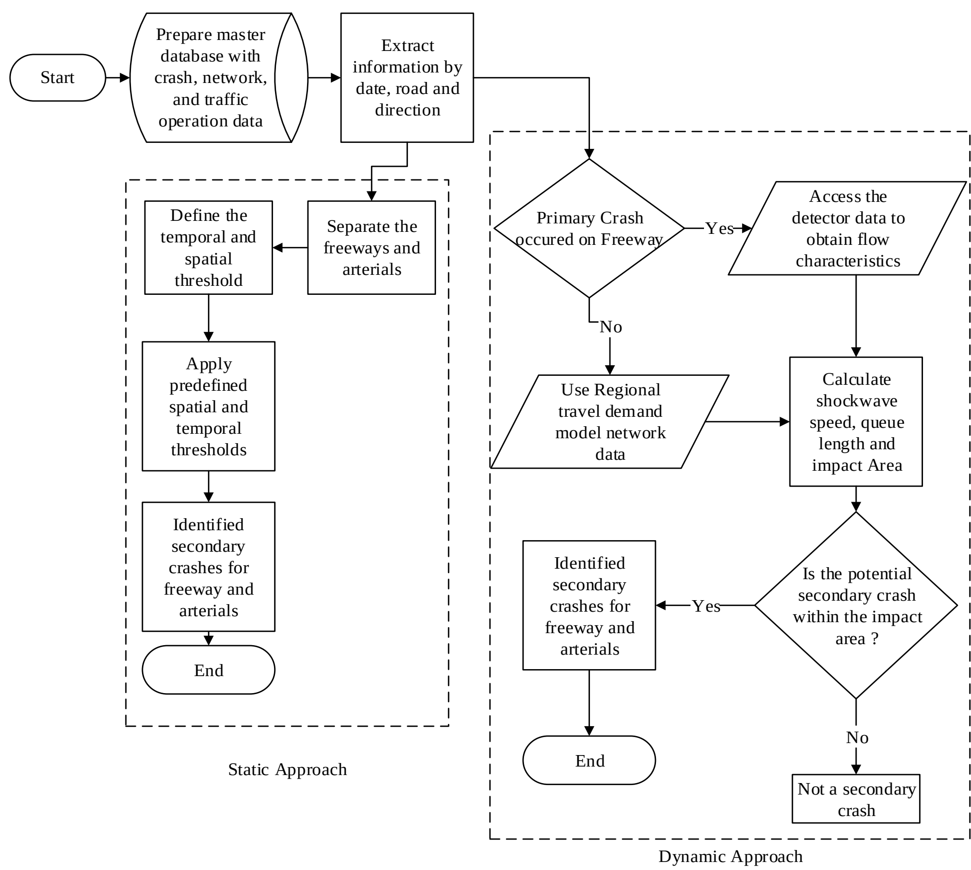


Figure 3-1 Flow chart showing the methodology.

For the static approach, in all five cases, spatiotemporal thresholds are predefined by the user. As an example, one can consider a one mile/one hour thresholds. For Cases-2 and 3, spatial threshold is applied in the opposite direction to account for the '*rubbernecking*' effect. Rubbernecking represents the phenomenon when drivers in the opposite direction slow down to observe the PC causing congestion, reduction in capacity, and associated delays (Saddi, 2009). Rubbernecking effects depends on the facility type, traffic conditions, type and severity of an incident, and has a significant potential of inducing SCs in the opposite direction of a PC (Saddi, 2009).

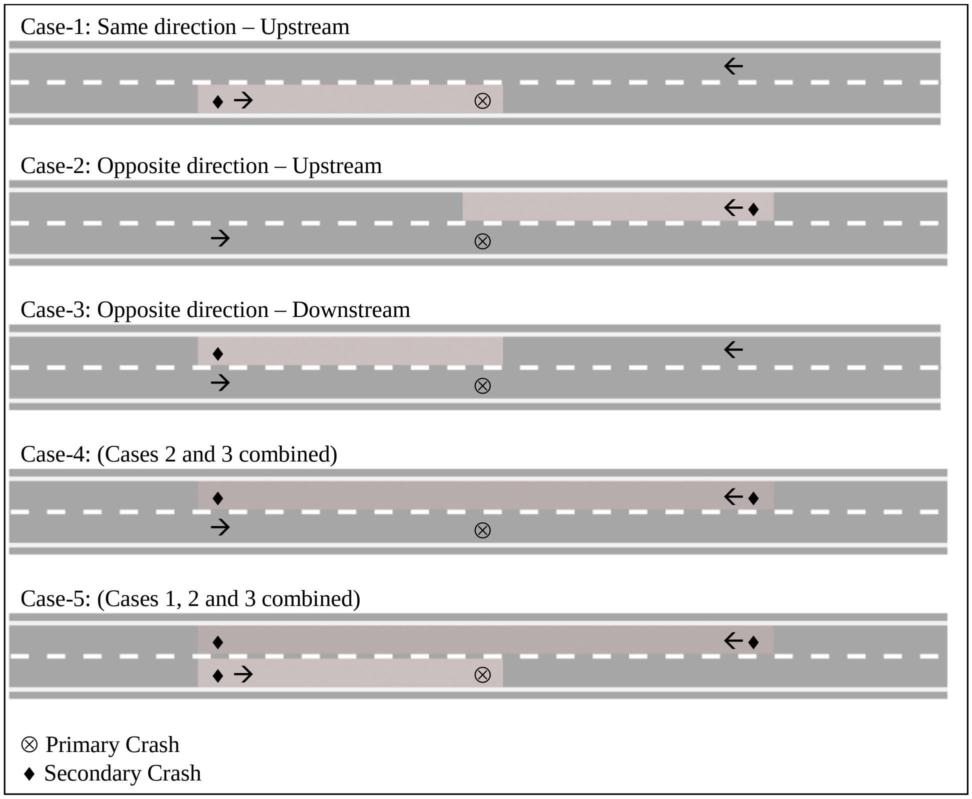


Figure 3-2 Pictorial representation of directionality and locations of SCs.

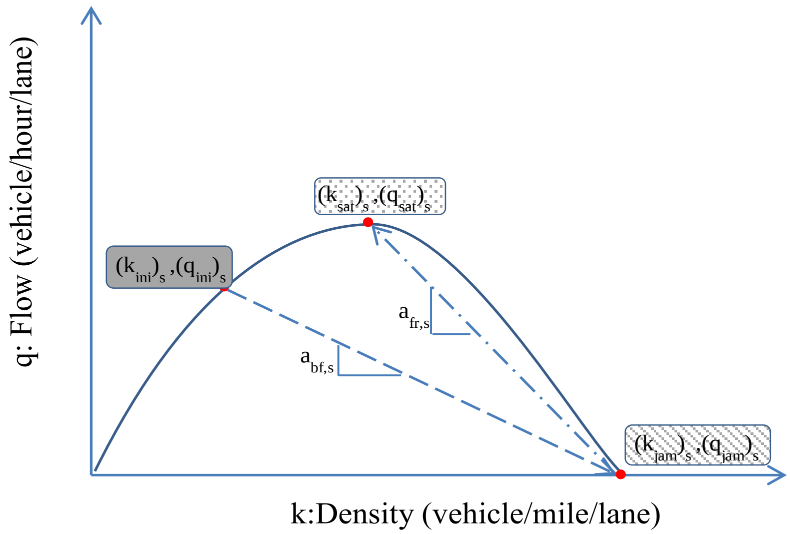
3.2 Dynamic Approach

The dynamic approach in SC identification aims to better capture effects of traffic characteristics (e.g. flow, speed, and density), that change over time and space, and affect both queue formation from a PC and SC occurrence. With a given state and lane specific traffic flow parameters, continuously monitored by closely spaced sensors or other devices near the crash location (flow, density, speed, number of lanes, location of the crash on a specific lane etc.), it is possible to calculate queue lengths using shockwave theory (Lighthill and Whitham, 1955). In this subsection we present a dynamic threshold SC incident identification approach to estimate the impact area of a PC created by a backward-forming and forward-recovery shockwave. Backward-forming shockwave affects the growth rate of the queue formed by the PC. Once the PC is cleared, a forward-recovery shockwave is set in motion and eventually catches up with the backward-forming shockwave resulting in dissipation of the queue. Next we discuss the steps required to estimate the impact area using the shockwave principle.

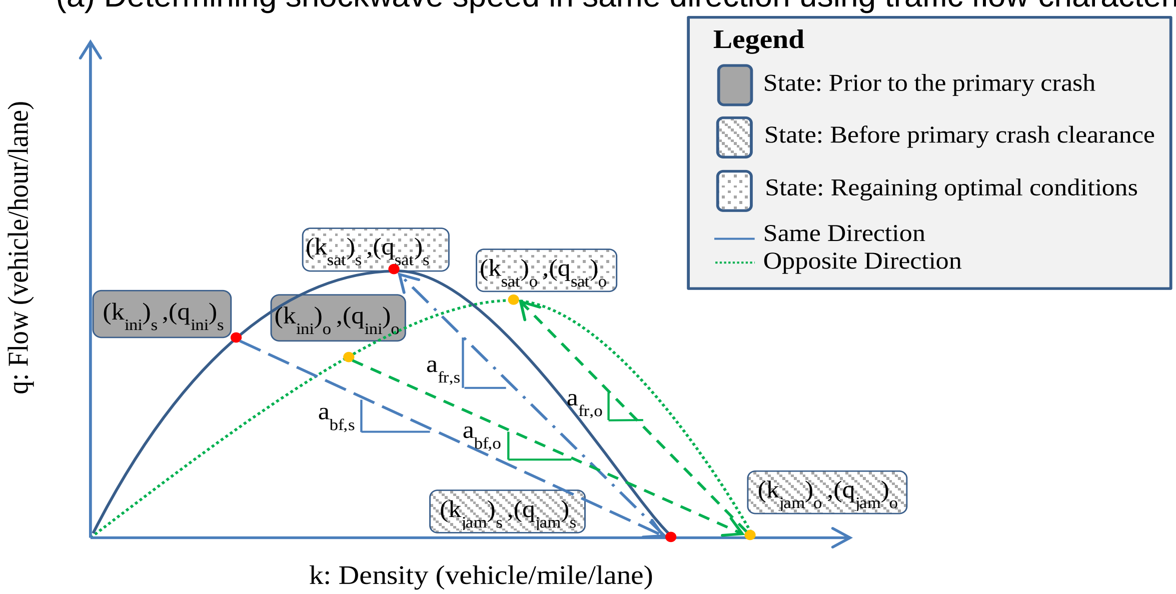
3.2.1 Estimation of backward- forming and forward-recovery shockwaves

A generalized density-flow curve is shown in Fig. 3-3(a) where $(k_{ini})_s$ and $(q_{ini})_s$ are the initial conditions of density and flow where the initial speed, $(u_{ini})_s$ is the slope of the curve. If one or more lanes are completely closed (often the case) due to a PC, then the traffic state is represented by $(k_{jam})_s$, and $(q_{jam})_s$ (until the clearance period),where $(k_{jam})_s$ represents jam density and both flow ($(q_{jam})_s$) and speed ($(u_{jam})_s$) are equal to zero. However, if density at this state is not equal to $(k_{jam})_s$, any flow/density state, represented by the parabola, can be used. Speed of the backward-forming shockwave, is equal to:

$$a_{bf,s} = \frac{(q_{ini})_s - (q_{jam})_s}{(k_{ini})_s - (k_{jam})_s}.$$



(a) Determining shockwave speed in same direction using traffic flow characteristics.



(b) Determining shockwave speed in opposite direction using traffic flow characteristics.

Figure 3-3 Shockwave speed for single and bi-directional traffic.

Once the PC is cleared, the queued traffic state will try to reach optimal conditions ($(k_{sat})_s$, $(q_{sat})_s$ and $(u_{sat})_s$) forming a forward-recovery shockwave with a speed of:

$$a_{fr,s} = \frac{(q_{jam})_s - (q_{sat})_s}{(k_{jam})_s - (k_{sat})_s}.$$

A similar approach can be adopted to analyze shockwaves in the opposite direction. Fig. 3-3(b) demonstrates traffic states for bi-directional traffic where $(k_{ini})_o$, $(k_{jam})_o$ and $(k_{sat})_o$ represent current, jam and optimal density states for the opposite direction, and $(q_{ini})_o$, $(q_{jam})_o$, and $(q_{sat})_o$ are the respective flow states.

3.2.2 Impact Area Estimation

Determining the impact area of a PC requires the clearance time (T_c) and the time difference between occurrence of PC and the "potential" SC ($t = t_2 - t_1$). The impact area (d) is defined as:

$$a_{fr,s} \times (t - T_c) \leq d \leq a_{bf,s} \times t, \text{ when } t > T_c$$

$$0 \leq d \leq a_{bf,s} \times t, \text{ when } t < T_c$$

In this research, when estimating the impact area, clearance time for the primary incident was available through an incident management database. Clearance time varies and depends on crash type and severity, number of vehicles involved, number of lanes, availability of shoulder area etc. Fig. 3-4 shows the impact area (shaded area between the backward-forming and forward-recovery shockwaves) which captures the portion of the queue, from the primary incident, which can induce a SC. Note that: a) the forward recovery shockwave does not set off until the primary incident is cleared (i.e. size of the impact area depends on the PC clearance time) and, b) higher recovery shockwave speed results in faster queue dissipation.

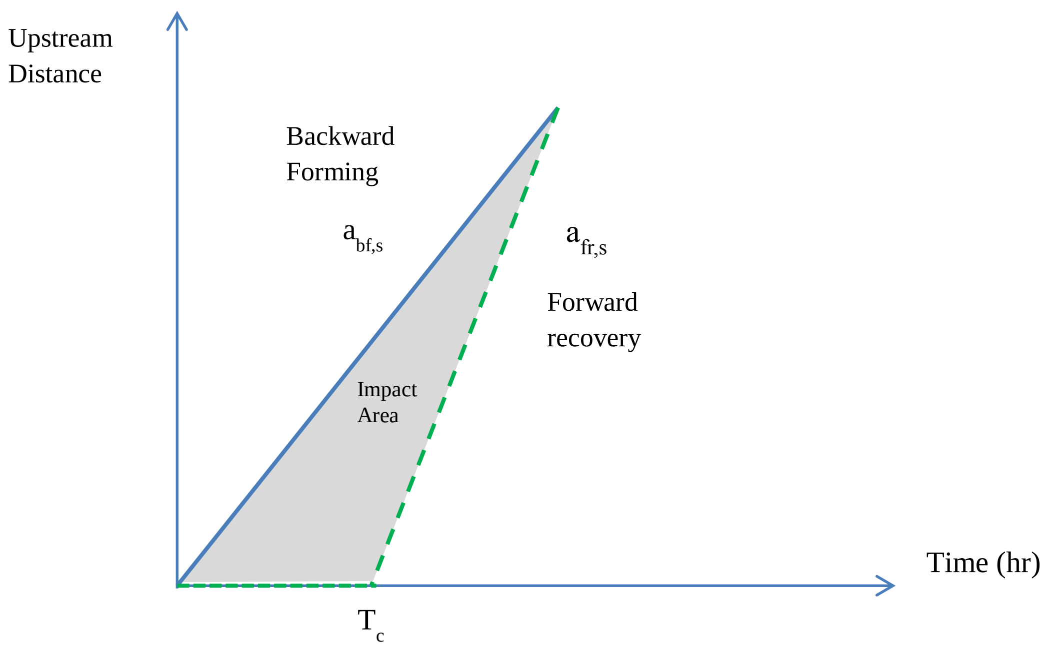


Figure 3-4 Graphical representation of impact area.

3.3 Scope of Case Study

The proposed methodology presented in the previous section was applied and evaluated using the transportation network of Shelby County, Tennessee (TN), in the United States (U.S.). Shelby County is an ideal case study candidate as the most populous county in the state of TN, home to one of the largest freight intermodal hubs in the US, and the largest metropolitan planning organization in the tri-state encompassing portions of Tennessee, Arkansas and Mississippi, with a significant portion of inter-state traffic. The following describes data collected for the case study:

- Roadway Network: A detailed transportation network (20,289 links/1,619 miles) with 20 different functional classes of roadways (1,337 miles of arterials and 282 miles of freeways) was available from TDOT (Fig. 3-5).

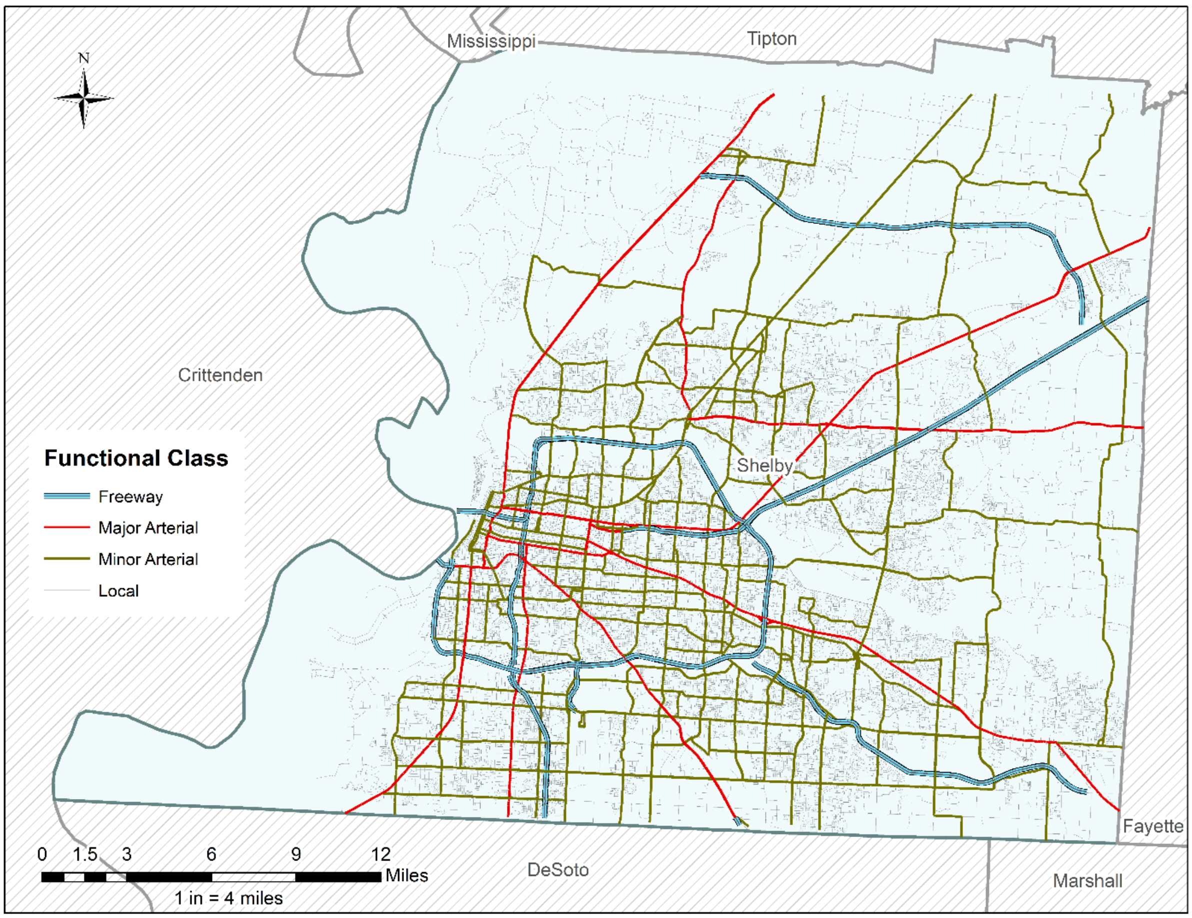


Figure 3-5 Roadway network in Shelby County.

- Crash data: Three years (2010-2012) of crash data, from the Tennessee Roadway Information Management System (TRIMS); a total of 91,325 crashes.
- Freeway Traffic Data: Lane specific traffic data by minute (speed, flow, occupancy etc.) aggregated into 15 minute intervals.

- <u>Arterial Traffic Data</u>: Traffic data on arterials were not available in such detailed manner as freeways. Link speed and flow were obtained from the Metropolitan Planning Organization (MPO) travel demand model.
- <u>Incident Management data:</u> Data on all reported incidents (e.g. time of crash occurrence, time taken for the rescue vehicle to reach incident location, clearance time, etc.) were available from the incident management system in TN.

A geodatabase was developed from these five data sets with facility types categorized into two groups: freeways or arterials. In this study, rural and urban interstates, and expressways were grouped into the freeways category, while rural and urban principal and minor arterials were grouped into the arterials category.

3.4 Secondary Crash Identification Algorithm (SCIA)

The algorithm developed to identify SC (***SCIA***), shown in Fig. 3-6, involves two major steps: a) crash pairing, and b) SC identification which are discussed in the following sections. Both steps are discussed in detail next.

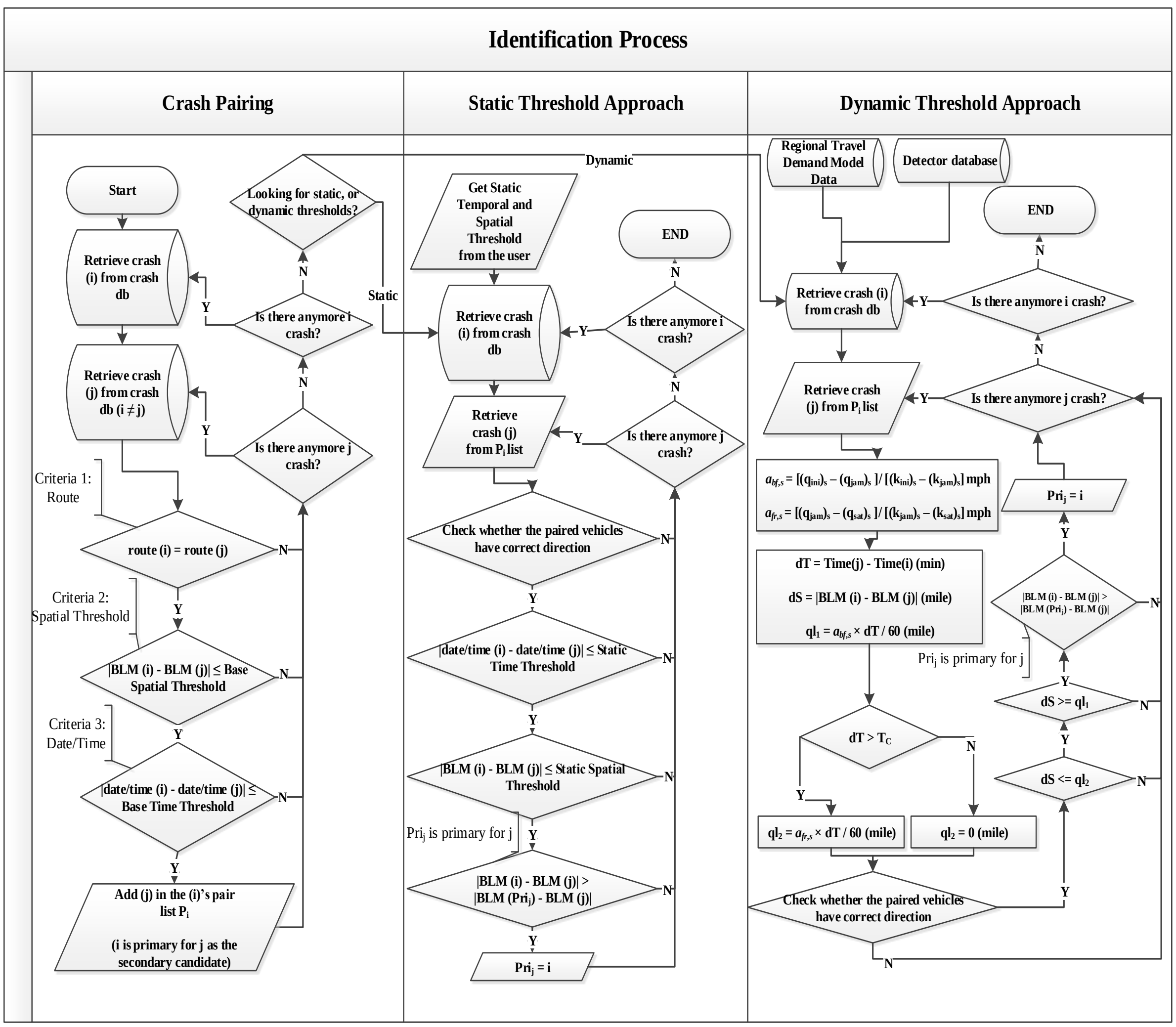


Figure 3-6 Flowchart representing the algorithms.

3.4.1 Step 1: Crash pairing

The first step of ***SCIA*** involves crash pairing which identifies candidate SCs, given a PC, using various criteria such as day of occurrence, route, and spatiotemporal thresholds. Accuracy of this procedure is crucial in reducing the complexity of the remaining steps of the algorithm. During pairing, any crash available in the database is considered as primary with all other crashes identified as candidate SCs to that particular crash. A crash is then considered as a SC and paired with a PC upon satisfying all the criteria. Distance between crashes was determined using the absolute difference in Beginning Log Mile (BLM) (as a reference point). The position of the paired crashes, with respect to each other, was determined using their direction, BLM and their respective coordinates.

3.4.2 Step 2: SC Identification

Once crash pairing is complete, SCs are identified using both the static and dynamic approach. For the static approach, only spatiotemporal thresholds were considered as criteria for identifying a SC. These thresholds can be set by the user. For the dynamic approach, traffic flow characteristics before the occurrence of a PC were required to estimate the impact area. These data were obtained either from detector datasets (for freeway) or the regional MPO travel demand model (for arterials). After the PC occurrence, one or more lanes are completely closed for the duration of the clearance time (T_c) and hence, jam condition is assumed. For this condition, we considered [$(q_{jam})_s$, $(u_{jam})_s$]=0 (i.e. there is no flow of traffic), and $(k_{jam})_s$ = 5280/25 = 211 veh/mi/lane assuming average vehicle length of 25 ft. including 5 ft. of safety distance between lead and following vehicle. After the primary incident is cleared, flow conditions will eventually reach saturated condition where $(q_{sat})_s$, $(u_{sat})_s$ are assumed to be 1900 veh/hr/lane, 65 mph (for freeway) or 1800 veh/hr/lane, 45 mph respectively (for arterial). $(k_{sat})_s$ is calculated accordingly using the basic density-flow-speed formula ($(k_{sat})_s=(q_{sat})_s/(u_{sat})_s$).

The pseudocode for SC identification is as following:

<u>Static Approach pseudocode (Case-1)*</u>

Let $\boldsymbol{R_i}$ = Route on which crash ***i*** is located $\boldsymbol{D_i}$ = Direction of the route for crash ***i***
    $\boldsymbol{T_i}$ = Date of occurrence of crash ***i*** $\boldsymbol{t_i}$ = Time of occurrence of crash ***i***
    $\boldsymbol{P_i}$ = Set of secondary crashes for ***i*** and set $\boldsymbol{P_i = \emptyset}$
For each crash ***i***, $\forall$ ***i*** $\in$ ***I***
        For each other crash ***j***, $\forall$ ***j*** $\in$ ***I*** and (i≠j)
Step 1: Check the route and direction for both crashes
        If $R_i = R_j$ & $D_i = D_j$ & crash ***j*** is in the upstream of crash ***i***
            Go to **Step 2,**
        Else Skip crash ***j*** and **Steps 2, 3** and continue with **Step 1**

Step 2: Check if two crashes are both occurring on the same day

If $T_i = T_j$

Go to **Step 3**

Else Skip crash ***j*** and **Step 3** and Go back to **Step 1**

Step 3: Check if spatiotemporal threshold is satisfied

If $|t_i - t_j| \leq$ Time Threshold & $|BLM_i - BLM_j| \leq$ Spatial Threshold

i is the primary crash and ***j*** is the secondary crash

$\boldsymbol{P_i = P_i \cup j}$

*Note: This pseudocode is for Case-1(static approach) only. For other cases, only direction criteria will be modified in Step 1.

Dynamic Approach pseudocode (Case-1)*

Let $\boldsymbol{R_i}$ = Route on which crash ***i*** is located $\boldsymbol{D_i}$ = Direction of the route for crash ***i***

$\boldsymbol{T_i}$ = Date of occurrence of crash ***i*** $\boldsymbol{t_i}$ = Time of occurrence of crash ***i***

$\boldsymbol{dT} = |t_i - t_j|$ $\boldsymbol{dS} = |BLM_i - BLM_j|$

$\boldsymbol{P_i}$ = Set of secondary crashes for ***i*** and set $\boldsymbol{P_i} = \emptyset$

Step 0: Set the default parameters by given network facility type: q_{jam}, u_{jam}, k_{jam}, q_{sat}, u_{sat}

For each crash ***i***, $\forall\, \boldsymbol{i} \in \boldsymbol{I}$

For each other crash ***j***, $\forall\, \boldsymbol{j} \in \boldsymbol{I}$ and ($i \neq j$)

Step 1: Check the route and direction for both crashes

If $R_i = R_j$ & $D_i = D_j$ & crash ***j*** is in the upstream of crash ***i***

Go to **Step 2,**

Else Skip crash ***j*** and continue with **Step 1**

Step 2: Check if two crashes are both occurring on the same day

If $T_i = T_j$

Go to **Step 3**

Else Skip crash ***j*** and Go back to **Step 1**

Step 3: Obtain traffic volume (q_{ini}) and speed (u_{ini}) at the time of a given PC and calculate density $\boldsymbol{k_{ini}}$, $\boldsymbol{a_{bf,s}}$ and $\boldsymbol{a_{fr,s}}$

Step 4: Calculate

$\boldsymbol{ql_1} = a_{bf,s} \times dT/60$

If ($dT > T_c$)

$\boldsymbol{ql_2} = a_{fr,s} \times (dT - T_c)/60$

Else $\boldsymbol{ql_2} = 0$

Step 5: Check if the crash ***j*** is within the impact area

If $ql_2 \leq dS \leq ql_1$

i is the primary crash and ***j*** is the secondary crash

$\boldsymbol{P_i = P_i \cup j}$

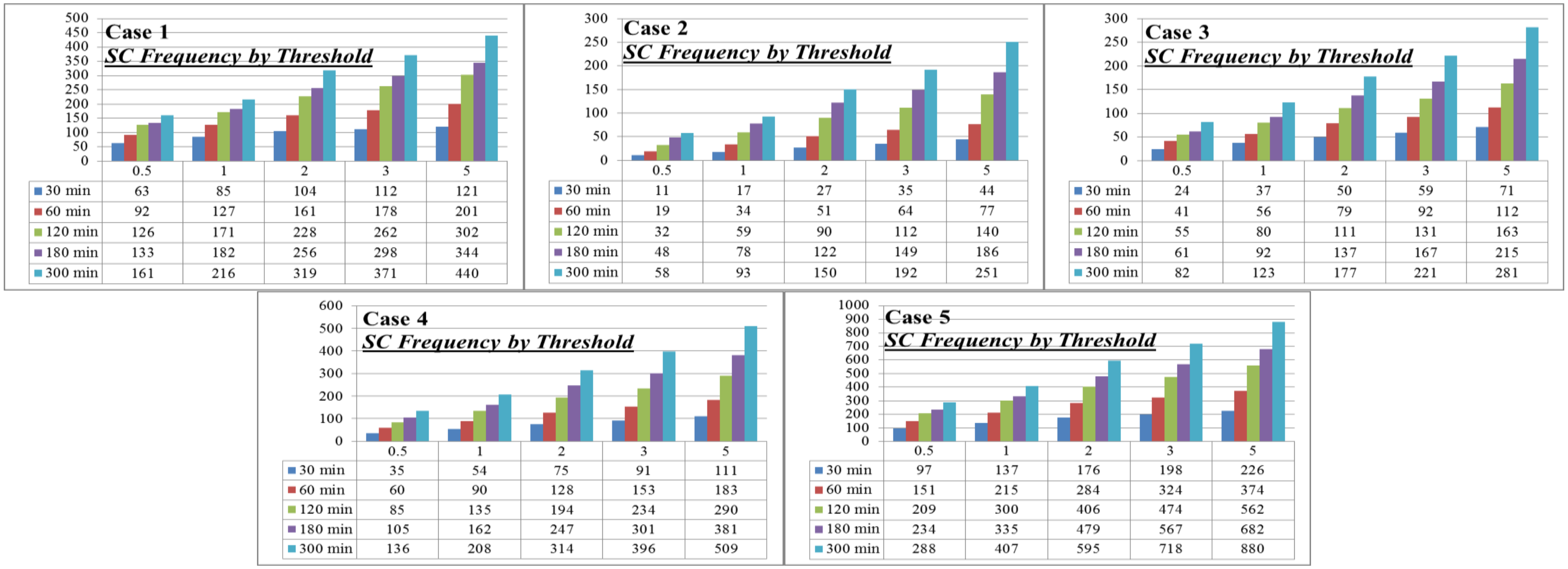
*Note: This pseudocode is for Case-1(dynamic approach) only. For other cases, only direction criteria will be modified in Step 1 and shockwave speed will be calculated for that particular direction in Step 3.

3.5 Results

SCs were classified into two categories based on facility type (i.e. SCs on freeways or arterials) to account for the significant differences in flow, speed and density characteristics of the two facility types. Additionally, incident management on urban arterial roadways area is considerably different than on freeways and will effect SC occurrence (Raub and Schofer, 1997). For the static approach different spatiotemporal thresholds were used to determine sensitivity and to assess over/under estimation of SC identification when compared to the dynamic approach. Temporal thresholds of 30, 60, 120, 180 and 300 minutes were used along with spatial thresholds of 0.5, 1, 2, 3 and 5 miles. Larger thresholds (e.g. over 120 minutes and 2 miles) were used to account for freeway queuing during peak periods.

3.5.1 Static approach

SCs identified for all five cases using different spatiotemporal threshold values by facility type (freeway and arterial) are presented in Fig. 3-7. It is observed that SC occurrences increase as the spatial threshold increases (for all cases and facility types). In general, higher number of SCs and higher rates are observed on arterials than freeways, which can be explained by the larger number of lane-miles covered by arterials. Note that Case-1 (same direction-upstream) has a significantly larger number of SCs for both facility types when compared to Cases-2 and 3 as a PC is more likely to cause congestion upstream in the same direction than the opposite which in turn may lead to SCs.



Case 1

	0.5	1	2	3	5
30 min	63	85	104	112	121
60 min	92	127	161	178	201
120 min	126	171	228	262	302
180 min	133	182	256	298	344
300 min	161	216	319	371	440

Case 2

	0.5	1	2	3	5
30 min	11	17	27	35	44
60 min	19	34	51	64	77
120 min	32	59	90	112	140
180 min	48	78	122	149	186
300 min	58	93	150	192	251

Case 3

	0.5	1	2	3	5
30 min	24	37	50	59	71
60 min	41	56	79	92	112
120 min	55	80	111	131	163
180 min	61	92	137	167	215
300 min	82	123	177	221	281

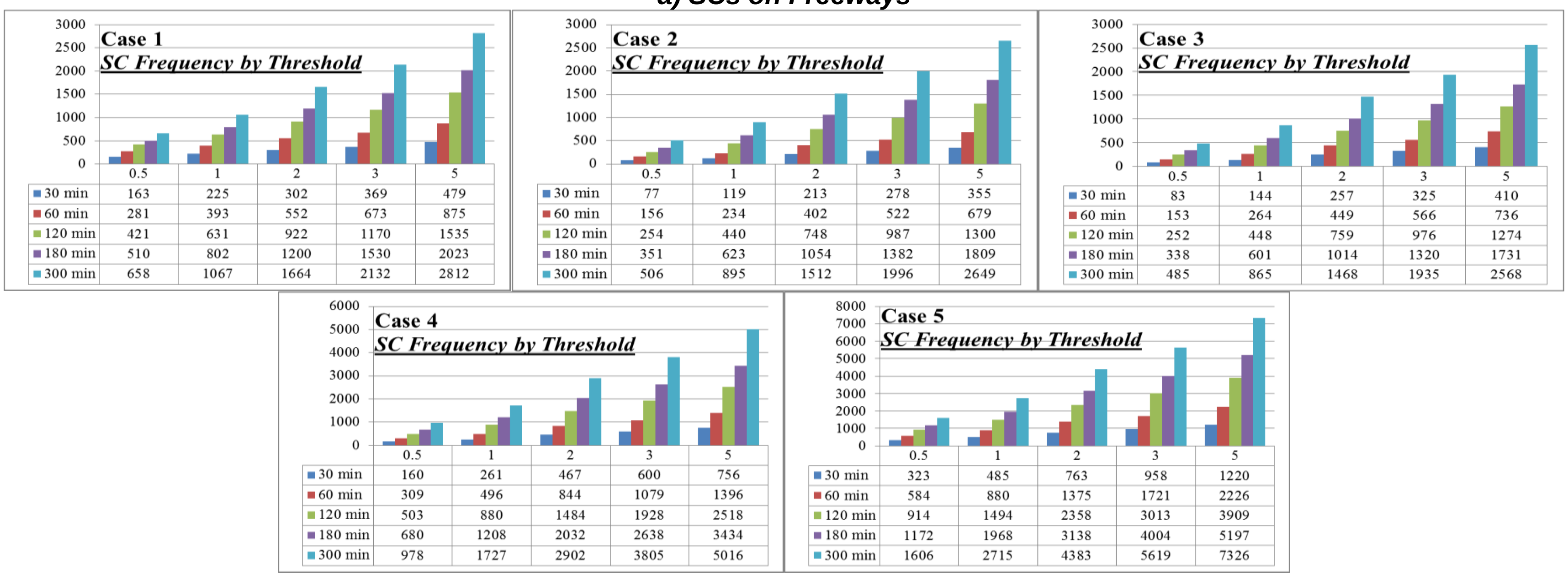
Case 4

	0.5	1	2	3	5
30 min	35	54	75	91	111
60 min	60	90	128	153	183
120 min	85	135	194	234	290
180 min	105	162	247	301	381
300 min	136	208	314	396	509

Case 5

	0.5	1	2	3	5
30 min	97	137	176	198	226
60 min	151	215	284	324	374
120 min	209	300	406	474	562
180 min	234	335	479	567	682
300 min	288	407	595	718	880

a) SCs on Freeways



Case 1

	0.5	1	2	3	5
30 min	163	225	302	369	479
60 min	281	393	552	673	875
120 min	421	631	922	1170	1535
180 min	510	802	1200	1530	2023
300 min	658	1067	1664	2132	2812

Case 2

	0.5	1	2	3	5
30 min	77	119	213	278	355
60 min	156	234	402	522	679
120 min	254	440	748	987	1300
180 min	351	623	1054	1382	1809
300 min	506	895	1512	1996	2649

Case 3

	0.5	1	2	3	5
30 min	83	144	257	325	410
60 min	153	264	449	566	736
120 min	252	448	759	976	1274
180 min	338	601	1014	1320	1731
300 min	485	865	1468	1935	2568

Case 4

	0.5	1	2	3	5
30 min	160	261	467	600	756
60 min	309	496	844	1079	1396
120 min	503	880	1484	1928	2518
180 min	680	1208	2032	2638	3434
300 min	978	1727	2902	3805	5016

Case 5

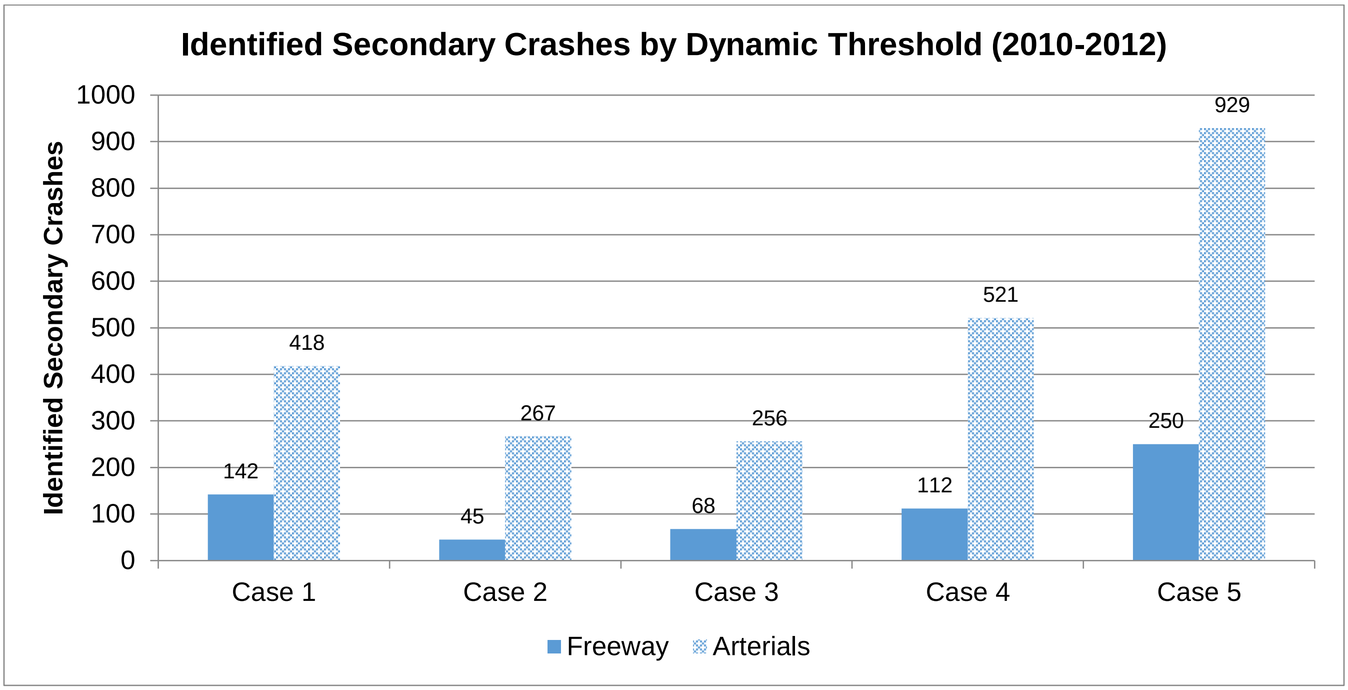
	0.5	1	2	3	5
30 min	323	485	763	958	1220
60 min	584	880	1375	1721	2226
120 min	914	1494	2358	3013	3909
180 min	1172	1968	3138	4004	5197
300 min	1606	2715	4383	5619	7326

b) SCs on Arterials

Figure 3-7 SCs identified using static approach (freeways and arterials).

3.5.2 Dynamic approach

Frequencies of SCs, identified using the dynamic approach, for all five directionality/location cases for freeways and arterials, are shown in Fig 3-8. For freeways, Case-1 exhibits a higher number of SCs when compared to Cases-2 and 3 combined, while SCs for Case-3 results in a higher frequency than Case-2 (142 SCs identified on freeways for Case-1 as compared to 45 and 68 for Case-2 and 3 respectively). The same trend is not observed for arterials as the rubbernecking effect is more prominent. A total of 1,179 SCs (freeways and arterials combined) are identified using the dynamic approach (Fig 3-8) which is comparable to the 1,095 crashes (Case-5) identified for one mile and one hour static threshold (Fig. 3-7).

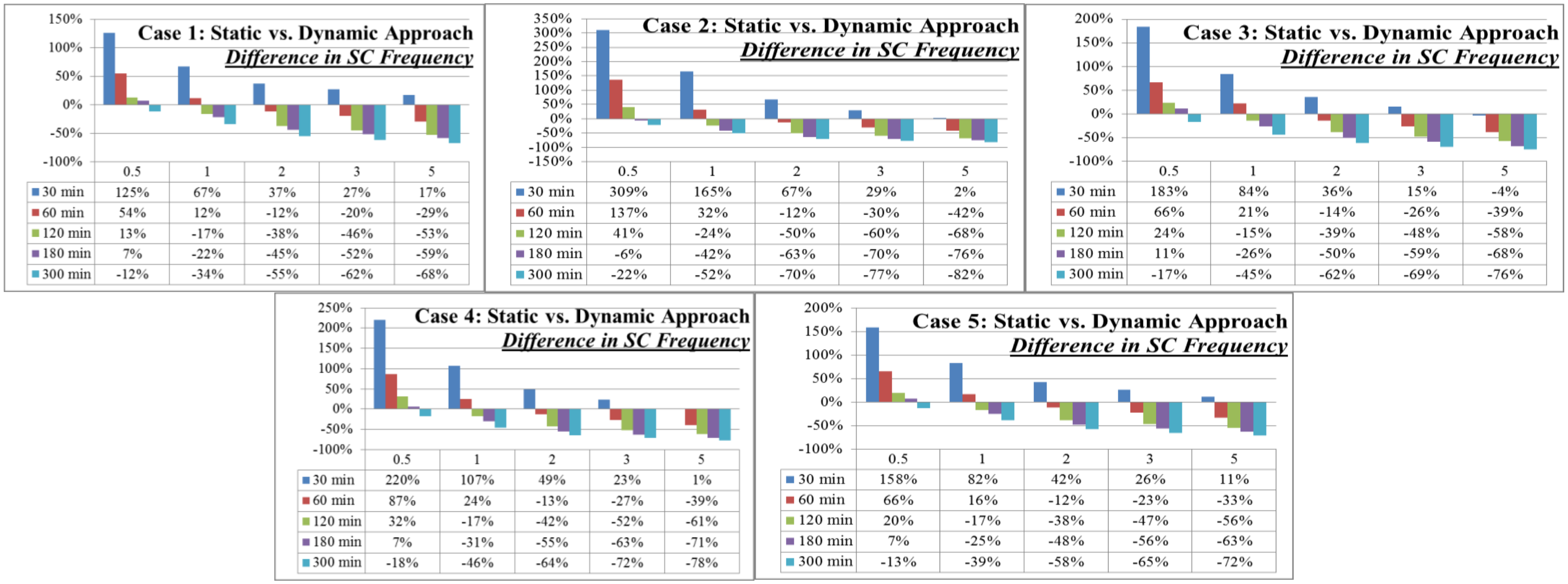


Note: Duplicate secondary crashes are removed from case 4 and case 5 as several secondary crashes overlapped with multiple case scenarios.

Figure 3-8 SCs identified using dynamic threshold (freeway and arterials).

3.5.3 Static vs. Dynamic approach: SC frequencies

Comparison of both approaches in terms of SCs identification is presented in Fig. 3-9. Results shown in Fig. 3-9(a) (freeways) and 3-9(b) (arterials) reveal that the static approach over-estimates SC frequencies as spatiotemporal thresholds increase. As expected, for low spatial/temporal thresholds (e.g. 30, 60 min and 0.5, 1 mile) the static approach underestimates SC frequencies. Overall, when comparing results from the static and dynamic approach, the number of SCs identified using the latter are significantly less when compared to SCs for larger thresholds used by the static approach.



Case 1:

	0.5	1	2	3	5
30 min	125%	67%	37%	27%	17%
60 min	54%	12%	-12%	-20%	-29%
120 min	13%	-17%	-38%	-46%	-53%
180 min	7%	-22%	-45%	-52%	-59%
300 min	-12%	-34%	-55%	-62%	-68%

Case 2:

	0.5	1	2	3	5
30 min	309%	165%	67%	29%	2%
60 min	137%	32%	-12%	-30%	-42%
120 min	41%	-24%	-50%	-60%	-68%
180 min	-6%	-42%	-63%	-70%	-76%
300 min	-22%	-52%	-70%	-77%	-82%

Case 3:

	0.5	1	2	3	5
30 min	183%	84%	36%	15%	-4%
60 min	66%	21%	-14%	-26%	-39%
120 min	24%	-15%	-39%	-48%	-58%
180 min	11%	-26%	-50%	-59%	-68%
300 min	-17%	-45%	-62%	-69%	-76%

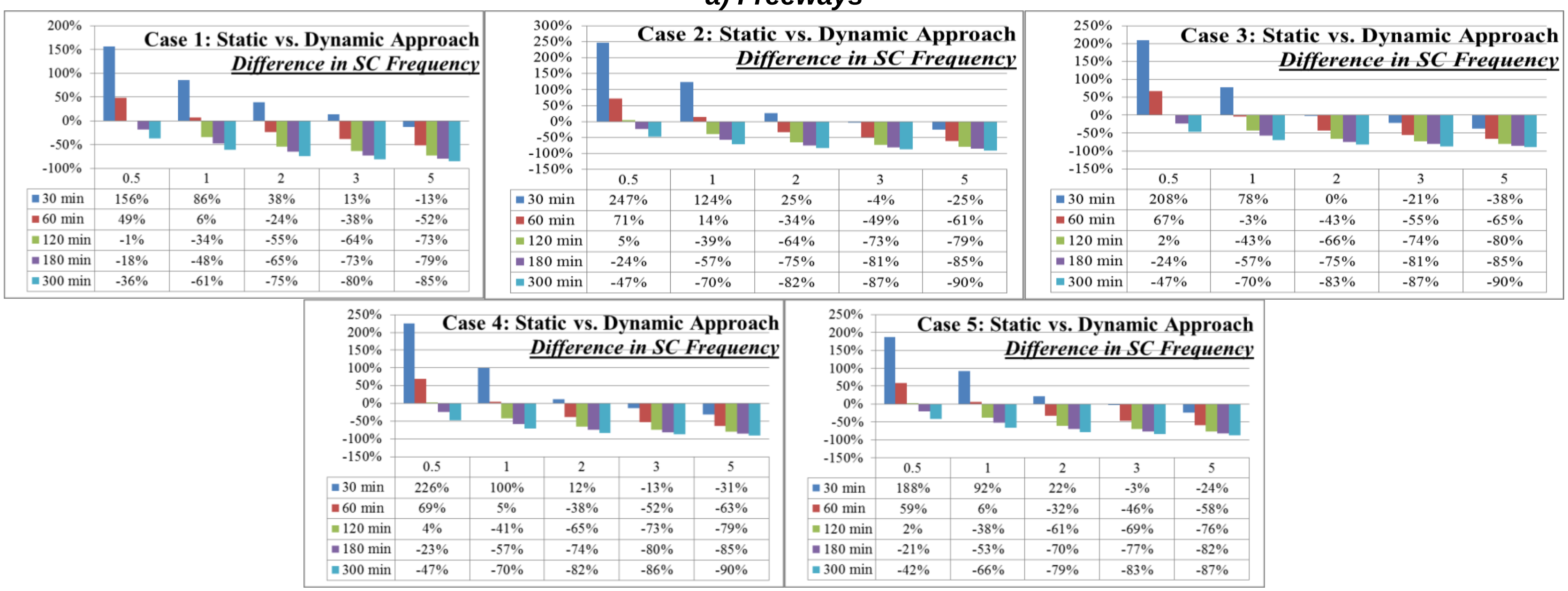
Case 4:

	0.5	1	2	3	5
30 min	220%	107%	49%	23%	1%
60 min	87%	24%	-13%	-27%	-39%
120 min	32%	-17%	-42%	-52%	-61%
180 min	7%	-31%	-55%	-63%	-71%
300 min	-18%	-46%	-64%	-72%	-78%

Case 5:

	0.5	1	2	3	5
30 min	158%	82%	42%	26%	11%
60 min	66%	16%	-12%	-23%	-33%
120 min	20%	-17%	-38%	-47%	-56%
180 min	7%	-25%	-48%	-56%	-63%
300 min	-13%	-39%	-58%	-65%	-72%

a) Freeways



Case 1:

	0.5	1	2	3	5
30 min	156%	86%	38%	13%	-13%
60 min	49%	6%	-24%	-38%	-52%
120 min	-1%	-34%	-55%	-64%	-73%
180 min	-18%	-48%	-65%	-73%	-79%
300 min	-36%	-61%	-75%	-80%	-85%

Case 2:

	0.5	1	2	3	5
30 min	247%	124%	25%	-4%	-25%
60 min	71%	14%	-34%	-49%	-61%
120 min	5%	-39%	-64%	-73%	-79%
180 min	-24%	-57%	-75%	-81%	-85%
300 min	-47%	-70%	-82%	-87%	-90%

Case 3:

	0.5	1	2	3	5
30 min	208%	78%	0%	-21%	-38%
60 min	67%	-3%	-43%	-55%	-65%
120 min	2%	-43%	-66%	-74%	-80%
180 min	-24%	-57%	-75%	-81%	-85%
300 min	-47%	-70%	-83%	-87%	-90%

Case 4:

	0.5	1	2	3	5
30 min	226%	100%	12%	-13%	-31%
60 min	69%	5%	-38%	-52%	-63%
120 min	4%	-41%	-65%	-73%	-79%
180 min	-23%	-57%	-74%	-80%	-85%
300 min	-47%	-70%	-82%	-86%	-90%

Case 5:

	0.5	1	2	3	5
30 min	188%	92%	22%	-3%	-24%
60 min	59%	6%	-32%	-46%	-58%
120 min	2%	-38%	-61%	-69%	-76%
180 min	-21%	-53%	-70%	-77%	-82%
300 min	-42%	-66%	-79%	-83%	-87%

b) Arterials

Figure 3-9 Static vs. Dynamic approach SC comparison .

3.5.4 Dynamic Approach: SC Distribution by Time of Day

Fig. 3-10 shows the time of day distribution of SCs identified using the dynamic approach on both freeways and arterials. Due to space limitations, only results for Case-1 are presented. Freeway facilities exhibit two distinct peaks: AM peak (between 8am- 9am), and PM peak (between 5pm-7pm). Both peak periods account for 59% of the total number of identified SCs for Case-1. On the other hand, arterials exhibit a very prominent PM peak (4pm-6pm) when compared to AM peak. SCs identified in the PM peak for arterials account for 48% of all SCs for Case-1. The reason that arterials have only one noticeable peak can be explained by the larger number of PCs occurring in the PM peak as compared to the AM peak. These results are in line with findings from the reviewed literature (Hirunyanitiwattana and Mattingly, 2006). Note that the majority of SCs observed late at night (10pm-3am) occurred during the last week of December and might be the results of high traffic from special events (Christmas break, winter weather etc.).

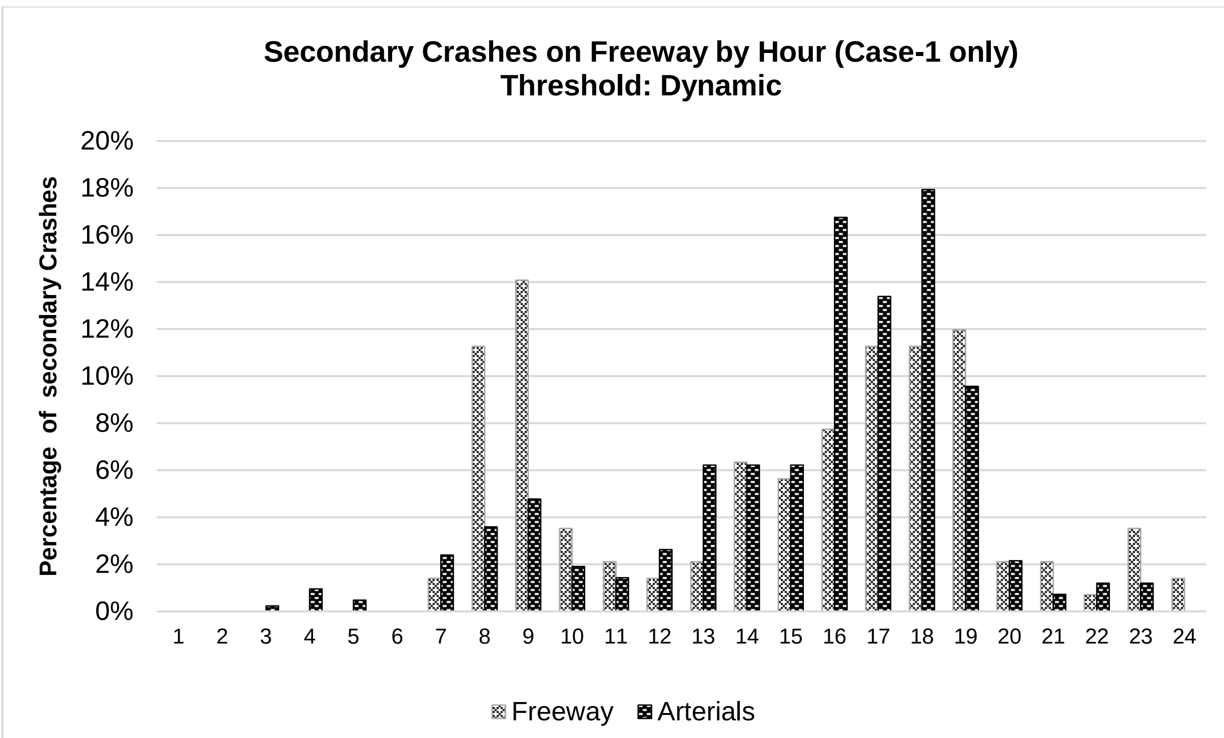


Figure 3-10 SCs (Case-1) by time of day using dynamic approach.

3.5.5 Dynamic Approach: SC Occurrence by Facility Type

Classifying SCs based on the facility types can support the assumption that SCs depend on intrinsic characteristics of facilities. The roadway network, used for case study, consists of approximately 285 miles (7.40%) of interstate/freeway facility but encountered 21.10% of SCs during 2010-12, whereas 56.40% of SCs occurred on 972 miles (25.25%) of major arterials as shown in Fig. 3-11. On the other hand, only 0.90% of SCs were identified on 2,163 miles (56.18%) of local/collectors during the same period. One of the primary reasons is the different travelling speed and traffic volume for different facility types. On freeways, interstates, and major arterials the average speed and traffic volume is much higher than on minor arterials and local collectors. Hence a primary crash on a facility with moderate to high speed and volume has a higher potential to induce SCs. The results are similar to findings in a past study (Hirunyanitiwattana and Mattingly, 2006).

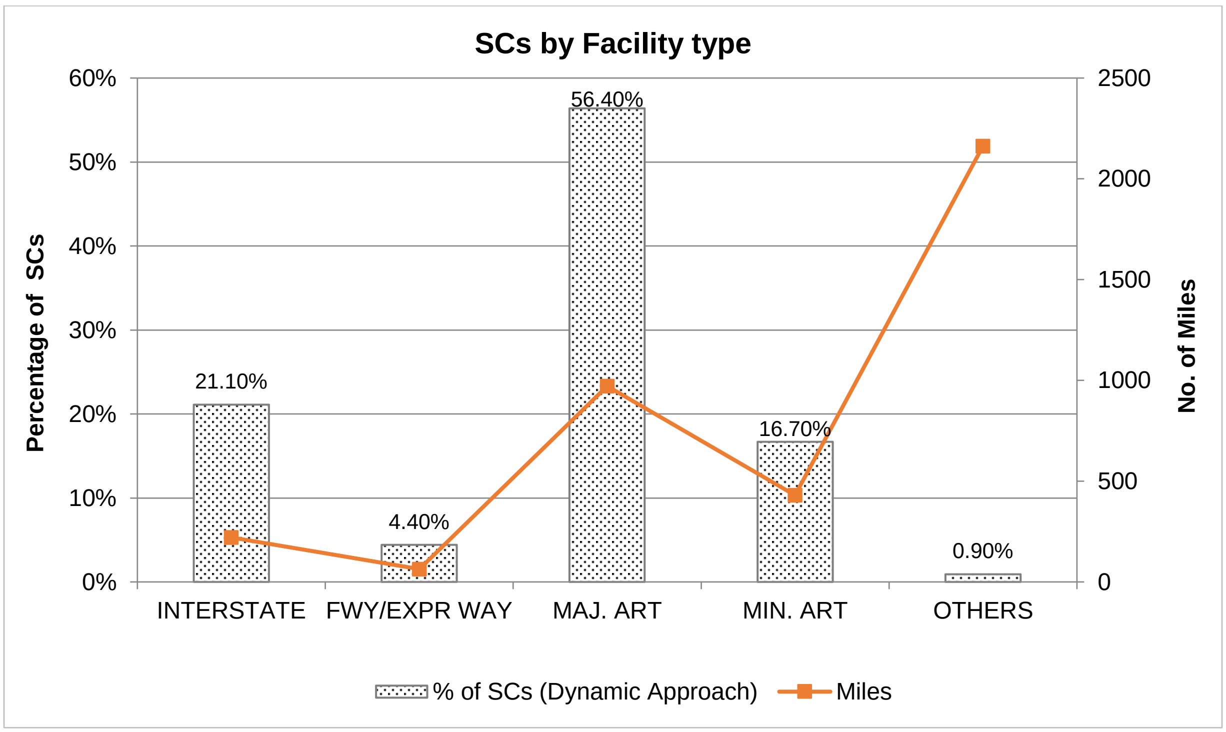


Figure 3-11 Identified SCs (using dynamic approach) by facility type.

3.5.6 Traffic Volume and SC Occurrence

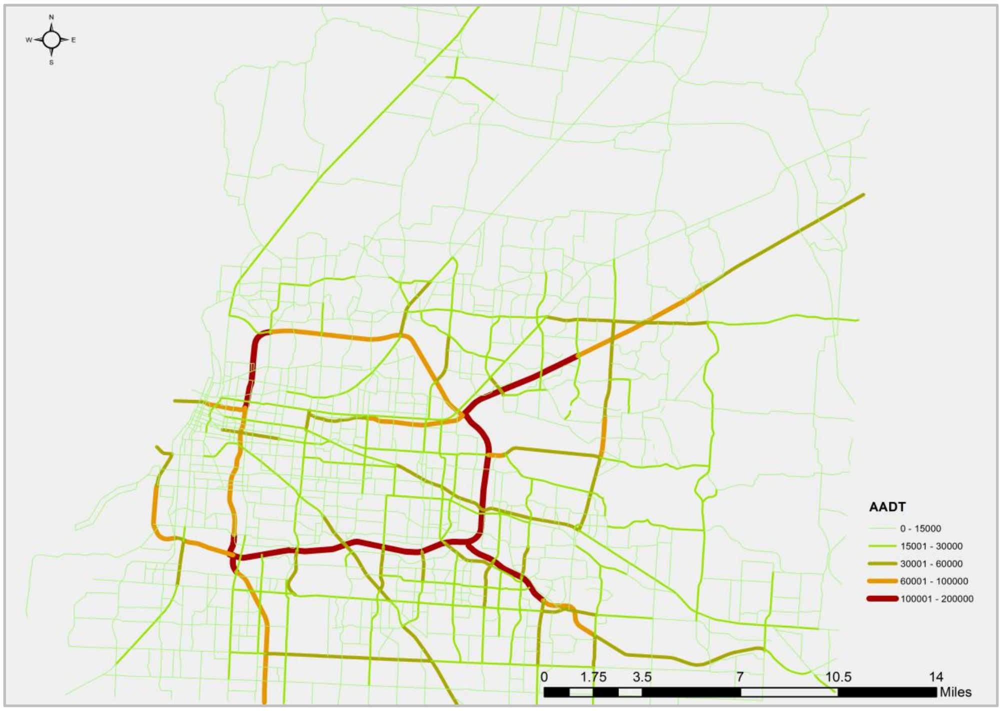
Freeway and interstates experience significantly higher Annual Average Daily Traffic (AADT) volumes when compared to major arterials. Fig. 3-12(a) shows AADT for the case study roadway network with moderate (orange) to high (red) volumes observed on freeway, interstates and arterials. Fig 3-12(b) presents SCs identified in the study area. Results shown in Fig. 3-12(a) and (b) indicate that several roadway facilities with moderate AADT experienced large number of SCs while a large portion of interstates encountered small number of SCs. It is found that moderately congested major arterials and freeway segments (as opposed to heavily congested segments) experienced very high occurrence of SCs. A possible explanation could be lower speeds and higher alertness of drivers for highly congested roadways; whereas on facilities with moderate congestion, higher speeds, and lower alertness increase the probability of PC and also the induced effect of a SC. These findings are similar to previous studies in the literature (Dixit et al., 2011; Schefer and Rietveld, 1994). The study by Dixit et al. (2011) reveals that roadway segments with lower free flow travel time (higher speeds) result in more severe crashes and hence, are more likely to induce a SC. Schefer and Rietveld (1994) investigate the relationship between congestion and crashes. The study finds that the lower speeds which are caused by congestion would lead to lower frequency of fatal crashes. As a result, there exists a parabolic relationship between density and severity of crashes on highways. When density is higher speed is expected to be lower leading to less severity and frequency of primary crashes. The lesser frequency of primary crashes can be attributed to alertness of the drivers at lower speed near jam density.

3.5.7 SCs Hotspots Map

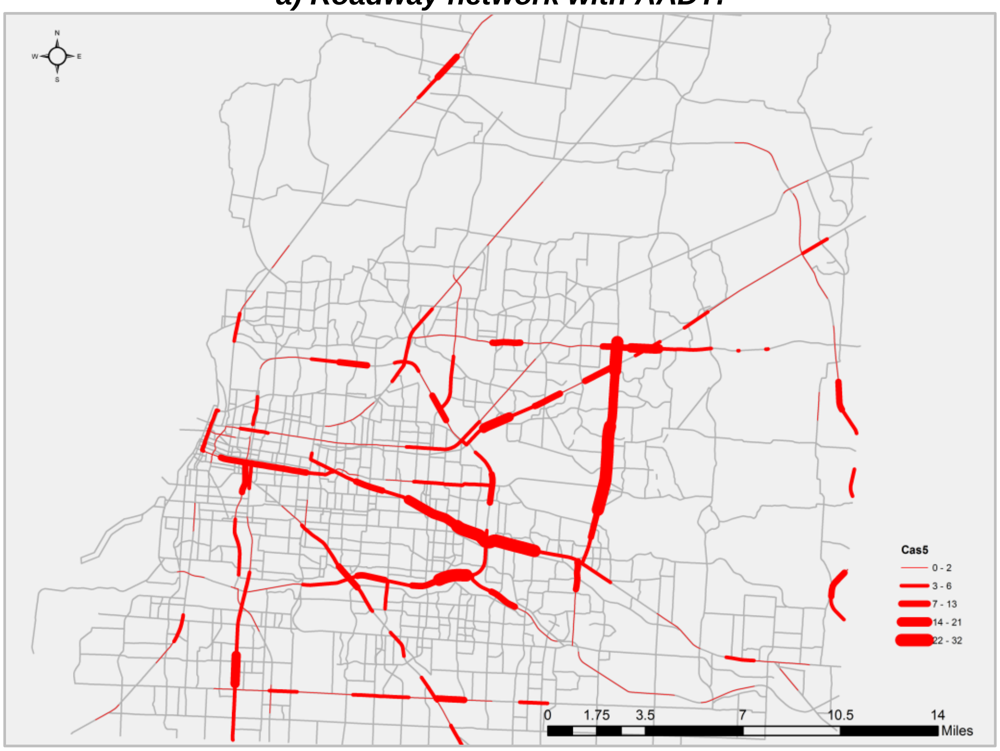
SCs hotspots map can be a useful visualization tool for various agencies and can assist in faster identification of problematic facilities as well as dissemination of results and possible remedial recommendations. Using results from the dynamic approach, hotspots maps, shown in Fig. 3-13, were developed for the case study network. For example, hotspot map for Case-5 (Fig. 3-13(e)) shows that the highest SC density occurs on two major arterials: Poplar Avenue (between *Perkins* and *Kirby Parkway*) and also on Germantown Parkway (between *Walnut Grove* and *Highway 64*). Another arterial with high number of identified SCs are a section of Highland St between *Southern Ave*. and *Poplar Ave.* Though traffic volume on those arterials is significantly less than on what usually observed on freeways, flow characteristics along with other primary contributing factors (e.g. geometric design and traffic operations) may have led to the high frequency of SCs. There are also some prominent hotspots on freeways, covering a relatively smaller region. For example, exit 16 and 17 of I-240E (toward *SR385* and *Mt. Moriah St.*) and at exit 10A of I-40W.

3.5.8 Validation

For any model or methodology, the ability to depict the real-life scenario accurately is important. From the literature review, no past studies have validated their methodology by comparing with observed number of SCs. In this study, the process of validation is carried out using the secondary incidents data from HELP database provided by TMC. The HELP trucks operate on assigned routes which are restricted to the core areas of each city, so that the operators can respond quickly to incidents that have the most impact on the total freeway system. In Shelby County, HELP truck serves 10 mile radius centering the TMC office at Memphis and mostly freeways. As a result any secondary incidents occurring on arterials are not reported in the HELP database. In the year 2011-2012, 16 incidents were identified as secondary. The proposed methodology using the dynamic approach was able to validate 13 of them as shown in Fig 3-14. To validate using a larger sample data, secondary incidents occurring in three other regions is also used.

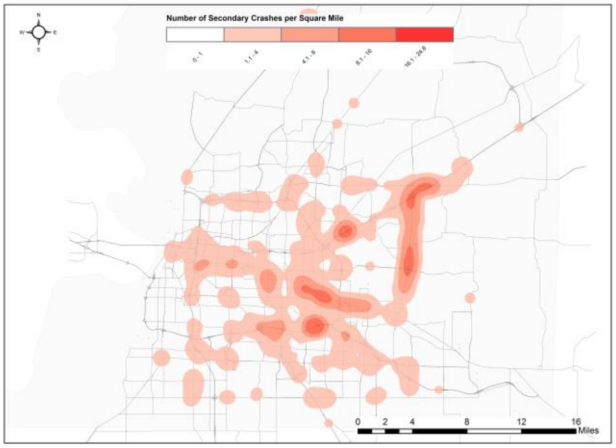


a) Roadway network with AADT.

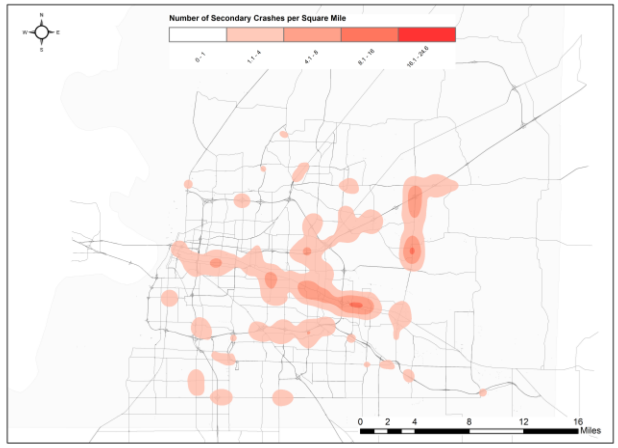


b) Identified SCs (Case-5).

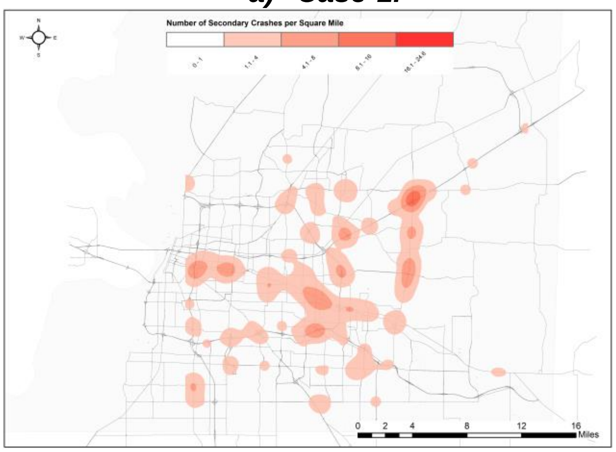
Figure 3-12 Comparing AADT and SCs.



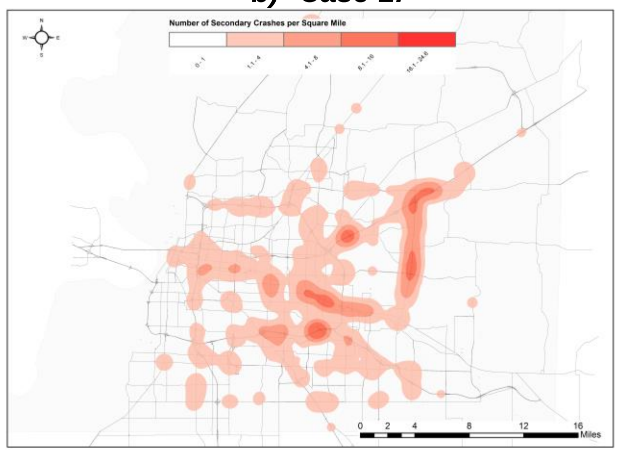
a) Case-1.



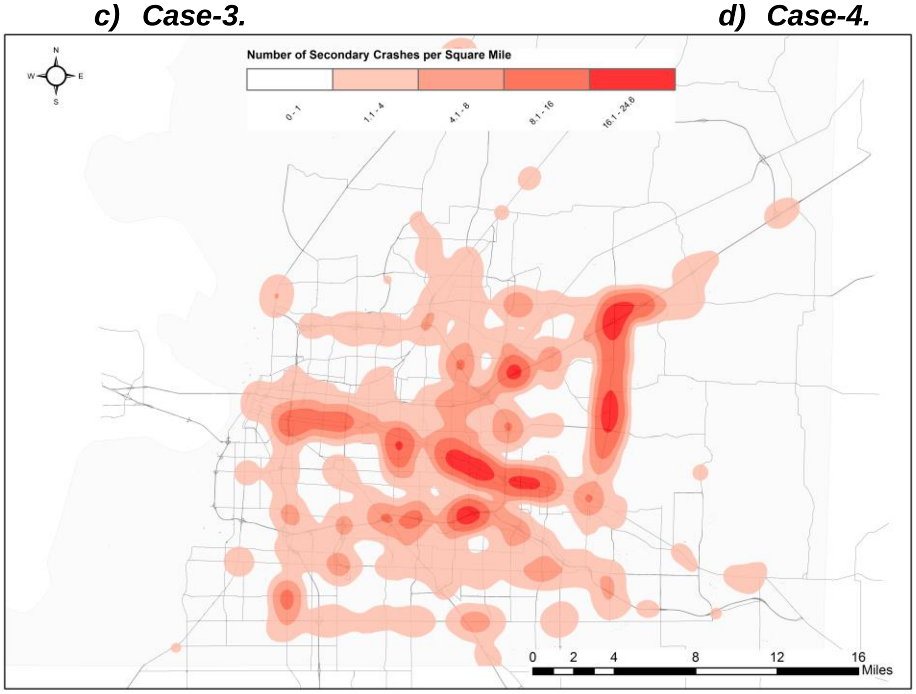
b) Case-2.



c) Case-3.

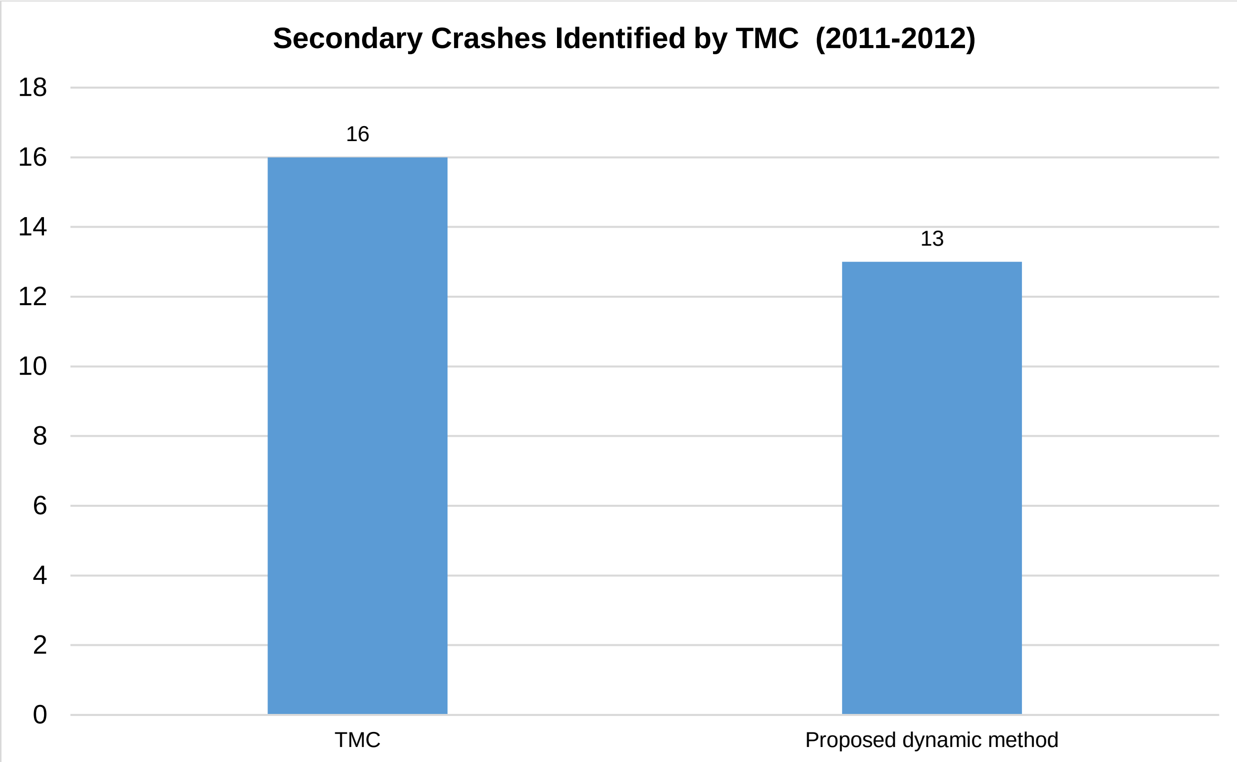


d) Case-4.



e) Case-5.

Figure 3-13 SCs hotspot map in Shelby County.



Note: Some SCs identified by TMC are excluded because of discrepancies in the data.

Figure 3-14 Validation using HELP database (2011-2012).

3.6 Discussions

For identification of SCs, static and dynamic approaches are used and results are compared with each other. Since static approach uses a predefined set of spatiotemporal thresholds, multiple combinations of spatio-temporal thresholds are used and compared to determine sensitivity. On the other hand, the dynamic approach used the shockwave principles to estimate the impact area of a PC and any crash occurring within that impact area is identified as a SC. The analysis of the results revealed the following:

- When compared to dynamic approach, the static approach consistently under and overestimated SC frequencies for small and large spatio-temporal threshold respectively. This phenomenon is expected as most SCs have a high probability of occurrence within the 30-60min and 0.5-1mile and a low probability of occurrence within the 300min and 5miles threshold. This relationship between static and dynamic approach is observed for both freeways and arterials.
- The characteristics of a facility type play a crucial role in inducing SCs. It is observed that facilities with moderate AADT (such as arterials) are quite likely to

encounter large number of SCs because of higher alertness of drivers on congested and low-speed roadways.

- SCs are more predominant during AM and PM peak hours, while arterials have a much prominent PM peak. Both peak periods account for up to 59% of the total number of identified SCs in the upstream of PC.

Based on the SC density a hotspot map is generated for the study area that shows the locations which are more likely to be encountering secondary crashes. These locations of the hotpots are of great importance to TIM agencies because investigating those locations to a great deal would reveal the primary contributing factors and also the strategies need to be undertaken in order to mitigate such incidents.

CHAPTER 4: DEVELOPMENT OF CRASH PREDICTION MODEL

4.1 Motivation

Occurrence of SCs significantly depends on characteristics of the PC, and associated traffic, highway geometry and environmental characteristics. In order to reduce SCs , it is important to identify the primary contributing factors associated with occurrence of a SC. The purpose of developing several crash prediction models is to determine those factors and also to estimate likelihood of SC occurrence given the characteristics of the PC. Prediction models can be really helpful for planners as well as TIM agencies to reduce the occurrence of SC by taking appropriate measures. If TIM agencies can obtain primary incident characteristics in real-time and quickly undertake necessary actions, they can reduce the impact of SCs in terms of congestion and safety.

4.2 Data

To develop different models, this study examined 74,806 primary incidents and their characteristics between years 2010 and 2012. Out of 74,806 incidents, 506 resulted in a SC and 22 induced multiple SCs. A map showing the location of the PC, SC and tertiary crash (crash induced by a SC) is shown in Fig 4-1.

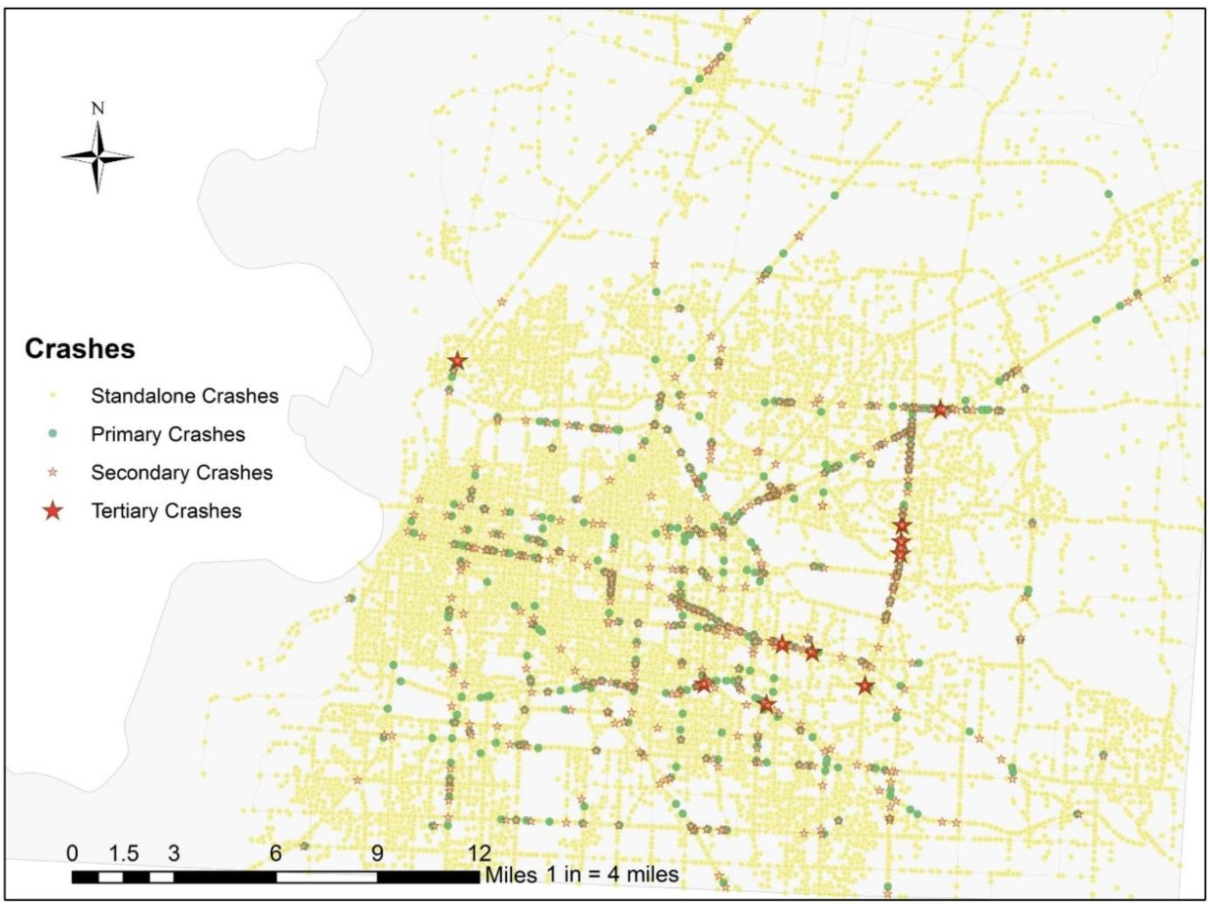


Figure 4-1 Map showing the locations of PC, SC and tertiary crash.

The potential independent variables considered for models are listed along with a brief description in Table 4-1. The descriptive statistics of these variables is presented below in Table 4-2.

Table 4-1 Potential independent and dependent variables.

Variable Description	***Type***	***Categories (if Applicable)***
Time of day when the PC occurred	Categorical	1-Peak hours (6-9am, 4-7pm) 0- Off-peak hours
Number of people injured in the PC	Continuous	
Number of people killed in the PC	Continuous	
Number of vehicles involved in the crash	Continuous	
Secondary crash occurrence (dependent variable)	Categorical	0-The PC did not result in a SC 1- The PC resulted in a SC
Severity of the PC	Categorical	1-Fatal 2-Incapacitating injury 3-Non-incapacitating injury 4-Property damage only (PDO)
Manner of collision	Categorical	0-No Collision 1-Rear-end 2-Head-on 3-Angle 4-Sideswipe 5-Others
Lighting condition	Categorical	1- Daylight or well-lighted 0- Not lighted/dawn/dusk
Weather condition	Categorical	1- Clear/good weather 0- Bad weather (snow, rain, sleet, fog etc.)
Average vehicle speed in the upstream of PC	Continuous	
Average flow of vehicles in the upstream of PC	Continuous	
Functional class of the roadway	Categorical	1- Freeway 0- Arterial
Annual Average Daily Traffic (AADT) of the roadway	Continuous	
$\log_{10}$(AADT)	Continuous	
Passenger car and pickups as a % of AADT	Continuous	
Single-Unit truck as a % of AADT	Continuous	
Multi-Unit truck as a % of AADT	Continuous	

It should be noted that some of the above factors are correlated with each other. Thus, only some of the correlated factors are expected to be selected by the model. The significance of these factors is determined as part of the model development as described in the following section.

Table 4-2 Descriptive statistics.

Variables	*Min*	*Max*	*Median*	*Mean*	*Standard Deviation*
Peak hour	0	1	0	0.390	0.488
Number of people killed	0	2	0	0.004	0.068
Severity of crash	1	4	4	3.828	0.618
Number of people injured	0	11	0	0.358	0.820
Number of vehicles involved	1	6	2	1.958	0.512
Incident type	0	5	1	1.894	1.405
Lighting indicator	0	1	1	0.928	0.259
Weather indicator	0	1	1	0.778	0.416
Avg. Speed of upstream traffic	0	85	36	38.814	12.913
Upstream flow	0	2040	317.96	382.508	297.535
Functional class	0	1	0	0.164	0.370
AADT	580	164150	26450	37832.351	35896.243
log (AADT)	2.76	5.21	4.42	4.43	0.36
Passenger car/pickups (% of AADT)	49	100	94	92.412	6.185
Single-unit truck (% of AADT)	0	19	3	3.193	1.774
Multi-unit truck (% of AADT)	0	41	3	4.394	5.347
Number of SCs	0	2	0	0.122	0.342
Number of SCs (Binary)	0	1	0	0.117	0.322

4.3 Model Estimation

The process of model estimation is started by taking all the explanatory variables (some with multiple categories) and then identifying the variables that have statistically significant relation with SC occurrence.

4.3.1 Linear Probability Model

Regression analysis methods, especially linear regression models, have increasingly been used in transportation research. Linear probability model (LPM) is very similar to Ordinary Least Square (OLS) regression with the assumption that dependent variable can only have two outcomes (0 or 1) and the independent variables are distributed following a normal distribution. To develop LPM, we needed a binary response variable and hence we created a dependent variable which had only two outcomes: A PC resulted in at least one SC and a PC didn't result in any SC. Table 4-3 shows the result obtained:

Table 4-3 Linear probability model.

Variable	*Coefficient*	*P-Value*	*VIF*
Constant	-0.03618	0.000	-
Number of vehicles involved	0.00475	0.000	2.10
Avg. speed of upstream traffic	-0.00008	0.075	2.86
Upstream flow	0.00003	0.000	1.53
log (AADT)	0.00719	0.000	2.06
Incident type: No collision	0.00558	0.000	2.24
Rear end	0.00310	0.000	1.15
Weather indicator	-0.00507	0.000	1.01
Functional class	-0.00735	0.000	4.00
No. of Observations (*N*): 74,806			
R^2 = 0.80%			

It can be observed that *number of vehicles involved, average speed of upstream traffic, upstream flow, log(AADT), No collision, weather* and *functional class* are found to be significant at 10% level of significance (α=0.10). The result shows that increase in number of total vehicle involved in PC increases the likelihood of SC occurrence. Also, a PC occurring on a roadway with high AADT and flow is more likely to induce a SC. Average upstream speed suggested a lower likelihood of a SC occurrence presenting the fact that the nature of PC is as such it does not have any significant impact on traffic flow characteristics. In terms of incident type, “No collision” and “Rear-end” are more likely to cause a SC in the upstream. In addition, a PC occurring on a freeway in a good weather condition is less likely to result in a SC. According to this model, Severity of crashes, traffic composition and lighting condition play no significant role. Another thing to be noted is that none of the explanatory variables are correlated with each other, as shown the variance inflation factor (VIF) which is less than the commonly used cut-off value of 10 (Kutner et al., 2004).

4.3.2 Binary Logit Model

A logistic regression model is developed to analyze the relationships between the primary incident characteristics and the possibility of secondary crash occurrence. In this model, we start the choice set with a binary response variable: *at least* *one SC* and *No SC.* The general form of incident occurrence probability in a logistic model is as follows (Ben-Akiva and Lerman, 1985):

$$P(y_i = 1|x_i) = p_i = \frac{e^{\alpha+\beta x_i}}{1 + e^{\alpha+\beta x_i}}$$

Where, where p_i is the probability that an instance *i* will occur, *α* is the constant, *β* is the vector of coefficients for independent variables, and x_i is the vector of independent variables.

The binary logit models developed is shown in Table 4-4. It should be noted that the reference category in this model is "*No SC"*. The coefficients of the variables obtained in the model are mostly as expected. It can be noticed that increased number of vehicles involved in crashes, relatively high AADT, increased upstream flow lead to higher likelihood of SC occurrence. On the other hand, good weather condition and PC occurring on a freeway result in less likelihood of SC occurrence. It is also observed that if the PC type is a rear-end collision or if the primary incident is not a collision with another vehicle then SCs are more likely to occur compared to other types of collision (e.g. angle, sideswipe, head-on etc.).

Table 4-4 Binary logit model.

Variable	***Coefficient***	***P-Value***
Constant	-14.59	-
Number of vehicles involved	0.3779	0.000
Upstream flow	0.0020	0.000
log (AADT)	1.8970	0.000
Weather indicator	-0.6648	0.000
Functional Class	-1.6120	0.000
Incident type: No collision	0.3910	0.038
Rear end	0.4331	0.000

No. of Observations (*N)*: 74,806
Final Log likelihood = -2898.375
Pseudo R^2 (ρ^2) = 0.077

4.3.3 Multinomial Logit Model

There are two types of multinomial logistic regression models heavily used: the Multinomial Logit (MNL) and the Conditional Logit (CL) model. MNL and CL have the same functional form although there are some fundamental differences between them. A MNL model can be described by a situation where the main influences on the choice outcome are the characteristics of the observations (for example: individual crash). This is in contrast to the CL model where the primary influences are the attributes of alternatives that vary for each alternative (alternate specific parameters). In this study, attributes of the crashes do not vary across outcomes. For example, the time of a crash does not vary across the possibility of inducing no, one and two or more SCs. But, it does among observations in the dataset as a whole. In practice, when estimating the model the model coefficients of the reference group are set to zero. Since there are three choices: *no SC* (reference category), *one SC* and *two SC* (referred as tertiary crash), only two distinct sets of parameters can be identified and estimated. Table 4-5 lists the MNL estimation results. The coefficients of the estimated model can be interpreted as follows. A positive significant coefficient on a variable indicates that the variable is associated with a higher probability of being in that group choice relative to the reference group and vice versa. When compared to binary logit model, similar inference can be made using the coefficients of the variables in this model.

Table 4-5 Multinomial Logit Model.

Variable	Multinomial Logit			
	Secondary Crash		***Tertiary Crash***	
	Coefficient	*P-value*	*Coefficient*	*P-Value*
Constant	-6.602	0.000	-11.637	0.000
AADT	1.023e-05	0.000	3.7639e-05	0.007
Functional Class	-1.362	0.000	-4.693	0.003
Number of vehicles involved	0.373	0.000	0.630	0.000
Upstream flow	2.087e-03	0.000	2.391e-03	0.000
Incident Type: Rear end	0.466	0.000	0.896	0.087
No collision	0.330	0.079	-	-
Weather indicator	-0.669	0.000	-	-

No. of Observation = 74,806
Final Log likelihood = - 3011.30
Pseudo R^2 (ρ^2) = 0.069
Likelihood ratio test : chi^2 = 443.53 (p-value=0)

The result obtained for MNL can be interpreted as follows. If PC occurred have the following characteristics:

- AADT:149,840
- Facility Type: Freeway
- Total number of vehicles involved:4
- Type of crash: Rear-end
- Weather condition: Rainy
- Upstream flow at the time of PC: 1,924

The utility expression of secondary crash can be evaluated to be:
-6.602 + (1.023e-05*149840)+(-1.362*1)+(0.373*4)+(2.087e-03*1924)+ (0.466*1)+ (0.330*0) + (-0.669*0) = -0.4578

Similarly, the utility function of tertiary crash is calculated to be:
-11.637 + (3.764e-05*149840)+(-4.693*1)+(0.630*4)+(2.391e-03*1924)+(0.896*1) = -2.6737
The probability of an incident occurring in MNL can be expressed as follows (Ben-Akiva and Lerman, 1985):

$$P(m) = \frac{e^{V_m}}{\sum_{k=1}^{M} e^{V_k}}$$

Where,

$P\ (m)$ = Probability of event '*m'* occurring

V_j = utility of event '*m*'

j = Set of choice, *j*=1,…*M*

Using this expression,

$$P(Secondary\ Crash) = \frac{e^{-0.4578}}{1+\ e^{-0.4578} +\ e^{-2.6737}} = 0.3718\ or\ 37.18\%$$

$$P(Tertiary\ Crash) = \frac{e^{-0.4578}}{1+\ e^{-0.4578} +\ e^{-2.6737}} = 0.0405\ or\ 4.05\%$$

$$P(Primary\ Crash) = \ 1 - \ (0.3718 + 0.0405) = \ 0.5877\ or\ 58.77\%$$

4.4 Discussions

The results from LPM and binary logit models indicate the following:

- Increased number of vehicle involved in PC leads to higher likelihood of SC occurrence because, larger the number larger the clearance time, duration and associated queue length.
- Roadway with high AADT is more likely to encounter SCs as it has higher probability to encounter PC and which turn can induce SC. But it should be noted that the results also indicate that Freeway is less likely to encounter SCs which can be explained by the fact that in Shelby county arterials cover much larger number of lane miles compared to freeways and interstates. Also, freeways have more number of lanes meaning if a crash happens entire roadway is less likely to be blocked which is not the case in Arterials. If the entire roadway is blocked it will lead to larger queue,
- If there is a high flow of traffic in the upstream at the time of PC, it is more likely to induce SCs. Though *time of day* was not found to be directly significant, traffic flow is directly related with *time of day* as peak hours encounter significant increase in flow. Section 5.4 explains discusses SC distribution during the hours of a day.
- Likelihood of SC occurrence also depends on the primary incident type. If the primary incident is a rear-end collision it is more likely to induce SCs. But severity of crashes has not been found to have significant effect on the SC occurrence.

It should be noted that the dependent variable is unbalanced as very small fraction (0.7%) of total crashes are secondary making it very difficult to estimate good models. Nonetheless, models are overall statistically significant and have reasonable goodness of fit.

For the crash prediction models, secondary crashes identified only for case-1 scenario is used because, SCs are more likely to occur in the upstream direction of a PC. Future research can be conducted to obtain the prediction models for other case scenarios. Also, to better understand the characteristics and their relation with SC, the endogeneity (interdependence between PC duration and SC occurrence) should be taken into account while developing models. Various other classical econometric model structures such as nested logit, cross nested logit, mixed logit and Bayesian econometric models can be analyzed in the future.

CHAPTER 5: SECONDARY CRASH IDENTIFICATION TOOL

In order to facilitate the process of analyzing crash data, a stand-alone tool was developed (Fig 5-1). The tool identifies SCs and provides a descriptive statistics for all the different cases (refer to five cases presented in Chapter 3) using an archived database. The tool can be downloaded using this link. This tool incorporates the both static and dynamic approach as mentioned in the proposed methodology. It also has the provision to allow the user set multiple criterions. Minimum requirement of successful installation requires the computer to have following minimum requirements:

- Windows Vista SP2, Windows 7 SP1, Windows 8, Windows Server 2008 SP2 Windows Server 2008 R2 SP1 and Windows Server 2012.
- .Net Framework 4.5.1.
- 20 MB Free disk space

The main window of the software tool is divided into two sections: 1) Left side: The set of input controls and a trigger to execute, and 2) Right side: The results will be produced in the right had portion of the software screen (see Fig 5-1). Description of each section is provided in more detail in the following quick steps.

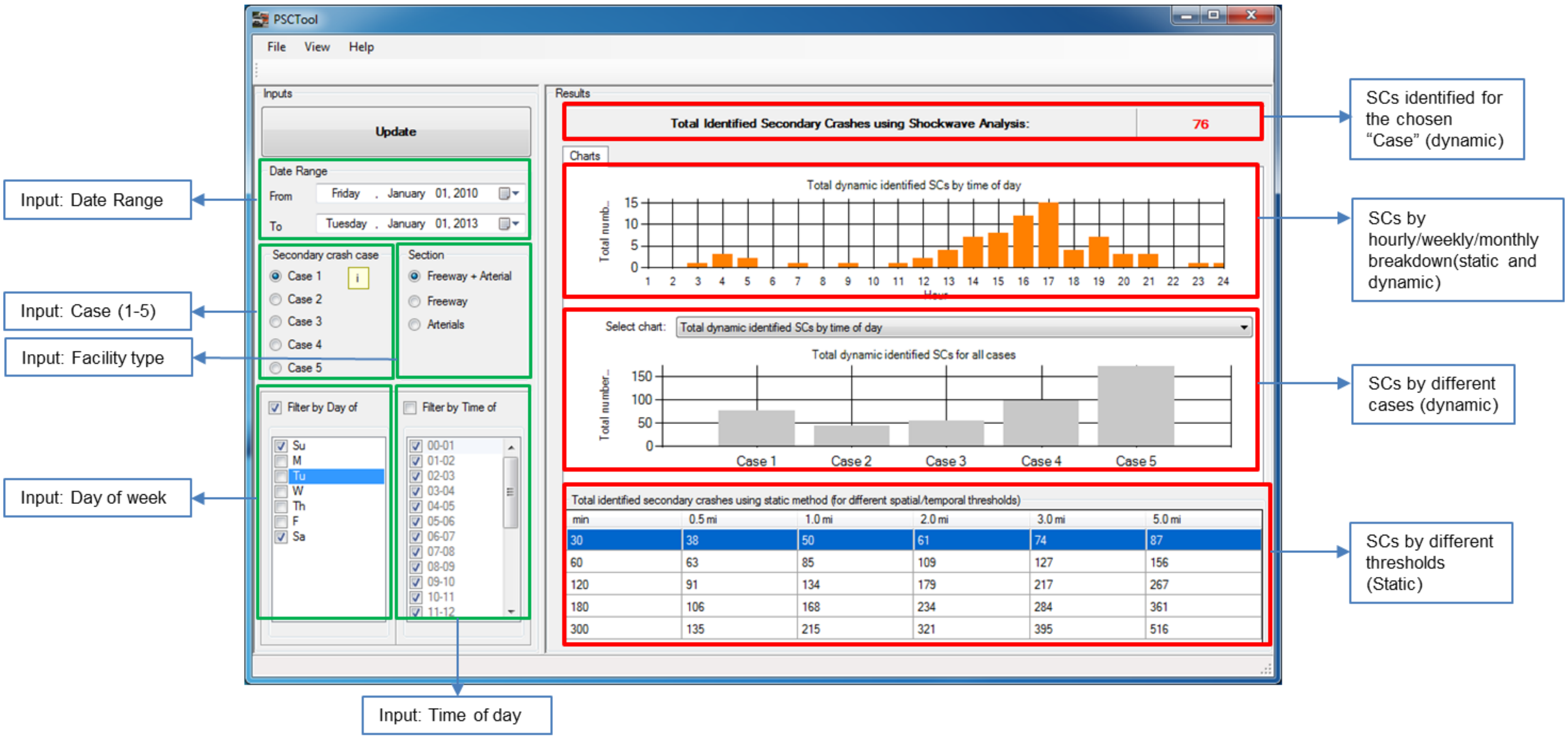


Figure 5-1 Overview of the secondary crash identification tool.

The steps to use the tool are summarized as follows:

Step 1: Specify the date range for the period of analysis. This can be performed by selecting the beginning date, and the end date using the dropdown calendar as shown in Fig 5-2.

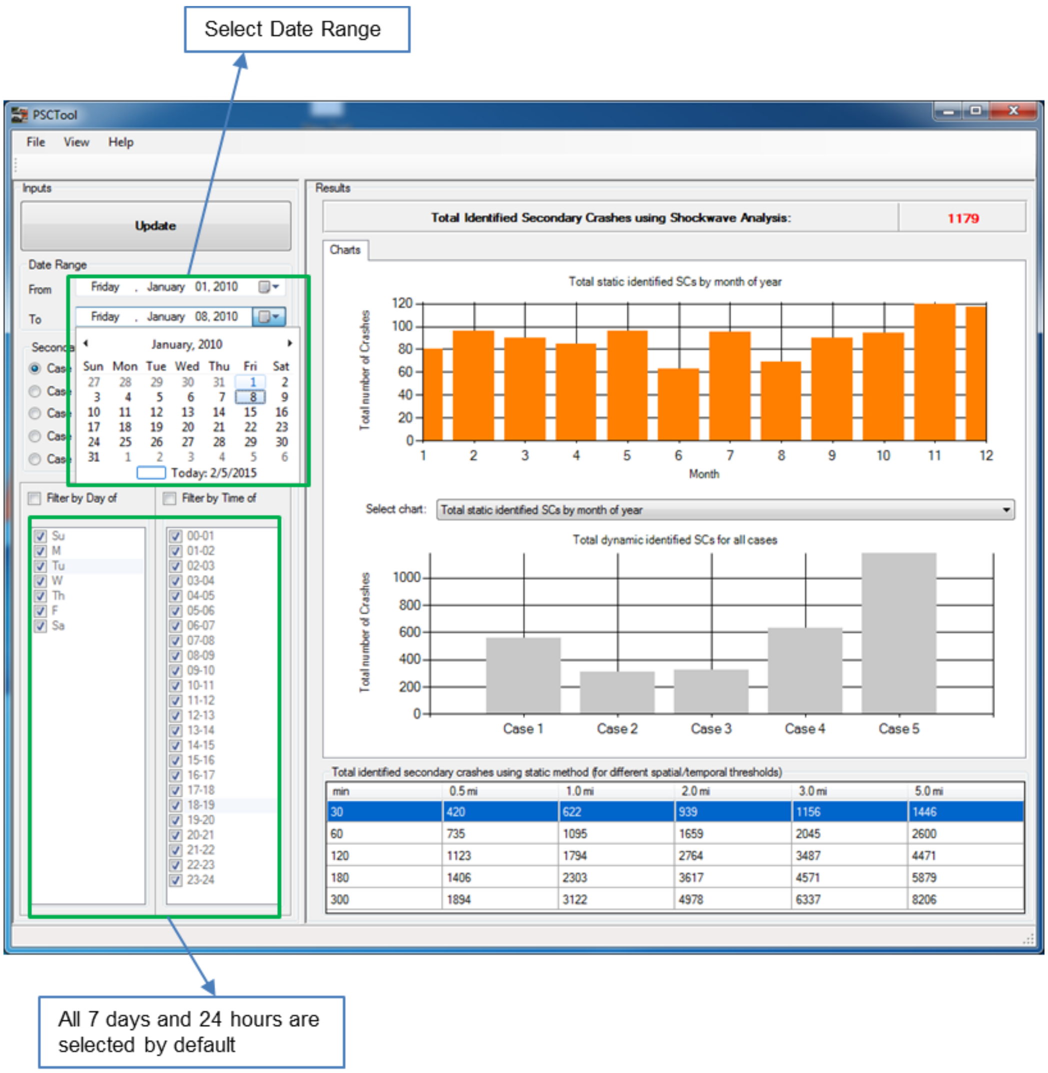


Figure 5-2 Specifying period of analysis.

Step 2: Select a particular case for analysis. Cases 1-5 are defined in the methodology section and the tool also provides a graphical representation of all the cases as shown in Fig 5-3.

Step 3: Specify the facility type by selecting Freeway and/or arterials.

Step 4: Select the days of week. If the checkbox 'Filter by Day of Week' is not checked, the algorithm will consider all seven days of the week starting from Sunday.

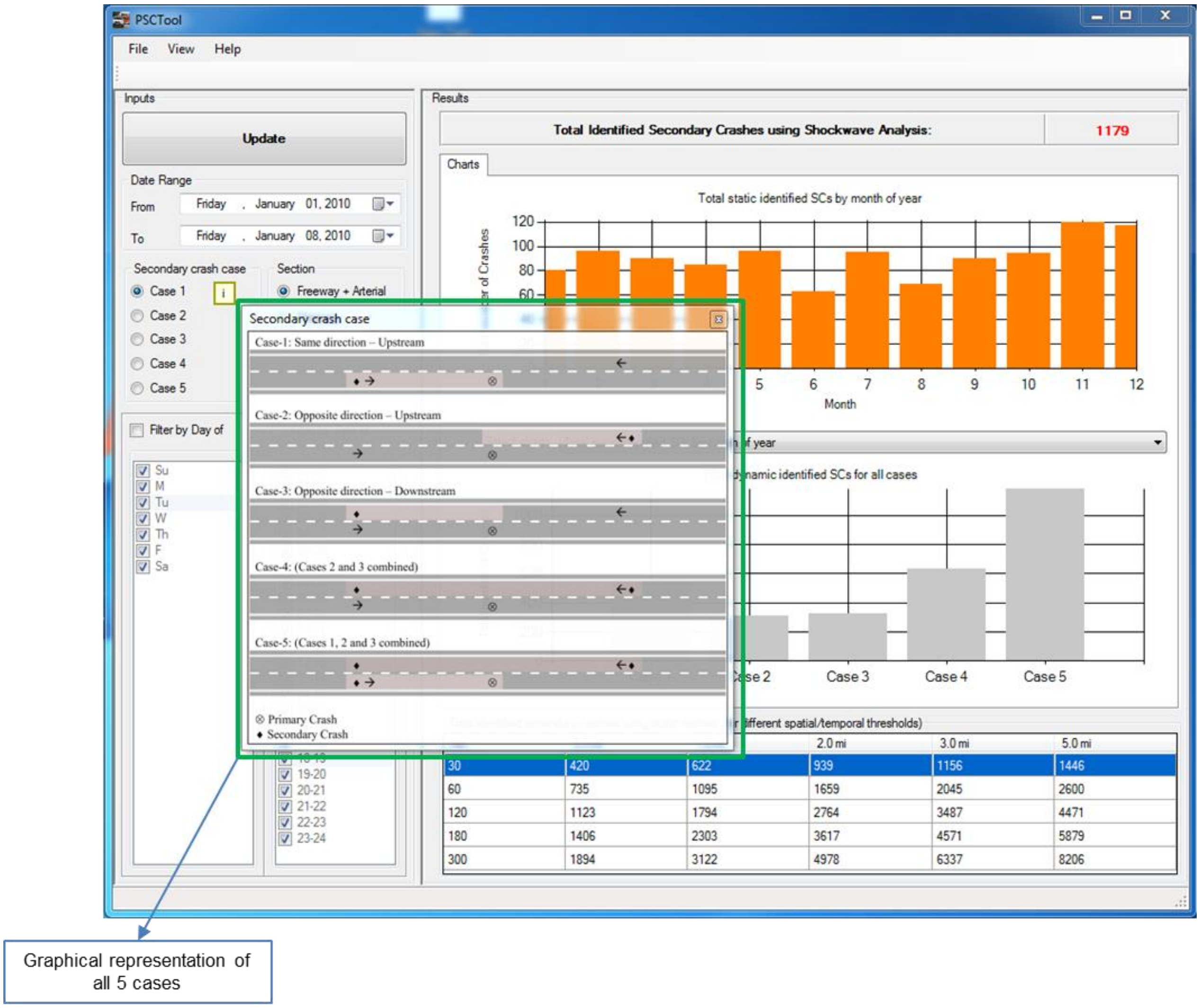


Figure 5-3 Graphical representation of Cases 1-5.

Step 5: Select the time of day. If only specific hours of day are to be analyzed, it can be done by this input section. If the checkbox 'Filter by Time of Day' is not checked, the algorithm will consider 24 hours by default.

Step 6: Click "Update". When the criterions are specified, click the "Update" button to perform the analysis.

Step 7: The results will be provided in the right side of the tool. Two charts are provided in the output section. The top-right section provides a breakdown of total crashes and SCs by time of day, by day of week, and by month of year using static or dynamic approach. The top chart can be changed by the drop down control on the bottom of the chart as shown in Fig 5-4. The bottom chart shows the total number of identified secondary crashes using dynamic method, separated by each case. The table in the bottom-right section provides total number of SCs identified using static approach for different spatiotemporal threshold.

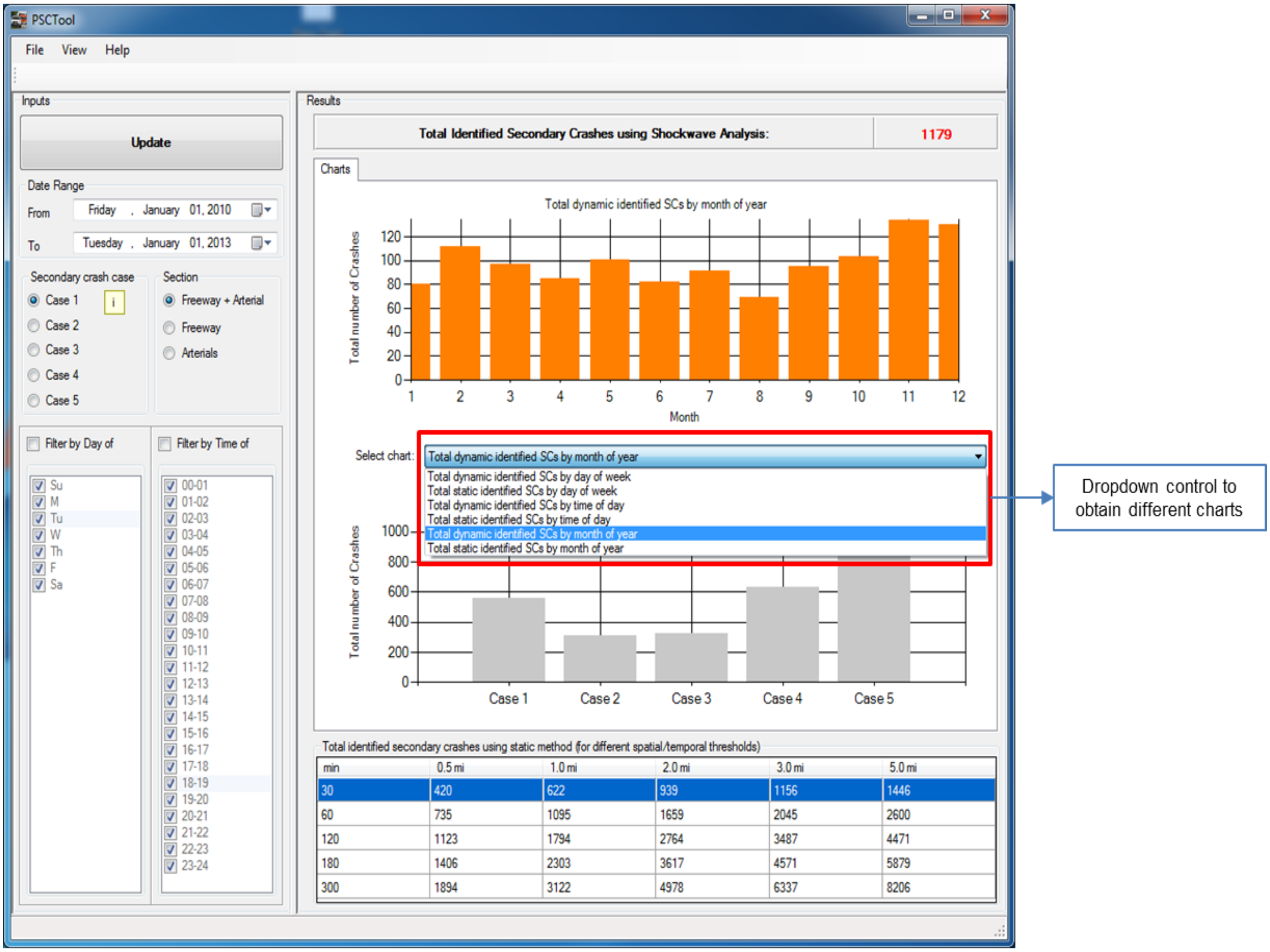


Figure 5-4 Producing different charts.

CHAPTER 6: CONCLUSIONS

This study proposed a Secondary Crash Identification Algorithm (SCIA) to for identification of SCs on large scale networks. SCIA is further categorized by static and a dynamic approach. Majority of past studies have proposed static approach and very limited suggested dynamic approach to identify SCs but, to date no robust methodology had been proposed that can identify SCs with considerable accuracy on large networks within an acceptable computation time. Most of the past studies were conducted on short segments of freeways in a small regional scale and the dynamic approach was used only where detailed data were available. High resolution data to capture the dynamic variation in traffic flow characteristics as a result of a primary incident are rarely available for arterials. These data limitations restricted the application of dynamic approaches to SCs identification on arterials.

For the static approach this study proposed five cases in an effort to consider all the different location and directionality combinations available when identifying SCs. The spatial threshold was applied in the opposite direction to capture effects of ‘rubbernecking’ which causes congestion and reduction in capacity in the opposite direction of the PC and can induce SC on arterials and even on freeways. For the static approach different spatiotemporal thresholds were used to evaluate their effect on the numbers of identified SCs. Temporal thresholds of 30, 60, 120, 180 and 300 minutes were used along with spatial thresholds of 0.5, 1, 2, 3 and 5 miles. The dynamic approach was based on shockwave principles and impact area analysis. A crash was identified as secondary if it occurred within the impact area of the PC. The proposed methodology was implemented in Shelby County, TN where SCs were identified for two types of facilities: freeway and arterials to account for the different traffic conditions and data availability for each.

The analysis of the results revealed that the static approach consistently under and overestimated SC frequencies for small and large spatio-temporal threshold respectively. This phenomenon is expected as most SCs have a high probability of occurrence within the 30-60min and 0.5-1miles temporal and spatial threshold respectively and a low probability of occurrence within the 300min and 5miles temporal and spatial threshold respectively. It was observed that characteristics of a facility type and time of day play a crucial role in inducing SCs. Results also revealed that facilities with moderate AADT (such as arterials) are quite likely to encounter large number of SCs. To identify the locations where SCs are more likely to occur, a hotspot map was developed for the study area based on the density of SCs. The proposed methodology can identify SCs and network wide hotspots to assist transportation agencies in the decision making process to mitigate such incidents.

Further a set of prediction models were developed to determine the causality of various factors to occurrence of SC. The results from SC prediction models revealed that some of the factors that lead to SC occurrence are: increased number of vehicles involved in PC, relatively high AADT, large flow of vehicle in the upstream of PC, PC occurring during bad weather condition (e.g. rain fog, snow, sleet etc.) and some particular type of primary incident such as “rear-end”. It was also observed that freeway is less likely to encounter SCs which can be explained by the fact that in Shelby county arterials cover much larger number of lane miles compared to freeways and interstates. The roadway network of Shelby County consists of approximately 285 miles of interstate/freeway facility but encountered 21.10% of SCs during 2010-12, whereas 56.40% of SCs occurred on 972 miles of major arterials. Also, freeways have more number of lanes meaning if a crash occurs, entire roadway is less likely to be blocked which is often not the case for arterials. If the entire roadway is blocked, it induces larger queue and hence more likely to lead to SCs.

For future research, SCIA can be validated using the HELP database from TMC. It will help to determine the accuracy of the identification process and try to eliminate type I and II error in the methodology. Marginal effects and elasticities of the independent variables should also be considered to identify each of their individual effect on SC occurrence. Also, to better understand the crash characteristics and their relation with SC, the endogeneity (interdependence between PC duration and SC occurrence) should be taken into account while developing models, using a 2SLS method. Further, other classical econometric and Bayesian models can be developed in the future for predicting SC occurrence.

ACKNOWLEDGEMENT

This research was partially funded by WisDOT, TDOT, and USDOT. Also, computational facilities at Intermodal Freight Transportation Institute (IFTI) at University of Memphis are greatly acknowledged. Further acknowledgement includes the support of Pooled Fund Program: TPF-5(274) http://www.pooledfund.org/Details/Study/507. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the above-mentioned agencies.

REFERENCES

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete choice analysis: theory and application to travel demand. MIT press.

Chang, G.-L., Rochon, S., 2003. Performance Evaluation of CHART–Coordinated Highways Action Response Team–Year 2002. Final Rep. Md. State Highw. Adm.

Chimba, D., Kutela, B., Ogletree, G., Horne, F., Tugwell, M., 2014. Impact of Abandoned and Disabled Vehicles on Freeway Incident Duration. J. Transp. Eng. 140, 04013013. doi:10.1061/(ASCE)TE.1943-5436.0000635

Chung, Y., 2010. Development of an accident duration prediction model on the Korean Freeway Systems. Accid. Anal. Prev. 42, 282–289. doi:10.1016/j.aap.2009.08.005

Dixit, V.V., Pande, A., Abdel-Aty, M., Das, A., Radwan, E., 2011. Quality of traffic flow on urban arterial streets and its relationship with safety. Accid. Anal. Prev. 43, 1610–1616. doi:10.1016/j.aap.2011.01.006

Dunn, W.M., Latoski, S.P., 2003. SAFE AND QUICK CLEARANCE OF TRAFFIC INCIDENTS. NCHRP Synth. Highw. Pract.

Hagen, L.T., 2005. Best Practices for Traffic Incident Management in Florida.

Hirunyanitiwattana, W., Mattingly, S.P., 2006. Identifying Secondary Crash Characteristics for California Highway System. Presented at the Transportation Research Board 85th Annual Meeting.

Imprialou, M.-I.M., Orfanou, F.P., Vlahogianni, E.I., Karlaftis, M.G., 2014. Methods for Defining Spatiotemporal Influence Areas and Secondary Incident Detection in Freeways. J. Transp. Eng. 140, 70–80. doi:10.1061/(ASCE)TE.1943-5436.0000612

Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. Accid. Anal. Prev. 23, 239–255. doi:10.1016/0001-4575(91)90003-N

Karlaftis, M.G., Latoski, S.P., Richards, N.J., Sinha, K.C., 1999. ITS impacts on safety and traffic management: an investigation of secondary crash causes. J. Intell. Transp. Syst. 5, 39–52.

Khattak, A.J., Wang, X., Zhang, H., 2010. Spatial Analysis and Modeling of Traffic Incidents for Proactive Incident Management and Strategic Planning. Transp. Res. Rec. J. Transp. Res. Board 2178, 128–137. doi:10.3141/2178-14

Khattak, A.J., Wang, X., Zhang, H., Cetin, M., 2011. Primary and Secondary Incident Management: Predicting Durations in Real Time.

Khattak, A., Wang, X., Zhang, H., 2009. Are incident durations and secondary incidents interdependent? Transp. Res. Rec. J. Transp. Res. Board 2099, 39–49.

Kutner, M.H., Nachtsheim, C., Neter, J., 2004. Applied linear regression models. McGraw-Hill/Irwin.

Kwon, J., Mauch, M., Varaiya, P., 2006. Components of congestion: delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. Transp. Res. Rec. J. Transp. Res. Board 1959, 84–91.

Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves. II. A theory of traffic flow on long crowded roads. Proc. R. Soc. Lond. Ser. Math. Phys. Sci. 229, 317–345.

Margiotta, R., Dowling, R., Paracha, J., 2012. Analysis, Modeling, and Simulation for Traffic Incident Management Applications.

Moore, J.E., Giuliano, G., Cho, S., 2004. Secondary accident rates on Los Angeles freeways. J. Transp. Eng. 130, 280–285.

Owens, N., Armstrong, A., Sullivan, P., Mitchell, C., Newton, D., Brewster, R., Trego, T., 2010. Traffic incident management handbook.

Ozbay, K., Kachroo, P., 1999. Incident management in intelligent transportation systems.

Pigman, J.G., Green, E.R., Walton, J.R., 2011. Identification of Secondary Crashes and Recommended Countermeasures.

Raub, R.A., 1997a. Secondary crashes: An important component of roadway incident management. Transp. Q. 51.

Raub, R.A., 1997b. Occurrence of secondary crashes on urban arterial roadways. Transp. Res. Rec. J. Transp. Res. Board 1581, 53–58.

Raub, R., Schofer, J., 1997. Managing Incidents on Urban Arterial Roadways. Transp. Res. Rec. 1603, 12–19. doi:10.3141/1603-02

Saddi, R.R., 2009. Studying the impacts of primary incidents on freeways to identify secondary incidents.

Schefer, D., Rietveld, P., 1994. Congestion and Safety on Highways; Towards an Analytical Model. Presented at the Proceedings of The Third International Conference on Safety and the Environment in the 21st Century: Lessons From The Past Shaping The Future, November 7-10, 1994, Tel Aviv, Israel.

Skabardonis, A., Petty, K., Varaiya, P., Bertini, R., 1998. Evaluation of the freeway service patrol (FSP) in Los Angeles. Calif. Partn. Adv. Transit Highw. PATH.

Sun, C.C., Chilukuri, V., 2010. Dynamic incident progression curve for classifying secondary traffic crashes. J. Transp. Eng. 136, 1153–1158.

Sun, C., Chilukuri, V., 2007. Secondary accident data fusion for assessing long-term performance of transportation systems. Midwest Trasportation Consortium, c/o Iowa State University.

Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., Halkias, B.M., 2010. Freeway Operations, Spatiotemporal-Incident Characteristics, and Secondary-Crash Occurrence. Transp. Res. Rec. J. Transp. Res. Board 2178, 1–9.

Yang, H., Bartin, B., Ozbay, K., 2014a. Mining the Characteristics of Secondary Crashes on Highways. J. Transp. Eng. 140, 04013024. doi:10.1061/(ASCE)TE.1943-5436.0000646

Yang, H., Ozbay, K., Morgul, E.F., Bartin, B., Xie, K., 2014b. Development of an On-line Scalable Approach for Identifying Secondary Crashes. Presented at the Transportation Research Board 93rd Annual Meeting.

Zhan, C., Gan, A., Hadi, M., 2009. Identifying Secondary Crashes and Their Contributing Factors. Transp. Res. Rec. J. Transp. Res. Board 2102, 68–75. doi:10.3141/2102-09

Zhan, C., Shen, L., Hadi, M.A., Gan, A., 2008. Understanding the Characteristics of Secondary Crashes on Freeways. Presented at the Transportation Research Board 87th Annual Meeting.

Zhang, H., Khattak, A., 2010. What is the role of multiple secondary incidents in traffic operations? J. Transp. Eng. 136, 986–997.

Zhang, H., Khattak, A., 2011. Spatiotemporal patterns of primary and secondary incidents on urban freeways. Transp. Res. Rec. J. Transp. Res. Board 2229, 19–27.

Zheng, D., Chitturi, M.V., Bill, A.R., Noyce, D.A., 2014. Secondary Crash Identification on a Large-Scale Highway System. Presented at the Transportation Research Board 93rd Annual Meeting.

APPENDIX A: LITERATURE REVIEW SUMMARY

Literature	Temporal threshold	Spatial Threshold	Additional Comments	Statistical Model	Dependent Variable	Independent Variables/ Contributing Factors
Hirunyanitiwattana (2006)	Clearance time < 60 minutes	Queue length< 2 miles in the upstream	Excluded secondary crashes in the opposite direction			▪ Urban vs. rural ▪ Time of day ▪ Crash severity ▪ Road classification ▪ Collision Type ▪ Primary Collision Factor
Sun and Chilukuri (2010)	Duration of the primary incident	Dynamic Threshold (Incident Progression Curve)	Used media Traffic Reports From Traffic Management Centers.	Polynomial model	Queue dissipation	
Zhan et al. (2008)	Clearance time+15 minutes	< 2 miles in upstream				▪ Time/Duration ▪ No. of lanes ▪ Peak Period ▪ Rollover ▪ No. of vehicles involved
Zhan et al. (2009)	The queue dissipation time of the primary incident	Maximum possible queue Length upstream		Logistic Regression Analysis	Occurrence of Secondary Crashes	▪ Time/Duration ▪ Peak Period ▪ Speeding ▪ Rollover ▪ Environmental ▪ Lane Closure ▪ Injury/Severity ▪ Vehicle type ▪ Location ▪ Traffic Volume
Pigman et. al (2011)	<80 minutes	<6000 ft for Same route and <1000 ft for side street/ intersection crashes	Reviewed crash data records from Kentucky's CRASH database			▪ Incident characteristics ▪ Incident detection efficiency ▪ Distribution of detection sources

						▪ Incident response efficiency ▪ Effectiveness of incident traffic management
Khattak et. al (2009)	Duration of the primary incident	Segment length where primary incident occurred (Segment Based) And Queue Length caused by primary incident (can spill over to upstream segments for the same direction), using segment length for opposite direction		Probit model with endogenous regressors	Occurrence of Secondary Events (2006 Data)	▪ Detection source ▪ Incident type ▪ Closure time ▪ Response vehicles ▪ Lanes affected ▪ Time of day ▪ AADT ▪ # of vehicles involved ▪ Duration of primary incident
				Occurrence of Secondary Events (2005 Data)	Partial Proportional Odds Models	▪ Primary incident type ▪ Incident duration ▪ Truck Involved ▪ # of vehicles, ▪ Presence of Outstate vehicle in primary ▪ Lane blockage in primary (%) ▪ Segment length ▪ Presence of Curve ▪ AADT/(Lane*1000)
				Time Gaps	Heckman Selection Model	▪ Primary incident type ▪ Incident duration ▪ Truck Involved ▪ # of vehicles, ▪ Presence of Outstate vehicle in primary ▪ Lane blockage in primary (%) ▪ Segment length ▪ Presence of Curve ▪ AADT/(Lane*1000)

				Event Duration	Linear Regression (ordinary and truncated)	▪ Primary incident type ▪ Incident duration ▪ Truck Involved ▪ Number of vehicles ▪ Presence Outstate vehicle in primary ▪ Lane blockage in primary (%) ▪ Segment length ▪ Presence of Curve ▪ AADT/(Lane*1000) ▪ On ramp ▪ Service Patrol detected ▪ Response time of service patrol for primary ▪ Response time of service patrol for secondary ▪ Time gap ▪ Secondary Lane blockage (%)
				Frequency of secondary Incidents	▪ Poisson Negative Binomial ▪ Zero inflated Poisson*	▪ Roadway segment length ▪ Number of lanes ▪ Curve or not ▪ AADT ▪ Congestion level (AADT per lane) ▪ Number of on-ramps/off-ramps ▪ Truck volume ▪ Distance to shopping center/ school/tunnel
Imprialou et. al (2014)	Real Influencing Area (RIA) (dynamic approach to detect dissipation and	RIA (Dynamic Threshold)	Using detectors The speed and The difference between the speed			

	propagation of real influence areas)		at successive downstream detectors			
Sun and Chilukuri (2007)	Duration of the primary Incident	Incident Progression Curve		Polynomial (2nd, 3rd and 4th degree)	Incident Duration and Queue Length (IPC Curves)	▪ Severity ▪ # of Vehicles ▪ v/c ratio
Khattak et. al (2010)	Duration of the primary incident	Segment length where primary incident occurred (Segment Based) And Queue Length caused by primary incident (can spill over to upstream segments for the same direction), using segment length for opposite direction		Poisson, zero-inflated Poisson, and negative binomial regression models	Frequency of secondary incidents from a macroscopic level	▪ Roadway length ▪ Traffic volume ▪ Number of on-ramps curve level ▪ Number of lanes ▪ Congestion level ▪ Truck volumes ▪ Roadway location
Zhang and Khattak (2011)	Duration of the primary incident	Queue Length		ordinary least – square regression models	the time and distance gap of secondary incidents	Primary incident characteristics
Yang et. al (2013)	Speed Contour Plot (SCP), Binary Speed Contour Plot (BSCP)	Speed Contour Plot (SCP), Binary Speed Contour Plot (BSCP)	Used Archived Sensor Data and NJTPK dispatcher data			
Raub (1997)	Clearance time+15 minutes	< 1 mile				

Karlaftis et.al (1999)	Clearance time+15 minutes	< 1 mile	Excluded non-crashes & secondary crashes in the opposite direction		▪ Clearance time ▪ Season ▪ Type of vehicle involved ▪ Lateral location of the primary crash
Chang (2003)	2 hours-same direction ½ hours -opposite direction	< 2 miles downstream for same direction or <0.5 mile downstream or upstream in the opposite direction			
Moore et al (2004)	Clearance time < 2 hours	Queue length< 2 miles	Excluded secondary crashes in the opposite direction		

APPENDIX B: DATA EXTRACTION FROM TRIMS

B.1 Pooling data

The trims database contains 31 tables which are listed in table B-1.

Table B-1 Trims database tables.

ID	Table Name	ID	Table Name
1	Bridgeview_Shelby	**17**	Ramp Traffic
2	Crash_mot_nonmot	**18**	Road History
3	Crash_Nonmotorist	**19**	Road Segment
4	Crash_Property_Owners	**20**	Road Segment Control Items
5	Crash_Truck-bus	**21**	Road Segment HPMS
6	Crash Driver	**22**	Road Segment Other
7	Crash Feature	**23**	Road Segment Road Names
8	Maintenance Feature	**24**	Road Segment Special
9	Maintenance Inventory	**25**	Road Segment US Routes
10	PPRM Projects	**26**	Road System
11	Railroad Crossing	**27**	Roadway Description
12	Ramp Interchange	**28**	Roadway Geometrics
13	Ramp Parent	**29**	Route Feature
14	Ramp Road Segment	**30**	Structures
15	Ramp Roadway Description	**31**	Traffic
16	Ramp route Feature		

Trims database provides two tools for data mining. The first tool is designed for desktop, which is working by sending the instructions defined by user to the trims online server, and obtains the results from it. The benefit of using the desktop tool, is ability to obtain large datasets without disconnection, download just the desired data and faster interface. The drawback of the desktop tool is lack of demonstration of data on map, inability to download the GIS results and problem in limited supported operating systems. The second tool is an online web based tool. The online tool has similar features to the desktop version with some differences. The main difference is the ability of online tool to demonstrate the queried data on map which was crucial in this study. The other feature that is missing from the desktop tool is ability to produce and download the queried shape file. The more extensive analyzes can be done by defining the relationship between the shape file(s) and other source of data. The shape files of crashes and network links were downloaded using the trims' online tool.

One problem of working with online version of trims is that it is not suitable for downloading large amount of data. One of the reasons is the timeout problem in server-client connection. So the tabular data were obtained using desktop version of trims. Also getting the results from the spreadsheet on dialog tool, produces text based results on most of cells of table. Besides the compression problem, it is not suitable for analyzing, modeling and any algorithm that may want to use. So designing a data structure which stores data as numerical codes (and its associated metadata to describe it) could help.

From TRIMS desktop, 31 different tables were downloaded in .csv format that contained data from January 1, 2010 until December 31, 2012. Some of the large tables were downloaded by segments. The segmented data later was appended in correct order to obtain the complete table. The downloaded tables were then imported into SQL server 2014. All the tables in trims database, contains a filed named 'mlink'. This field provides a unique identifier for joining the tables. List of all tables were obtained via TRIMS is provided in the table B-2.

Table B-2 List of tables from TRIMS database used in the study.

Table Name	Number of records
BRIDGEVIEW_SHELBY	1688
CRASH_DRIVER	168428
CRASH_FEATURE	1628
CRASH_MOT_NONMOT	15701
CRASH_NONMOTORIST	15249
CRASH_PROPERTY_OWNERS	158363
EXPORT_CR	8325
HELP_MARCH_2012	2210
HELP_MVC_R4_13	1249
HELP_SEPTEMBER_2012	2594
HELP_SEPTEMBER_2012old	2595
HELP_SVC_R4_13	317
INTERSECTION	42436
MAINTENANCE_FEATURE	26504
MAINTENANCE_INVENTORY	25447
PPRM_PROJECTS	476
Railroad Crossing	498
RAMP_INTERCHANGE	85
RAMP_PARENT	569
RAMP_ROAD_SEGMENT	569
RAMP_ROADWAY_DESCRIPTION	2845
RAMP_ROUTE_FEATURE	1138
RAMP_TRAFFIC	566
ROAD_HISTORY	620
ROAD_SEGMENT	16209
ROAD_SEGMENT_CONTROL_ITEMS	12913
ROAD_SEGMENT_HPMS	311
ROAD_SEGMENT_OTHER	12428
ROAD_SEGMENT_ROAD_NAMES	15194
ROAD_SEGMENT_SPECIAL_SYSTEMS	174
ROAD_SEGMENT_US_ROUTES	33
ROAD_SYSTEM	12355
ROADWAY_DESCRIPTION	75450
ROADWAY_GEOMETRICS	17914
ROUTE_FEATURE	92011
SHELBY_CR	91325
SHELBY_RAMPS	1565
STRUCTURES	1688
T_BL_LRS_COLUMNS	1277
TBL_LRS_CODEDESCRIP	4492
TRAFFIC	932
TRAFFIC_COUNTY_SHELBY	937
TRAFFIC_COUNTY_SHELBY_CREATE_RTE	454

B.2 Notable fields

The crash related tables have important information needed to identify primary and secondary crashes. Each crash associated with one unique id called 'CASENO_129'. This field is used to find the relationship between each crash, their attributes, driver's characteristics etc.

The table also contained spatial and temporal information. Temporal data is date and time of crash, with precision of minutes and also seconds. The spatial information includes geographic coordination, state and county, route name and other relevant data. Along the geographic information, LRS[2] data is also provided. The route id and BLM[3] and ELM[4] of LRS help to identify location of vehicles along each route and the distance between them. An example of BLM field of interstate I-40 and I-240 is shown in the Fig B-1. The variation in BLM value can be seen by changes in ranges of color. The BLM values increase from west to east on these interstates.

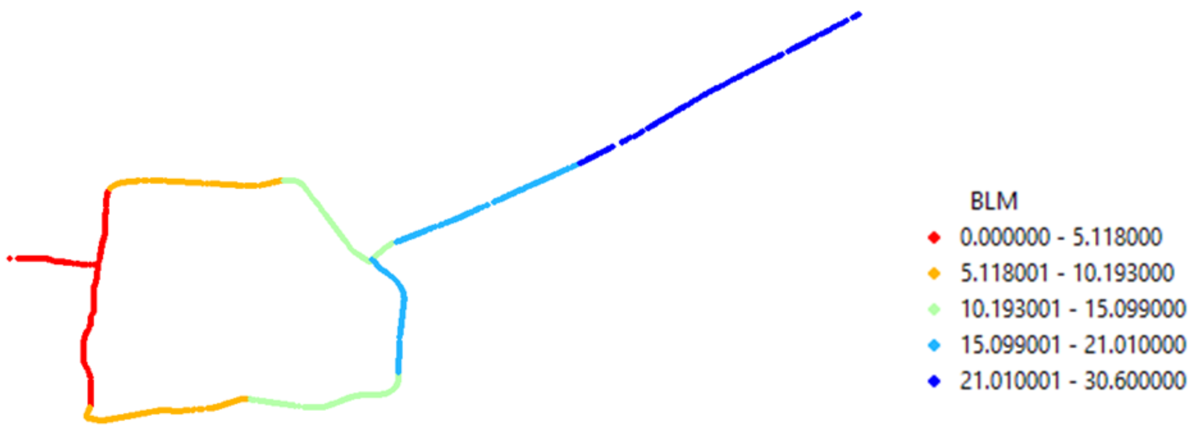


Figure B-1 Example of BLM field of interstate I-40 and I-240.

B.3 Detector data

The SCIA algorithm used shockwave principles and hence requires information about traffic condition at the time of primary crash. According to the detector database, I-40, I-240, I-55 and SR385 contain detectors. The database contained date/time stamp, location, lane number, counts and average speed as shown in the Table B-3.

[2] Linear Referencing System
[3] Begin Log Mile
[4] End Log Mile

Table B-3 Example of detector data.

Detector ID	Lane #	Longitude	Latitude	Date	Time	Count	Average Speed
62005601	1	-90.0424	35.15187	1/1/2012	0:00	44	66
62005602	2	-90.0424	35.15187	1/1/2012	0:00	168	70
62005603	3	-90.0424	35.15187	1/1/2012	0:00	76	82
62005701	1	-90.0345	35.16	1/1/2012	0:00	0	0
62005702	2	-90.0345	35.16	1/1/2012	0:00	0	0
62005703	3	-90.0345	35.16	1/1/2012	0:00	0	0
62005801	1	-90.0345	35.16	1/1/2012	0:00	148	31
62005802	2	-90.0345	35.16	1/1/2012	0:00	0	0
62005803	3	-90.0345	35.16	1/1/2012	0:00	0	0
62006101	1	-90.0216	35.1514	1/1/2012	0:00	108	75
62006102	2	-90.0216	35.1514	1/1/2012	0:00	252	76
62006103	3	-90.0216	35.1514	1/1/2012	0:00	116	74

The data obtained from detected database was joined to each crash, and obtaining the volume and speed of the flow at the time of primary crash. Another important table regarding the detectors contained their descriptive information. This table shows the identification, route and lane information, geographic coordination. The direction of detectors can be determined by interpreting the text field *'DETECTOR_NAME'*.as shown in Table B-4. Figure B-2 shows the histogram results for a particular detector (ID: 62014601) located at 'I-40 WB at US-64 Overpass' on January 30th, 2010. The peak hour in morning and afternoon can be identified by looking at this chart.

Table B-4 detector information.

DETECTOR_ID	DETECTOR_NAME	LANE_NUM	DEVICE_ID	X_COORD	Y_COORD
62000101	I-40 EB W/O Airport Rd. Intersection	1	20001	-90.22582	-90.22582
62000102	I-40 EB W/O Airport Rd. Intersection	2	20001	-90.22582	-90.22582
62000201	I-40 WB W/O Airport Rd. Intersection	1	20002	-90.22582	-90.22582
62000202	I-40 WB W/O Airport Rd. Intersection	2	20002	-90.22582	-90.22582
62005601	I-40 WB West End of Danny Thomas Intersection Ramps	1	20056	-90.04238	35.15187

62005602	I-40 WB West End of Danny Thomas Intersection Ramps	2	20056	-90.04238	35.15187
62005603	I-40 WB West End of Danny Thomas Intersection Ramps	3	20056	-90.04238	35.15187
62005701	VDS I-40 WB East of Danny Thomas	1	20057	-90.03450	35.16000
62005702	VDS I-40 WB East of Danny Thomas	2	20057	-90.03450	35.16000
62005703	VDS I-40 WB East of Danny Thomas	3	20057	-90.03450	35.16000
62005801	VDS I-40 EB East of Danny Thomas	1	20058	-90.03450	35.16000
62005802	VDS I-40 EB East of Danny Thomas	2	20058	-90.03450	35.16000
62005803	VDS I-40 EB East of Danny Thomas	3	20058	-90.03450	35.16000
62006101	I-40 EB Between Autumn Ave. and N. Parkway	1	20061	-90.02155	35.15140
62006102	I-40 EB Between Autumn Ave. and N. Parkway	2	20061	-90.02155	35.15140
62006103	I-40 EB Between Autumn Ave. and N. Parkway	3	20061	-90.02155	35.15140
62006201	I-40 WB Between Autumn Ave. and N. Parkway	1	20062	-90.02168	35.15195
62006202	I-40 WB Between Autumn Ave. and N. Parkway	2	20062	-90.02168	35.15195
62006501	I-40 EB Between Jackson Ave. and Vollintine Ave.	1	20065	-90.01853	35.16000

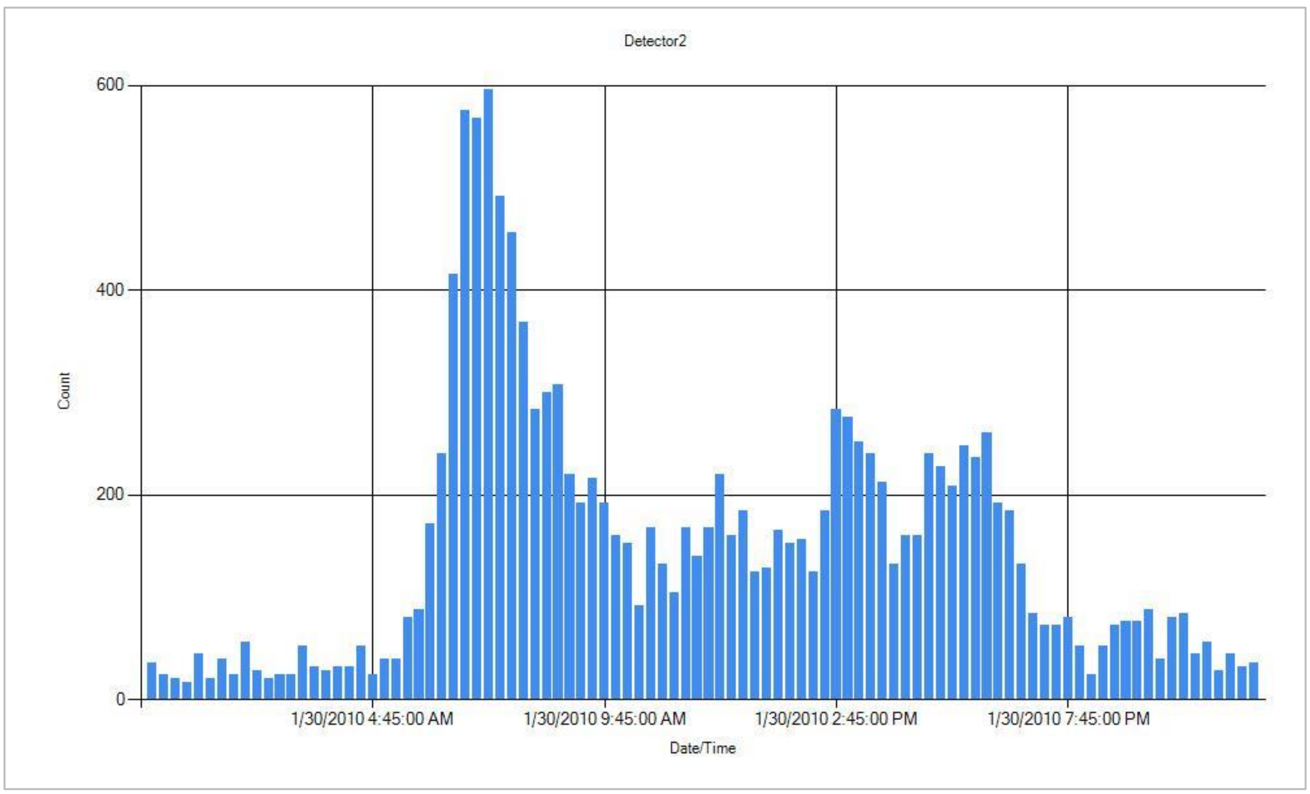


Figure B-2 Example of histogram of volume and average speed of a detector.

B.4 Data issues

Figure B-3 demonstrates the inconsistency in direction of BLM on Shelby county road sections. The low values of BLM showed with warmer colors. As the BLM value increases, the road section color gradually turns to cooler range. As it can be, the BLM is not always increasing in N/W directions. While the majority of the section follows this pattern, there are areas where the BLMs are increasing in S/E directions too. This issue is important in identifying the positions of the vehicles respect to each other (which one is following the other one). The positions of vehicles were determined by looking at their Latitude/Longitude and the direction of the flow (which was determined by direction of vehicle 1 involved in each case number involving multiple vehicles).

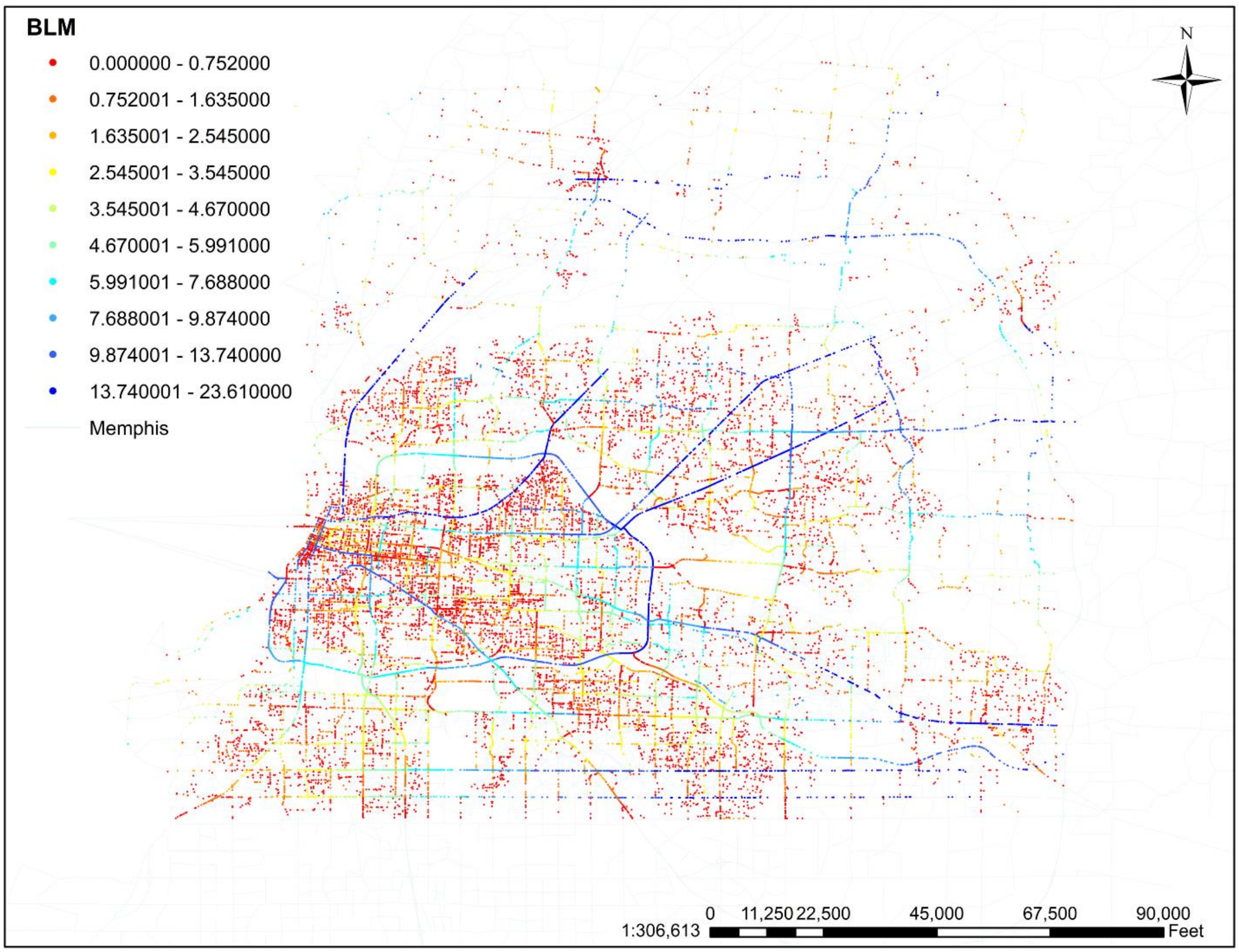


Figure B-3 The inconsistency of BLM value.

Issues regarding misclassification of records were also present in the data. For example, the location of a crash was categorized into categories such as: *Along Roadway, Ramp, Intersection,* and *Railroad Grade Crossing*. As it can be seen in the Fig B-4, the crash occurred on the interstate, but they were classified as "*on ramp*"(green triangle represent on ramp crashes)

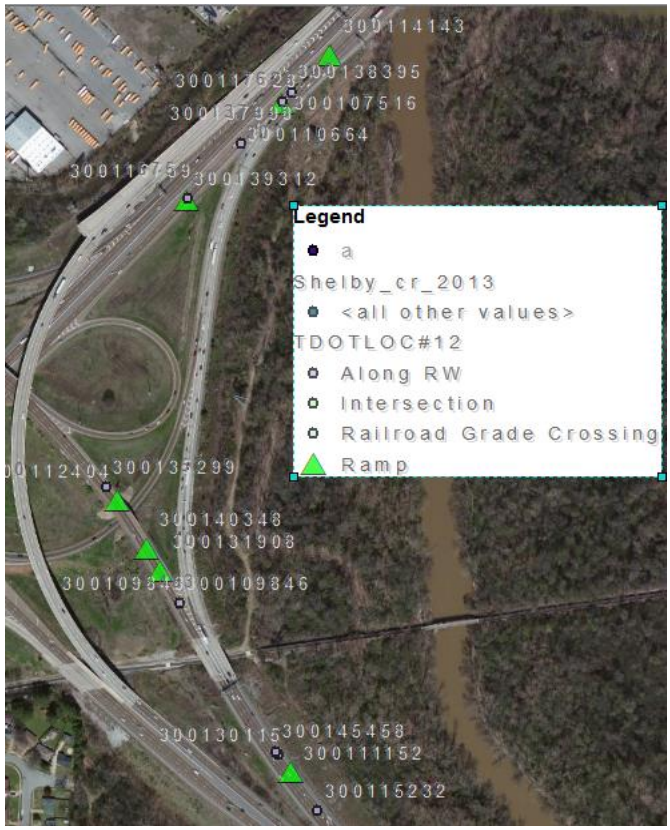


Figure B-4 Misclassification of crash location.

APPENDIX C: STATISTICAL ANALYSIS OF CRASH DATA

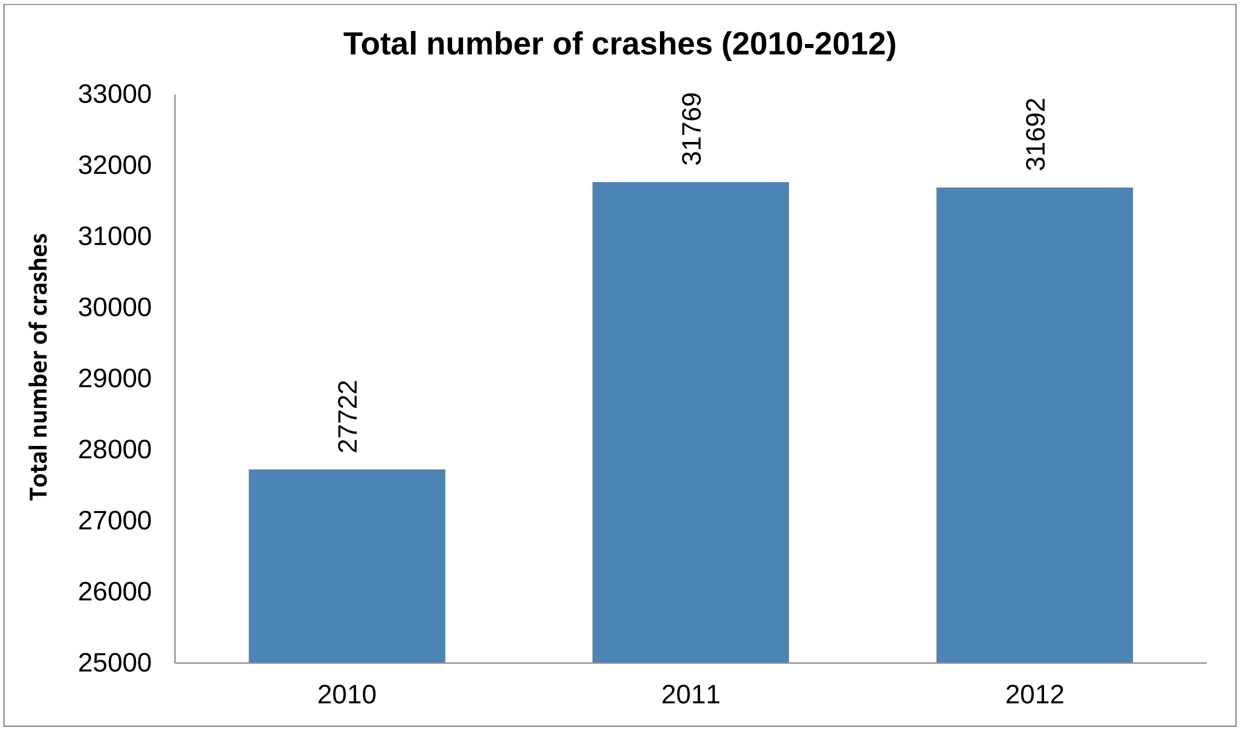


Figure C-1 Total number of crashes in Shelby County (2010-2012).

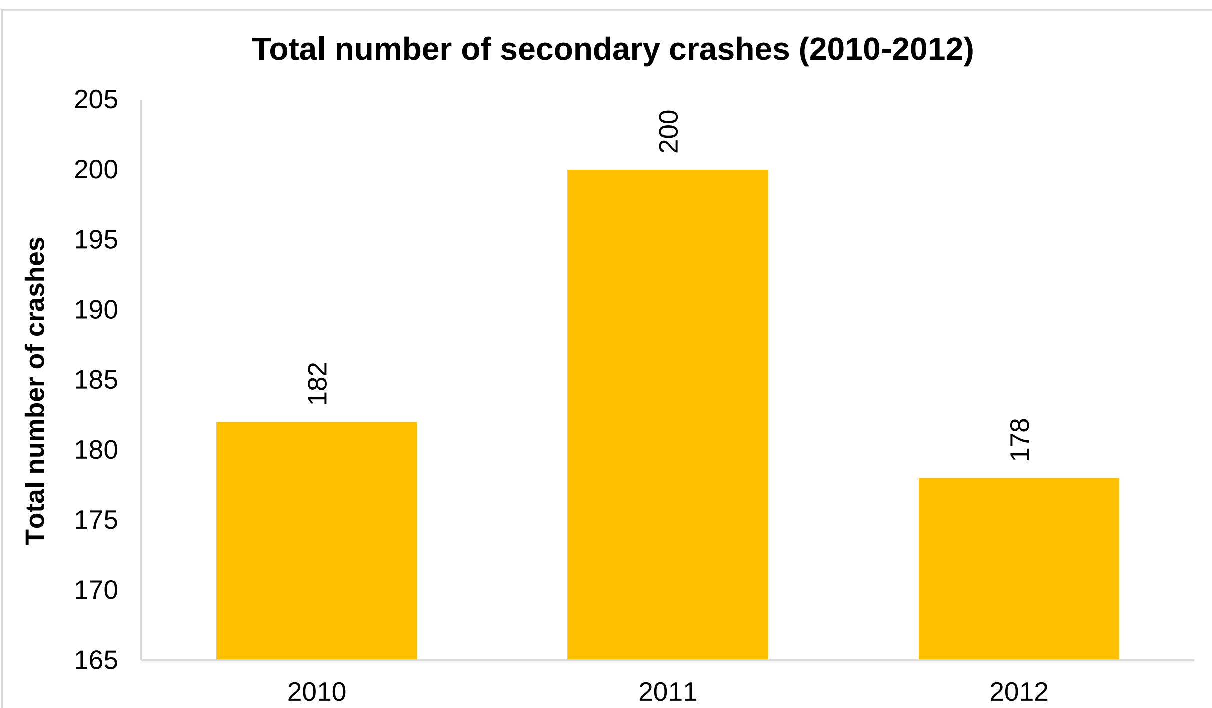


Figure C-2 Total number of SCs identified in Shelby County (2010-2012).

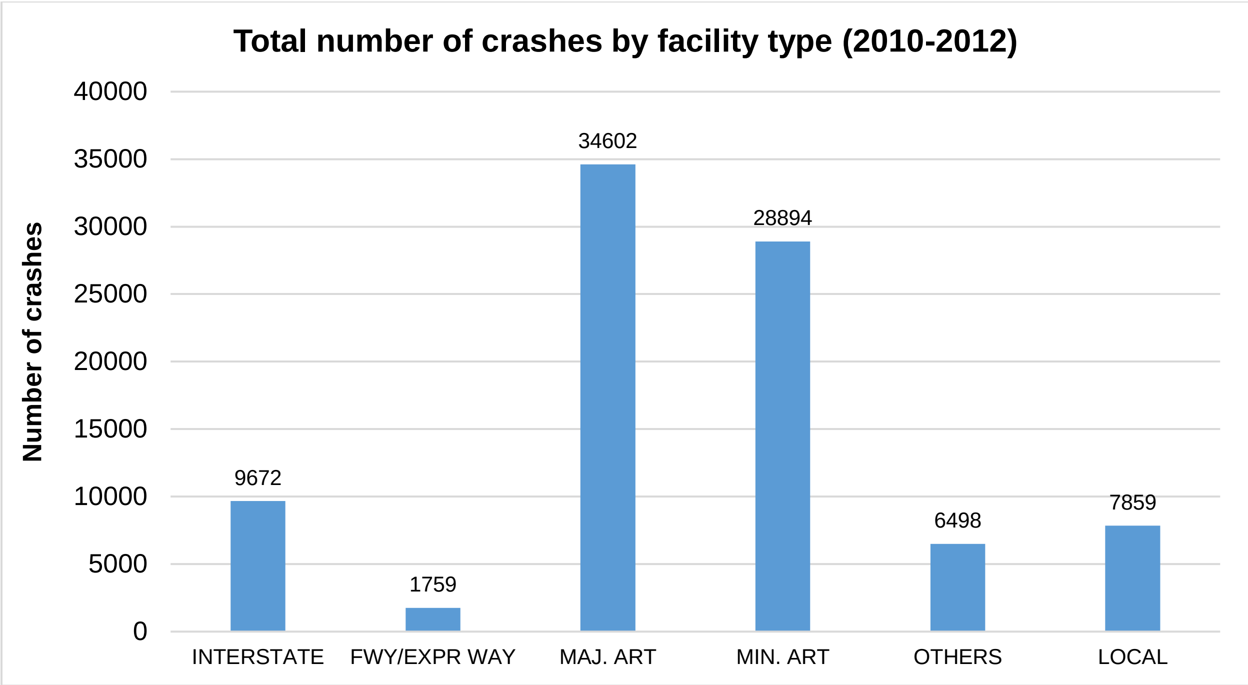


Figure C-3 Total number of crashes by facility type.

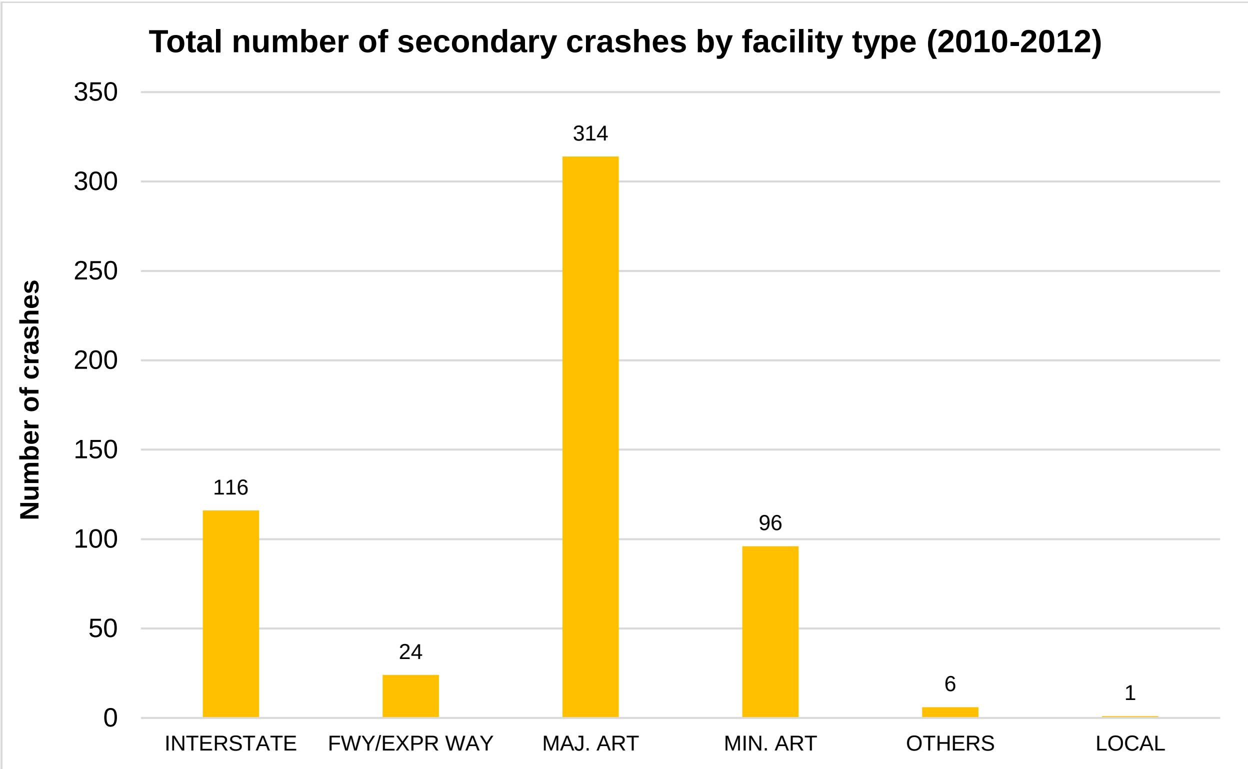


Figure C-4 Total number of SCs by facility type.

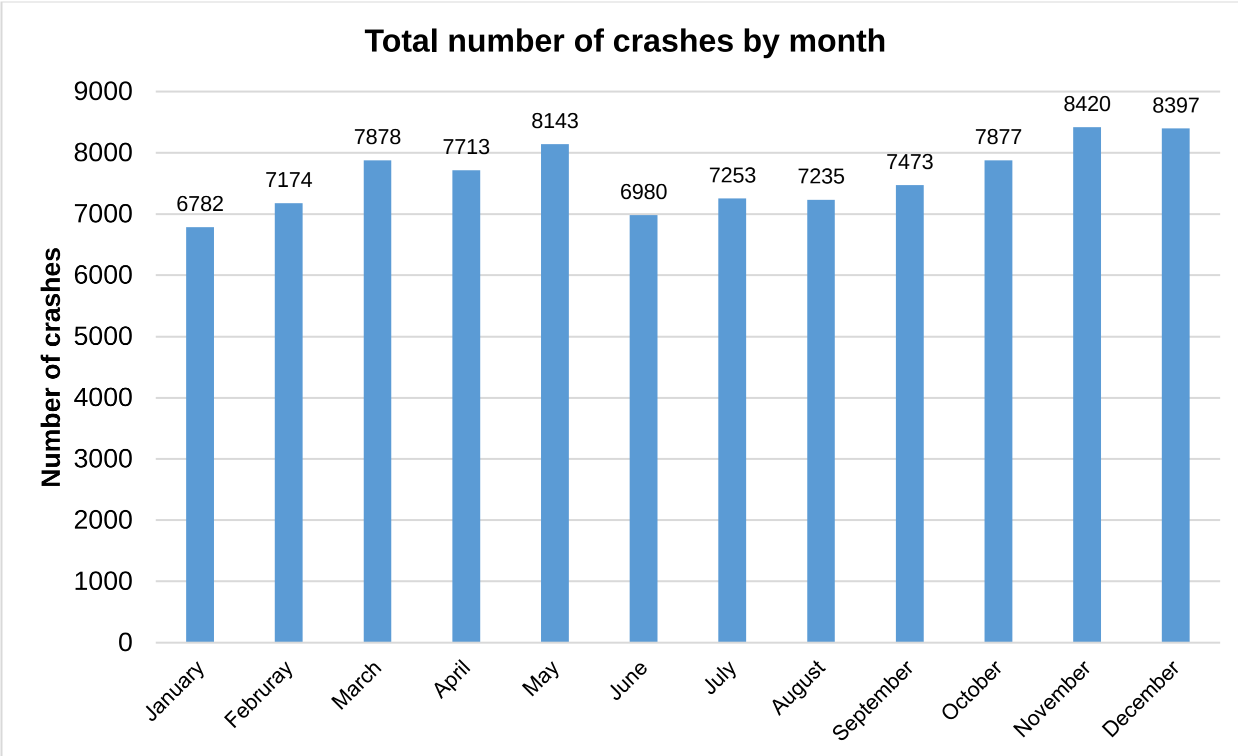


Figure C-5 Total number of crashes by month.

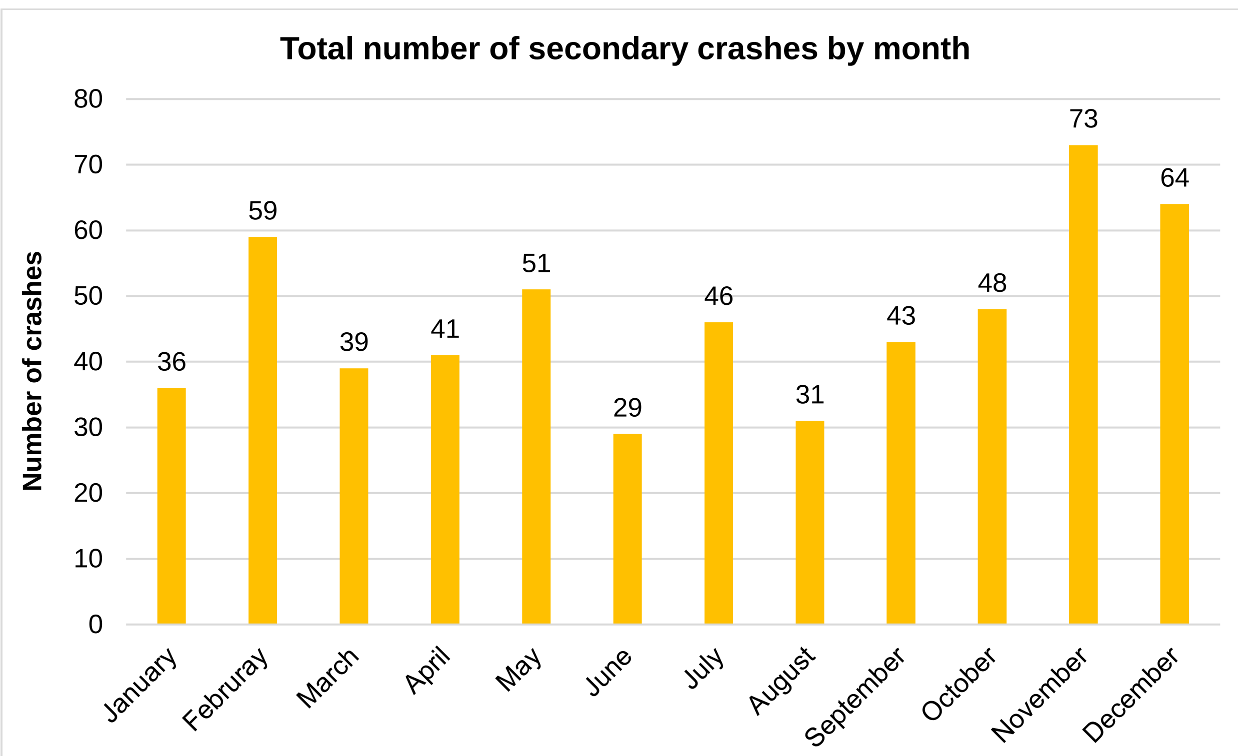


Figure C-6 Total number of SCs by month.

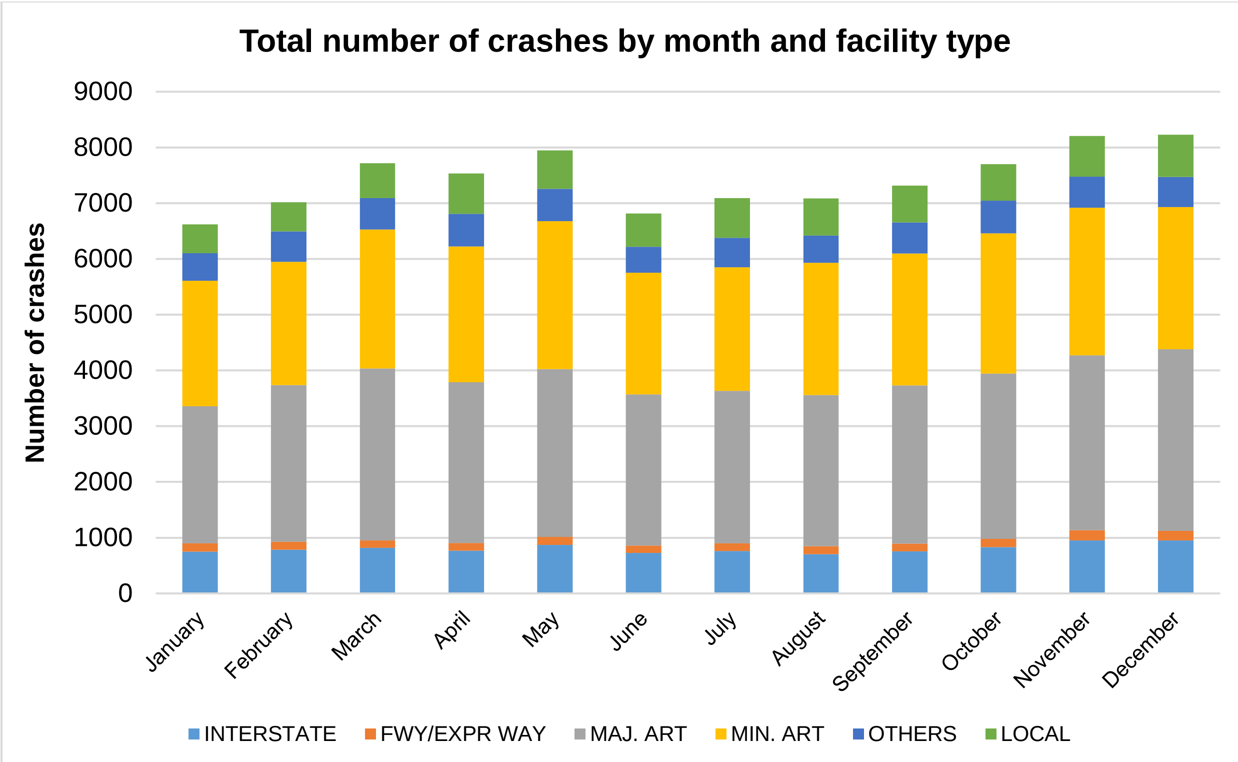


Figure C-7 Total number of crashes by month and facility type.

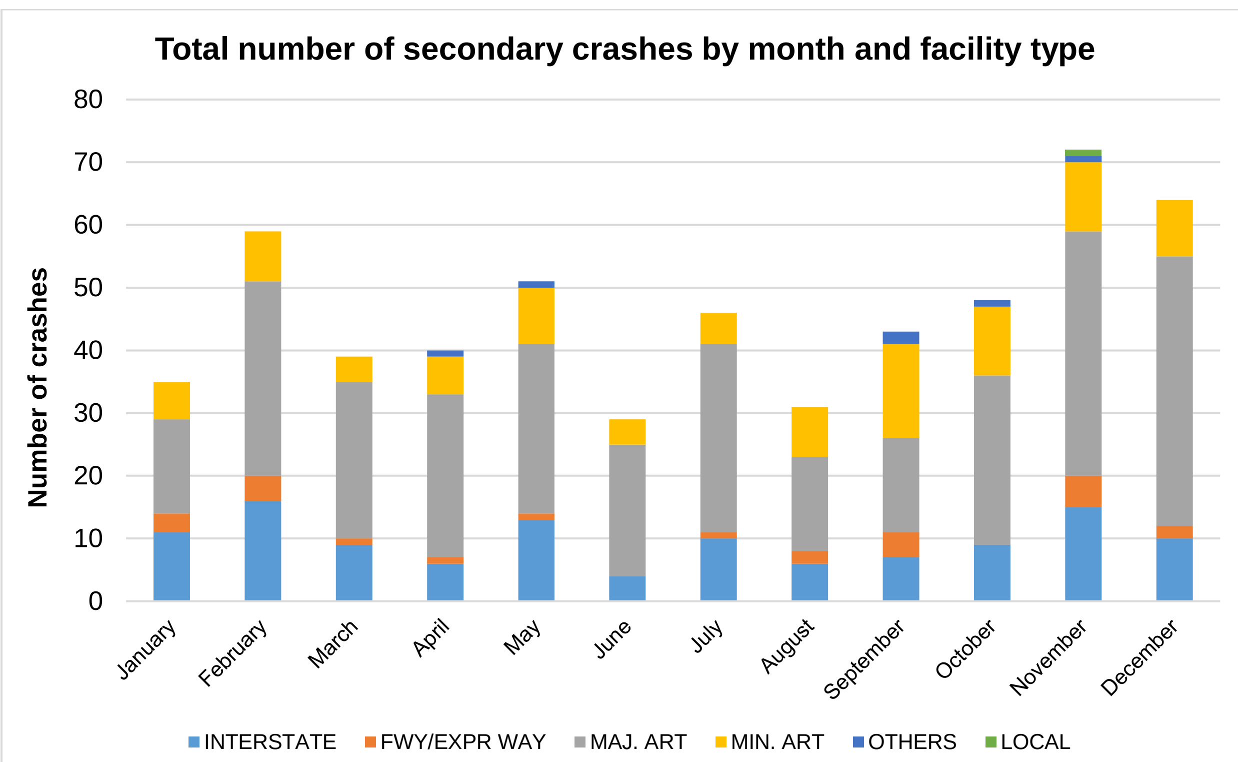


Figure C-8 Total number of SCs by month and facility type.

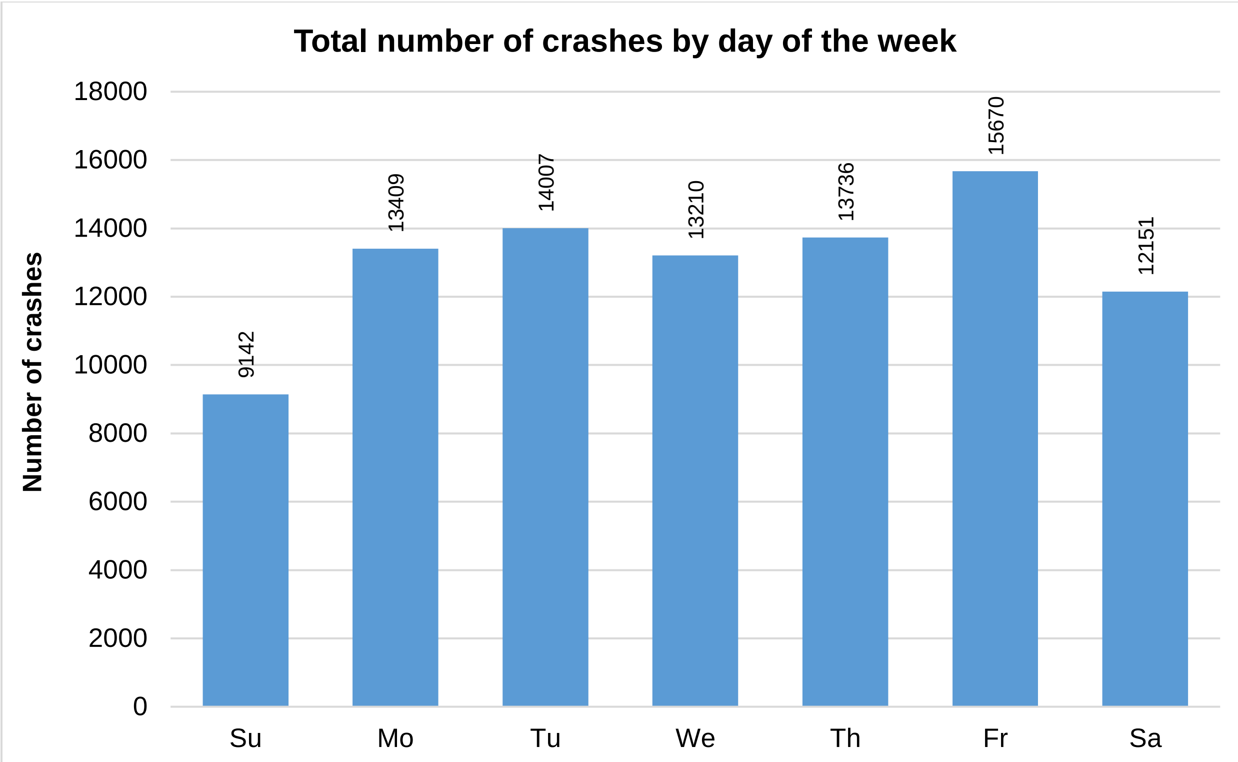


Figure C-9 Total number of crashes by day of the week.

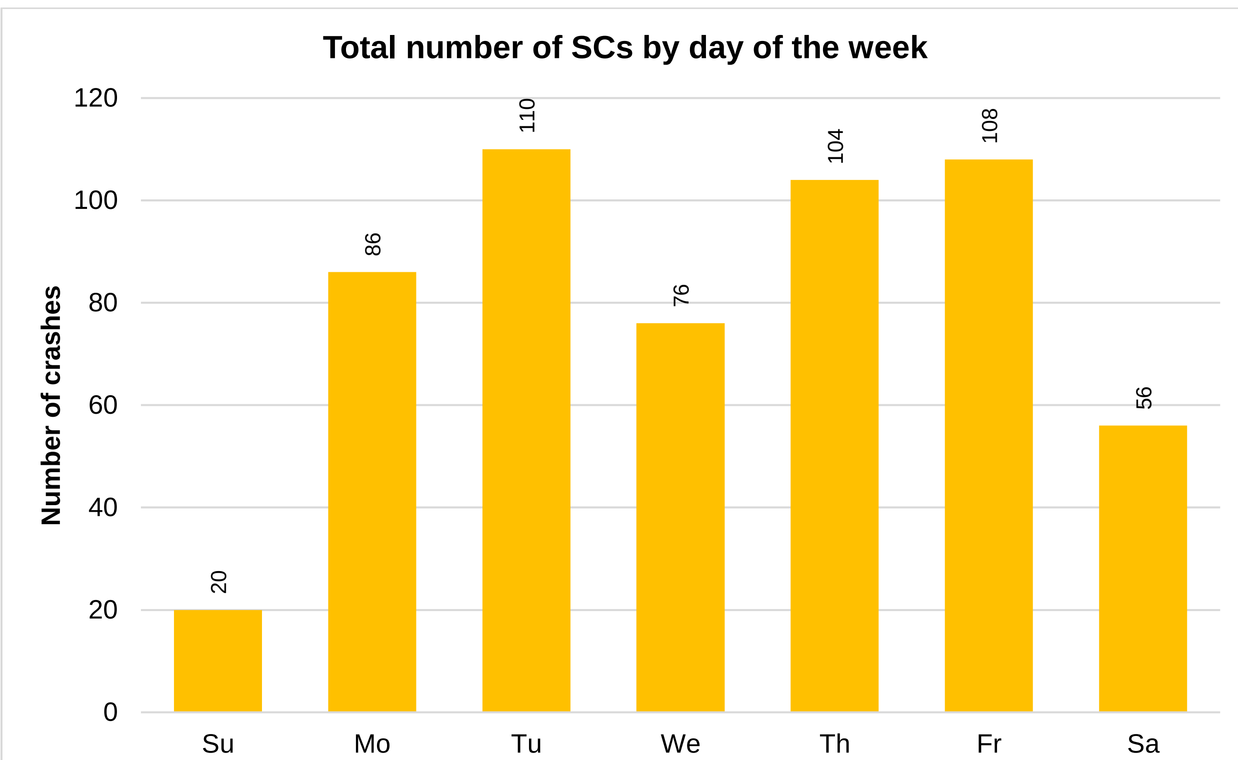


Figure C-10 Total number of SCs by day of the week.

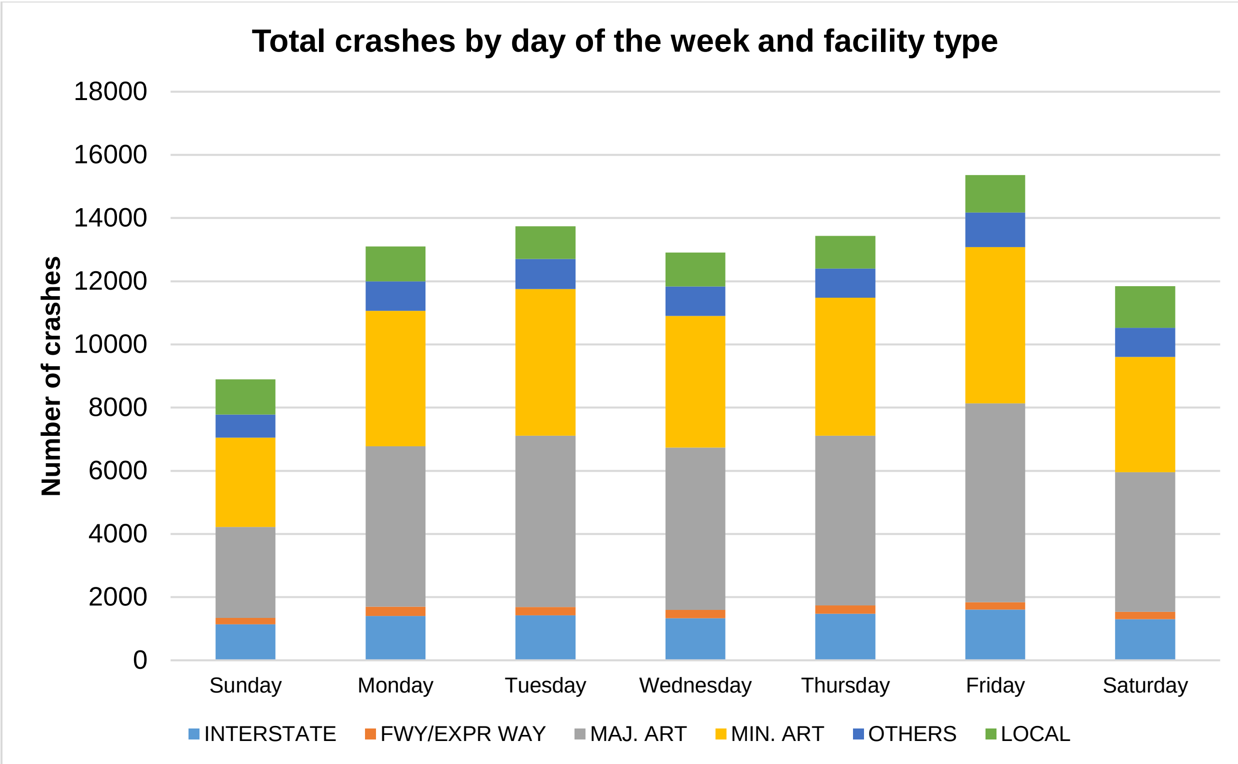


Figure C-11 Total number of crashes by day of the week and facility type.

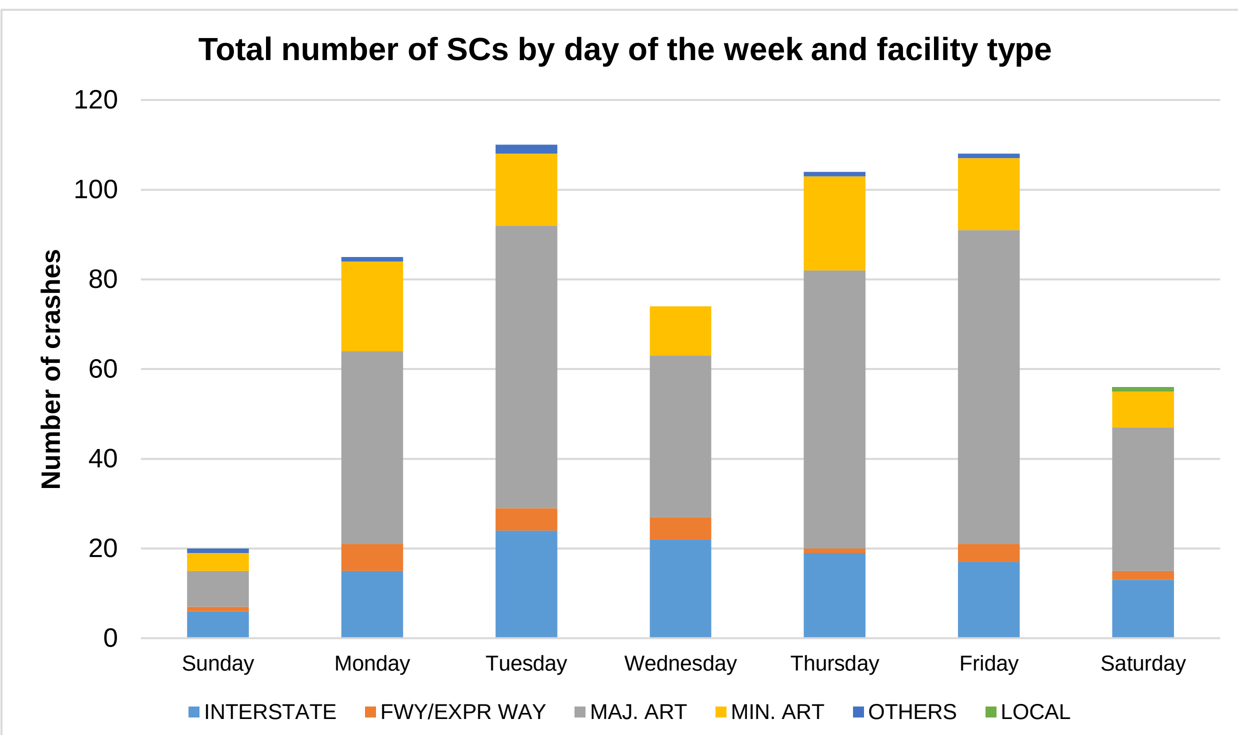


Figure C-12 Total number of SCs by day of week and facility type.

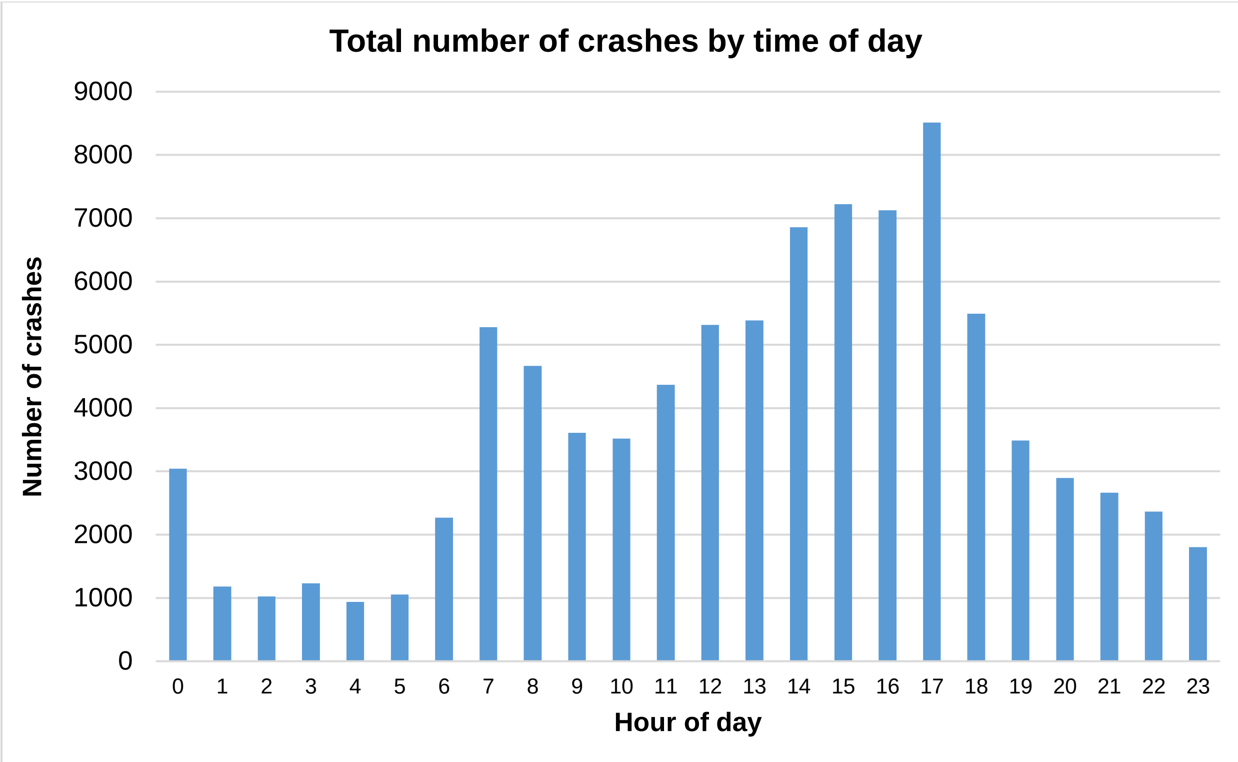


Figure C-13 Total number of crashes by time of day.

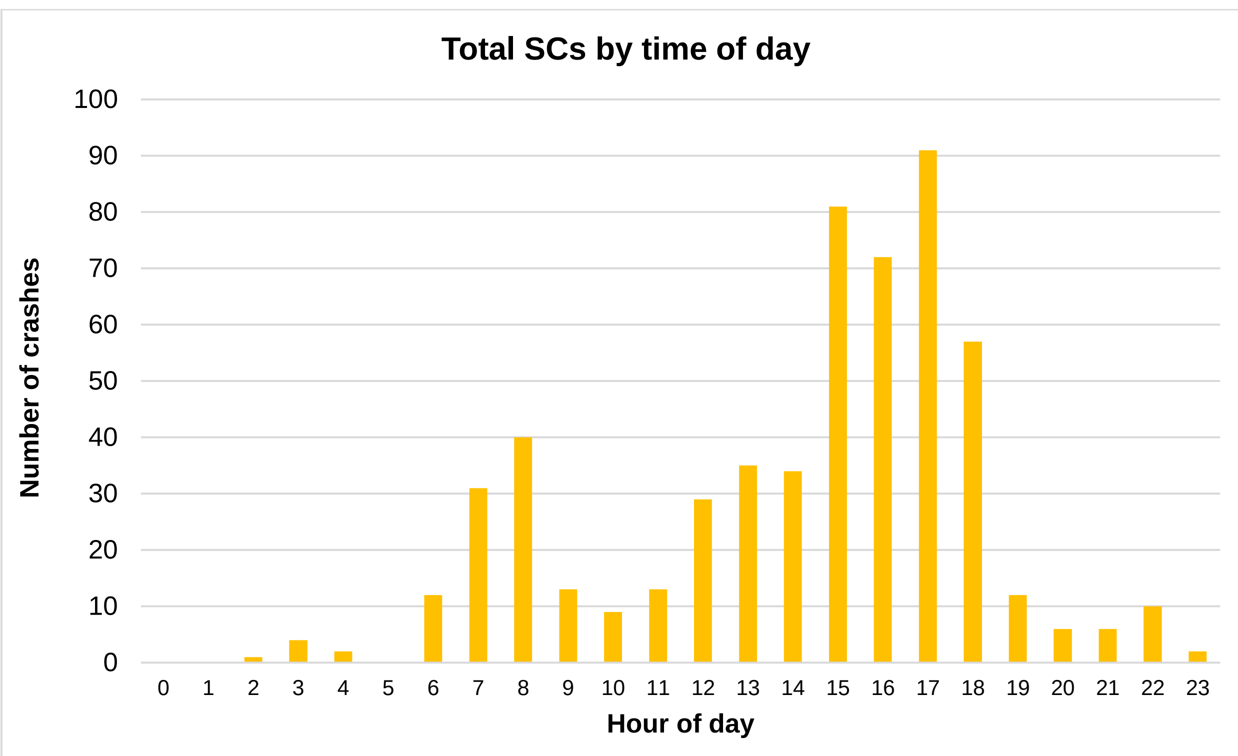


Figure C-14 Total number of SCs by time of day.

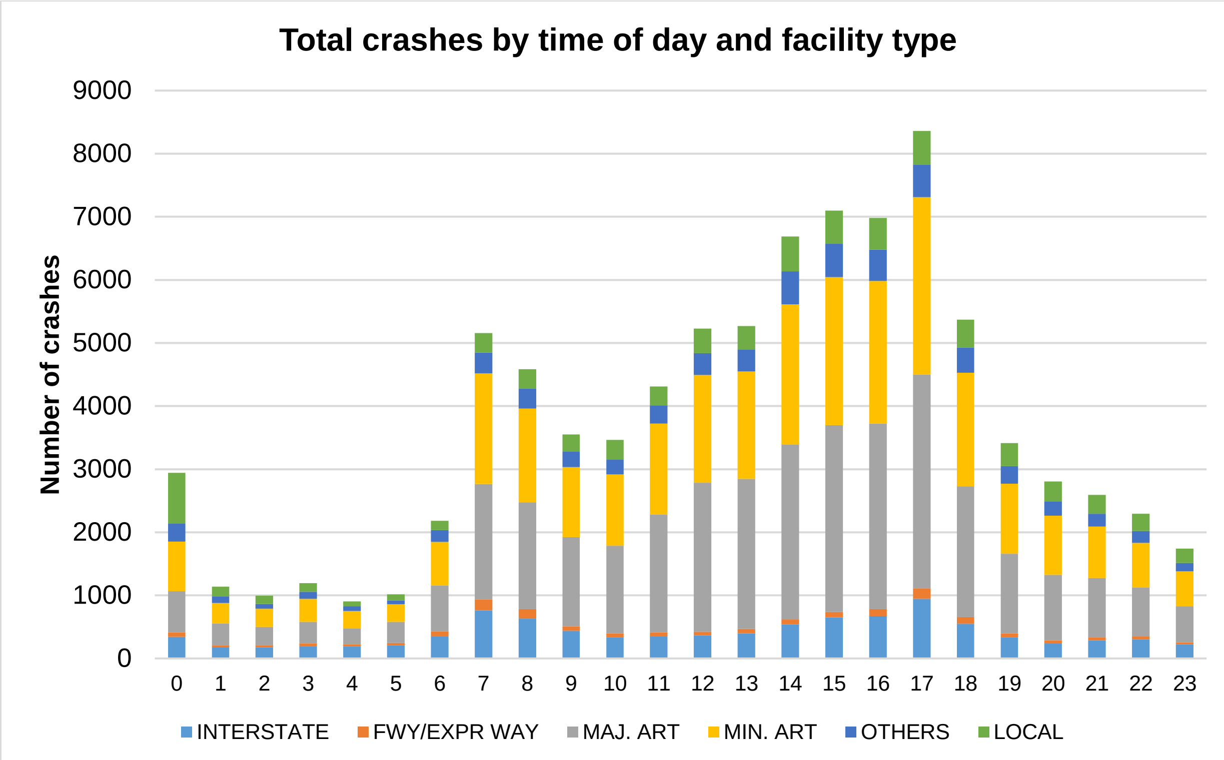


Figure C-15 Total number of crashes by time of day and facility type.

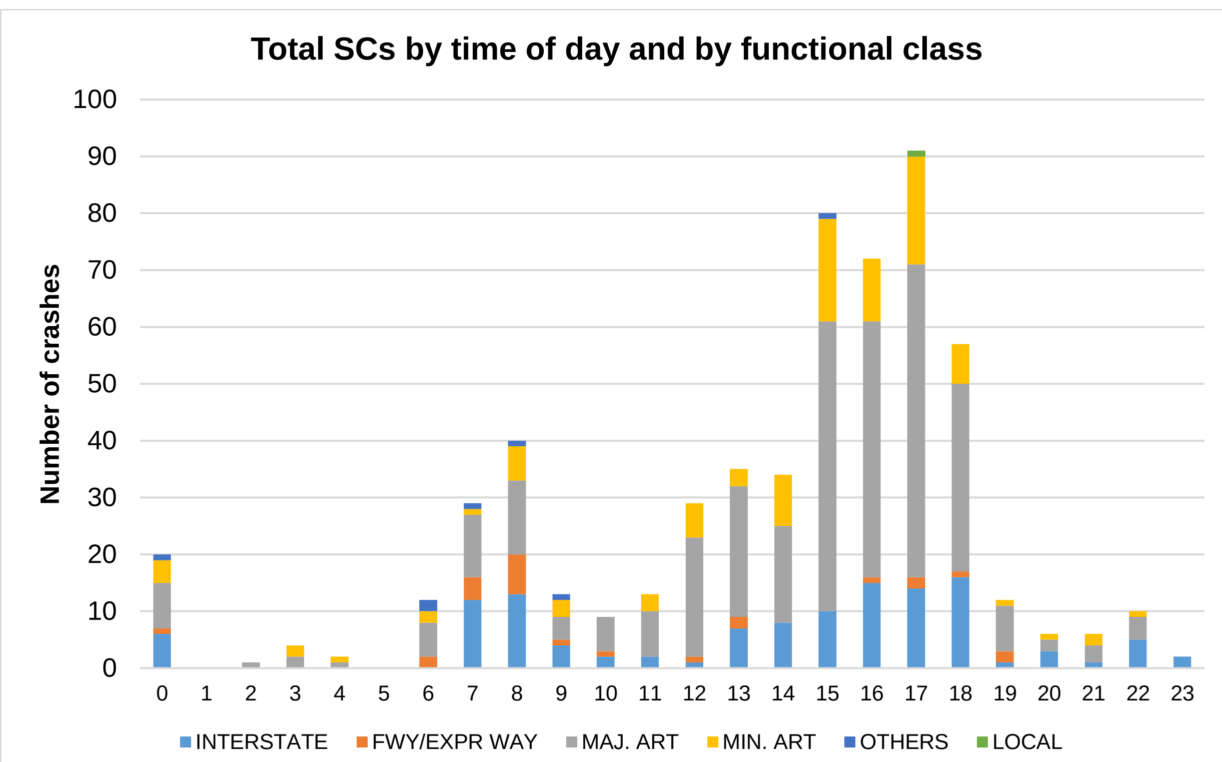


Figure C-16 Total number of SCs by time of day and facility type.

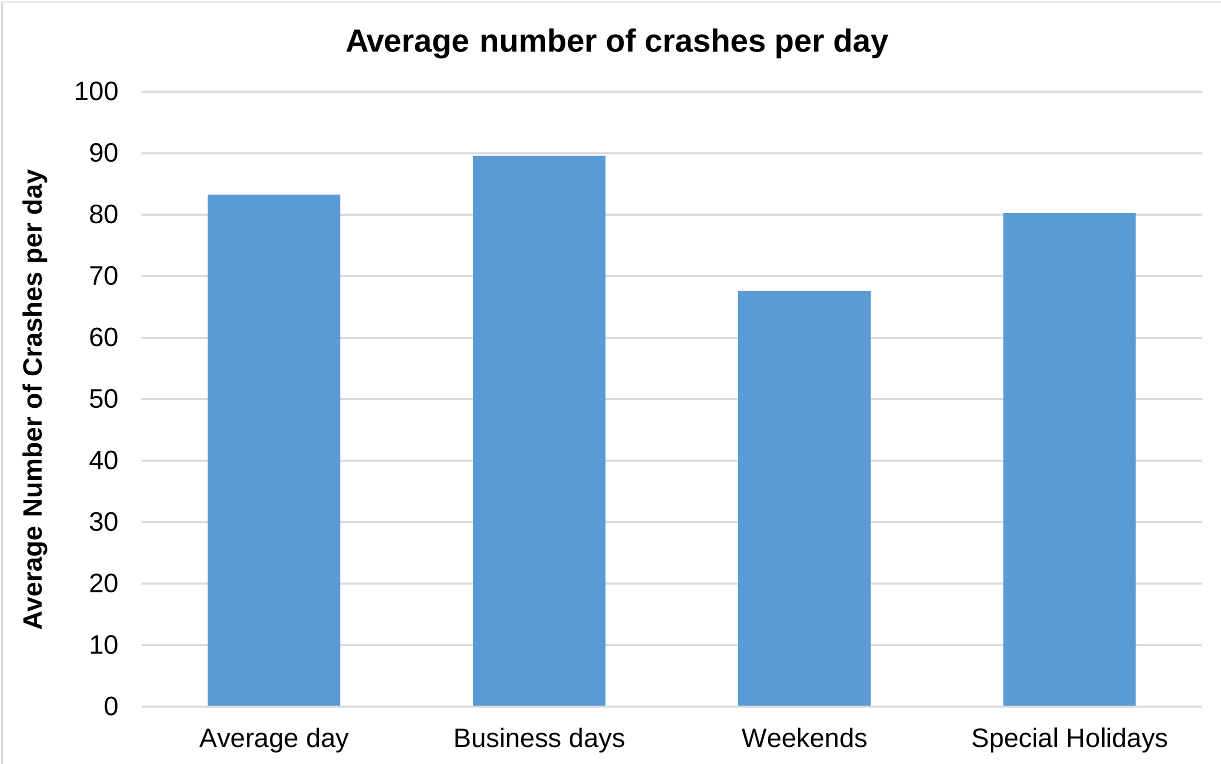


Figure C-17 Average number of crashes by holidays.

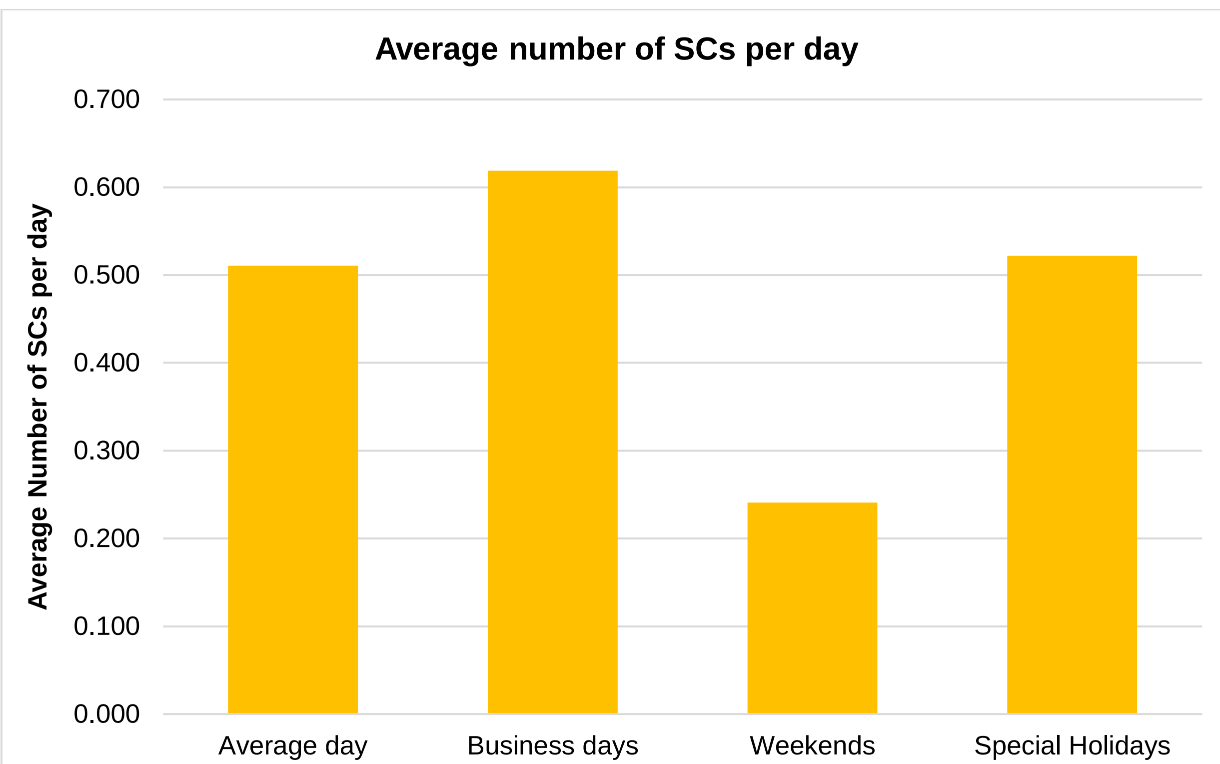


Figure C-18 Average number of SCs by holidays.

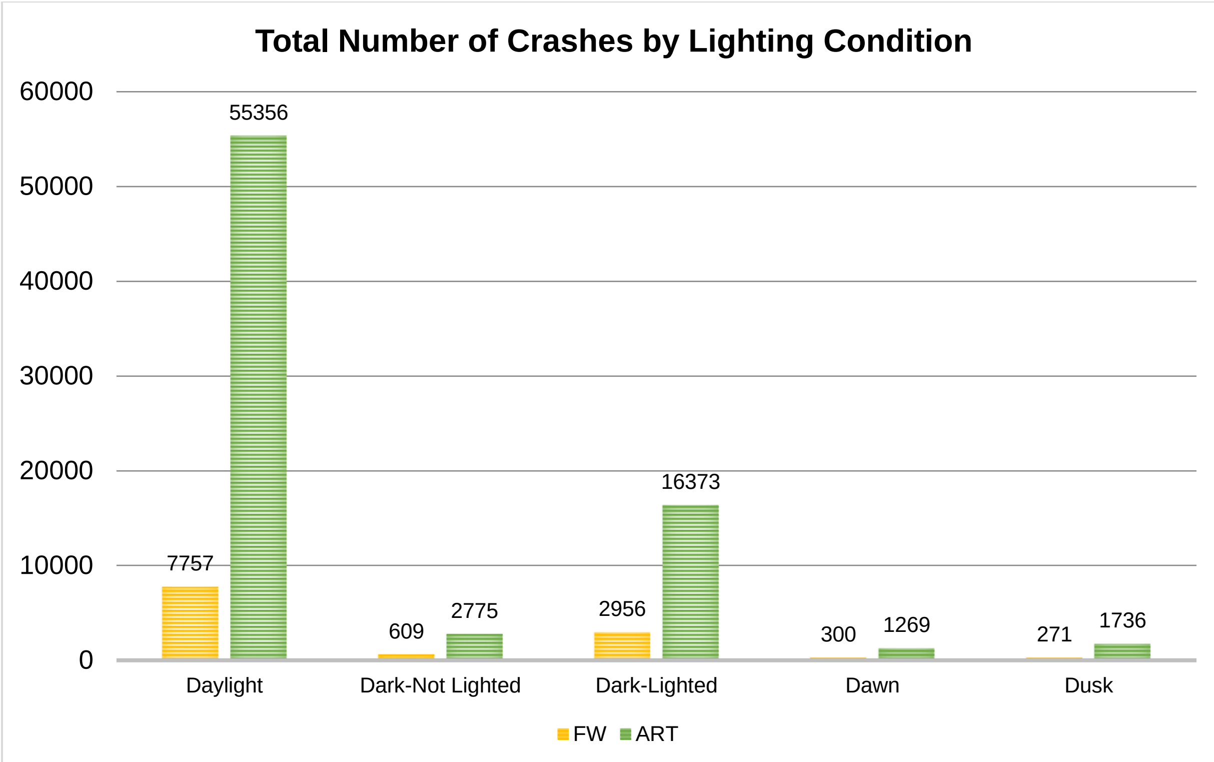


Figure C-19 Total number of crashes by lighting condition.

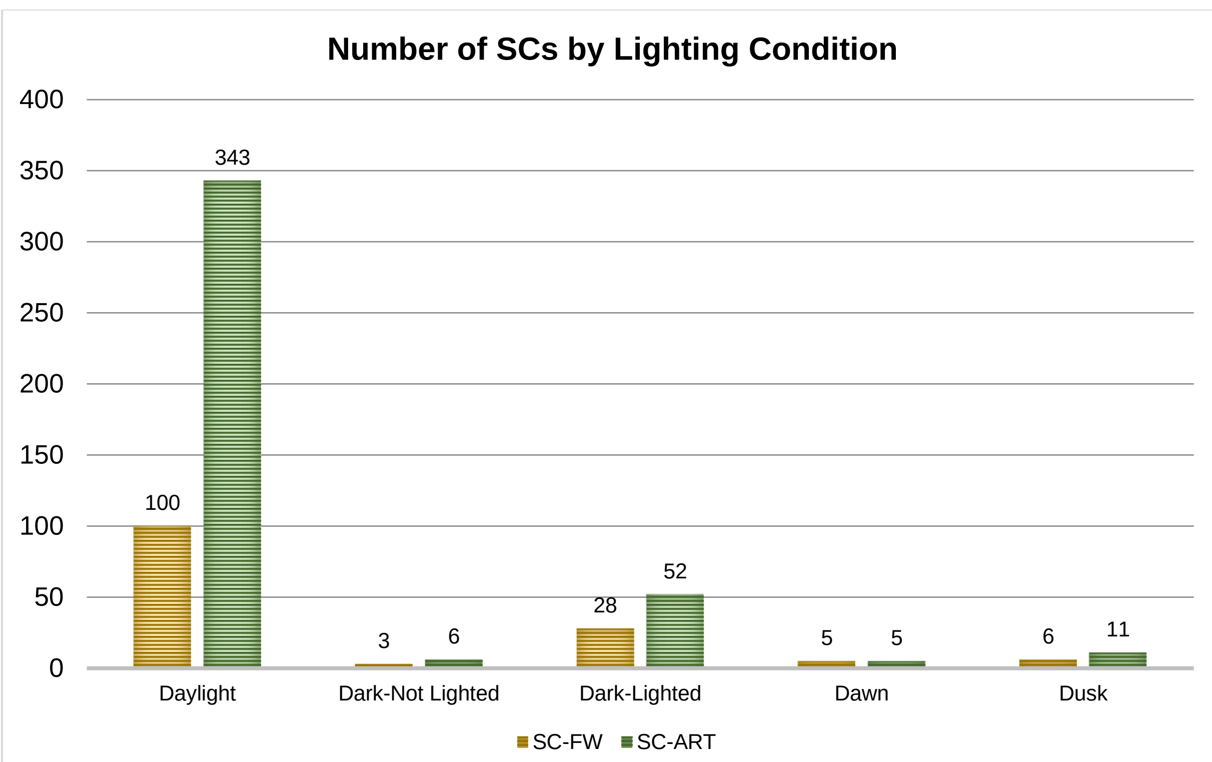


Figure C-20 Total number of SCs by lighting condition.

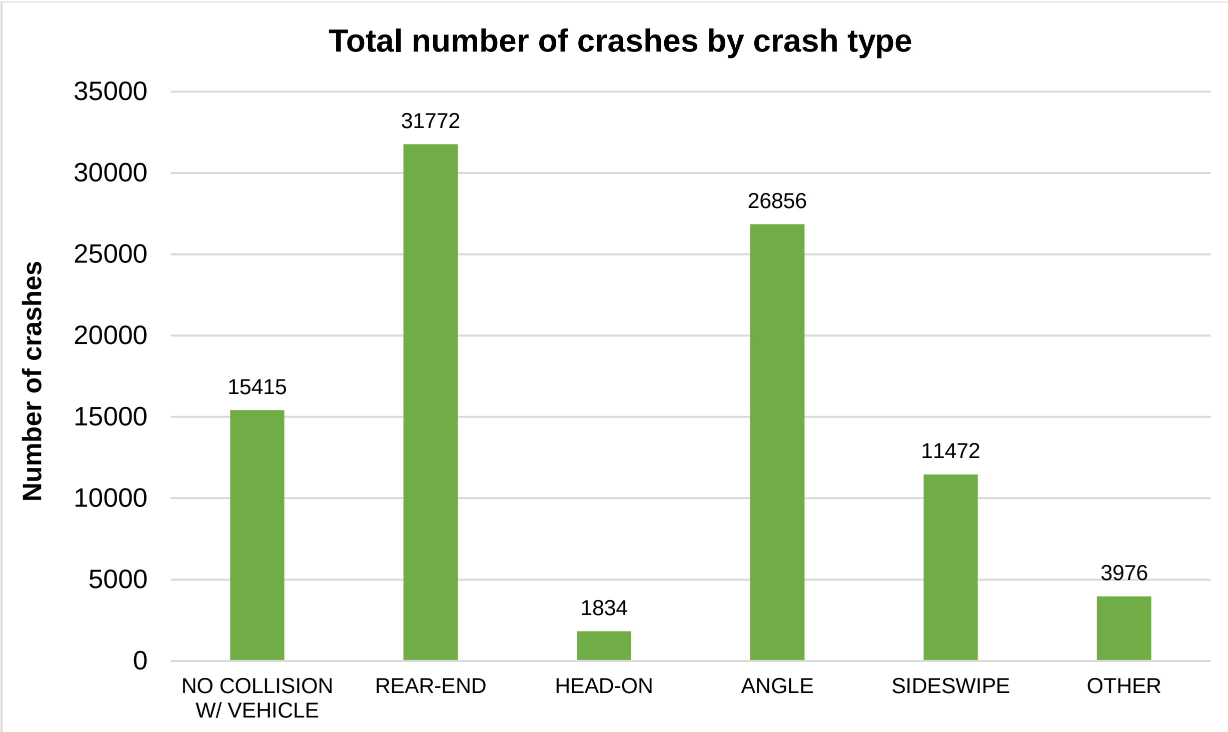


Figure C-21 Total number of crashes by crash type.

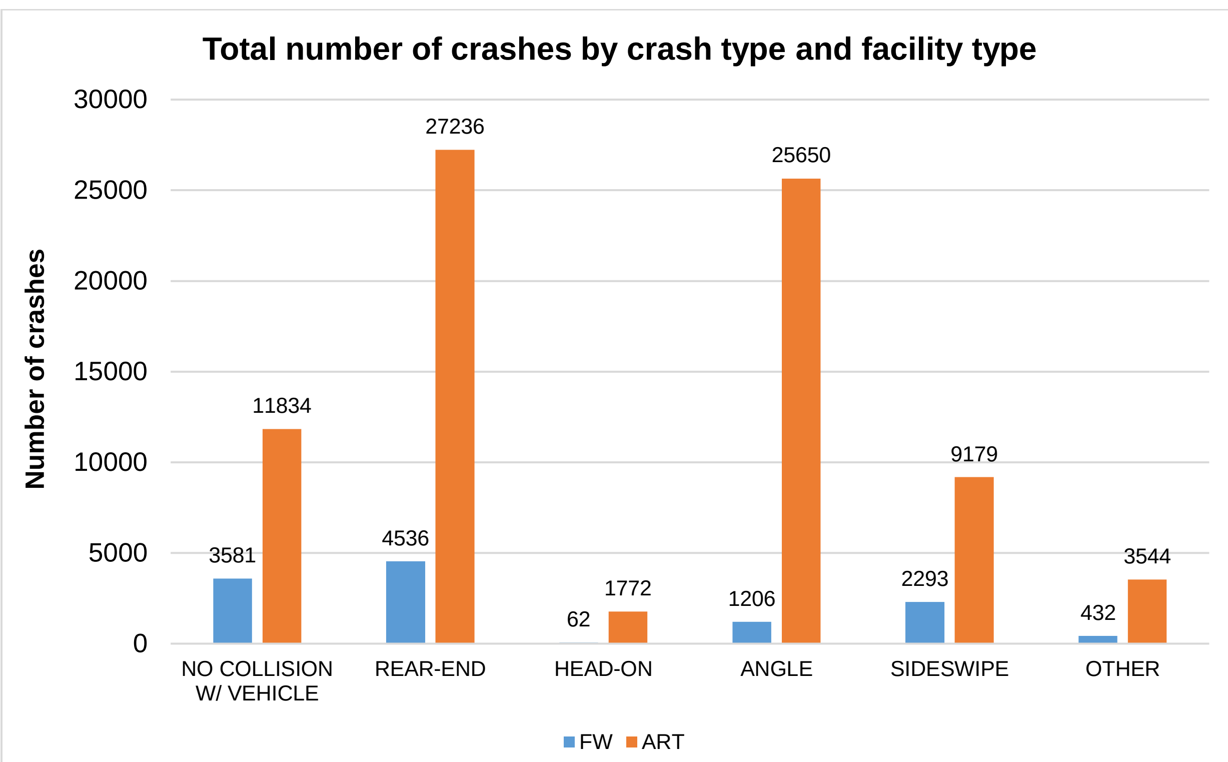


Figure C-22 Total number of crashes by crash type and facility type.

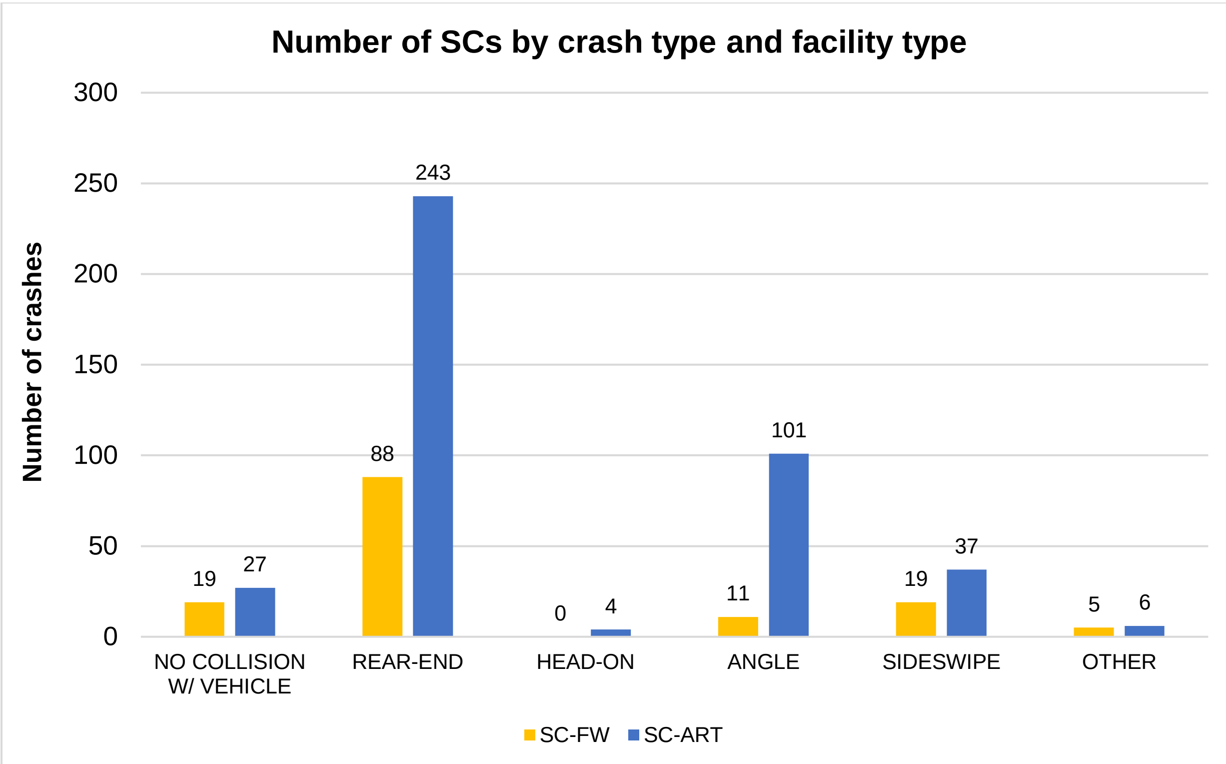


Figure C-23 Total number of SCs by crash type and facility type.

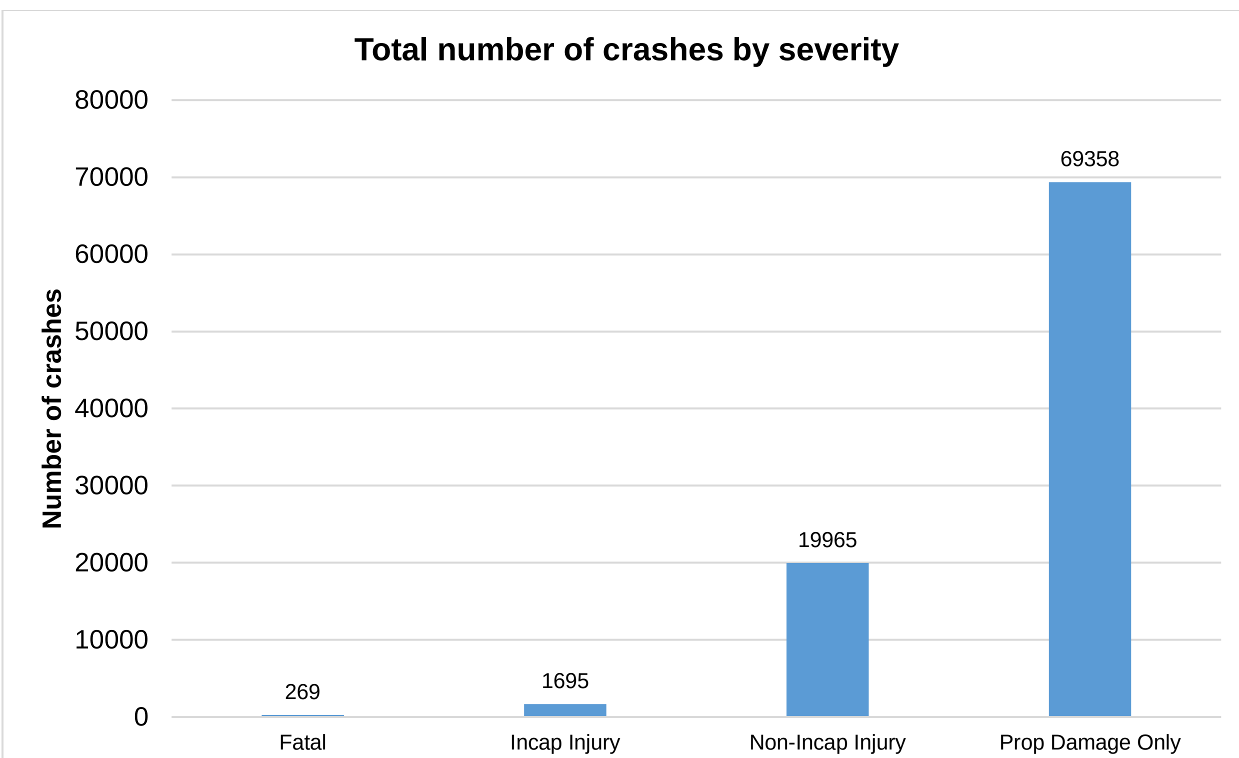


Figure C-24 Total number of crashes by severity.

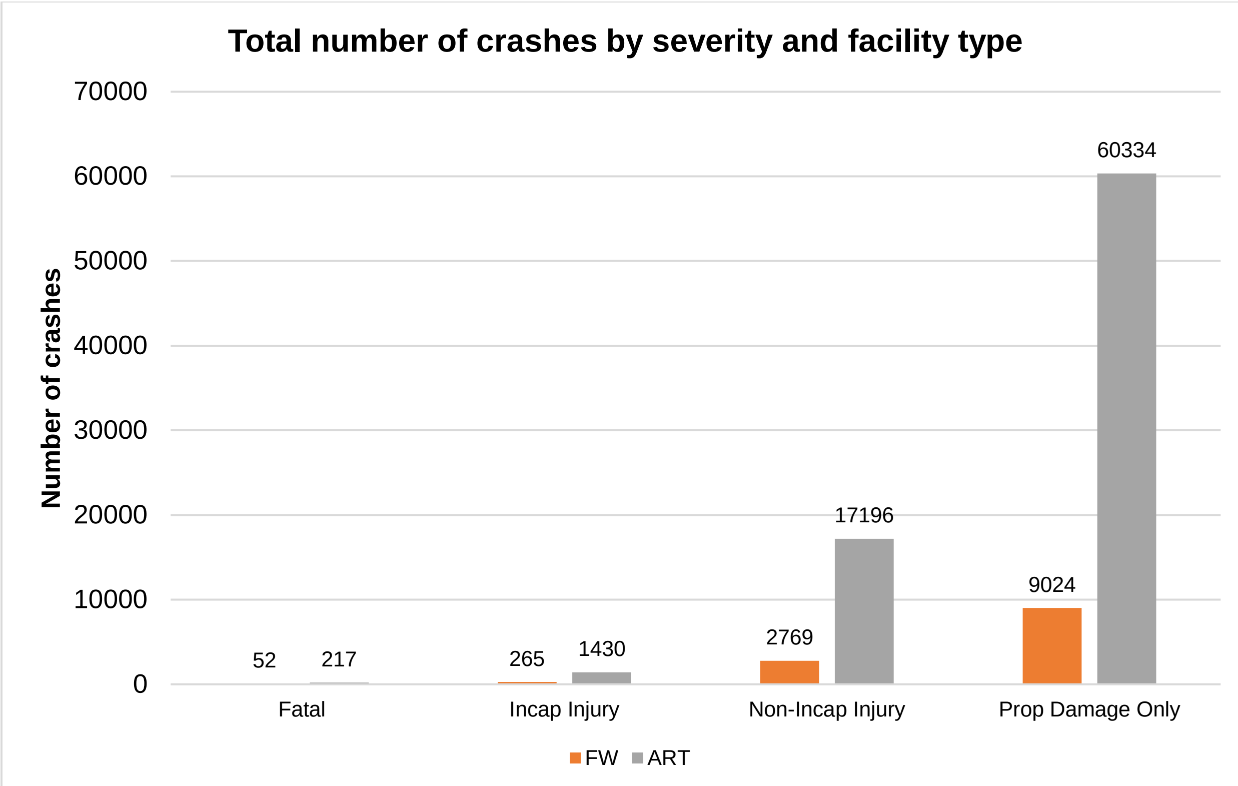


Figure C-25 Total number of crashes by severity and facility type.

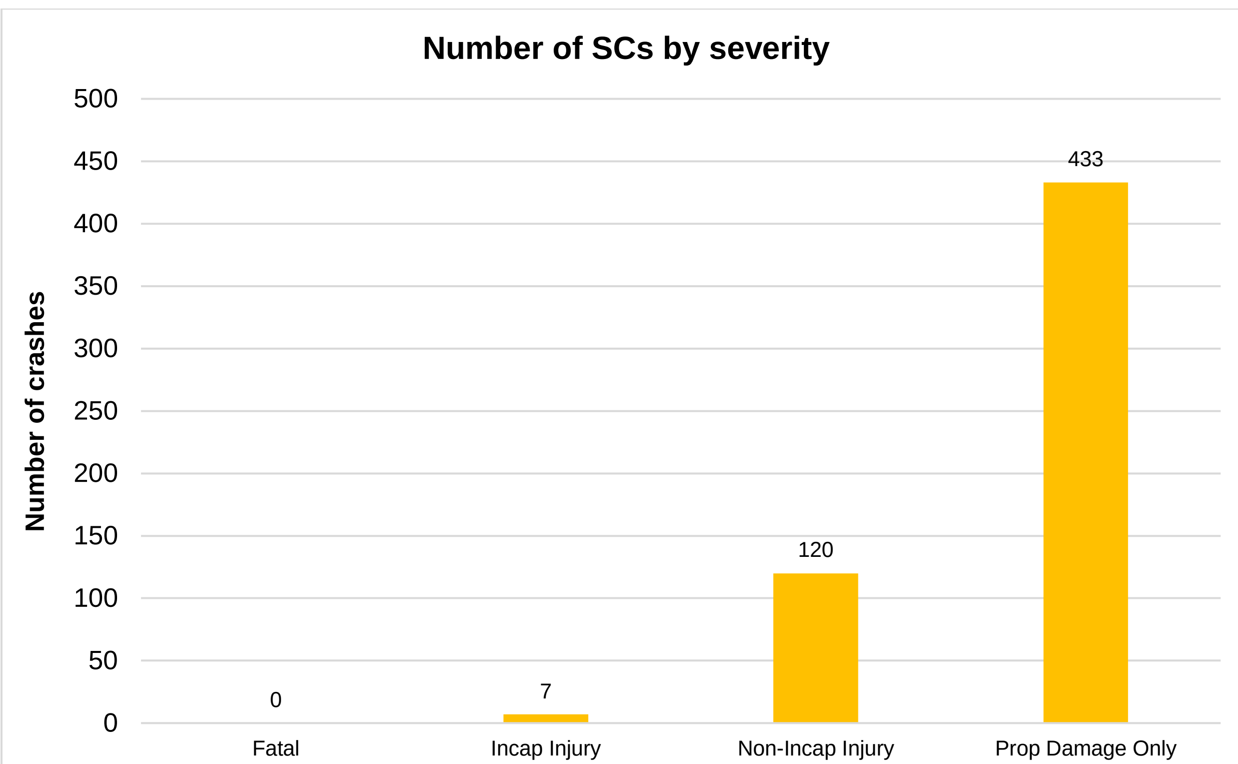


Figure C-26 Number of SCs by severity.

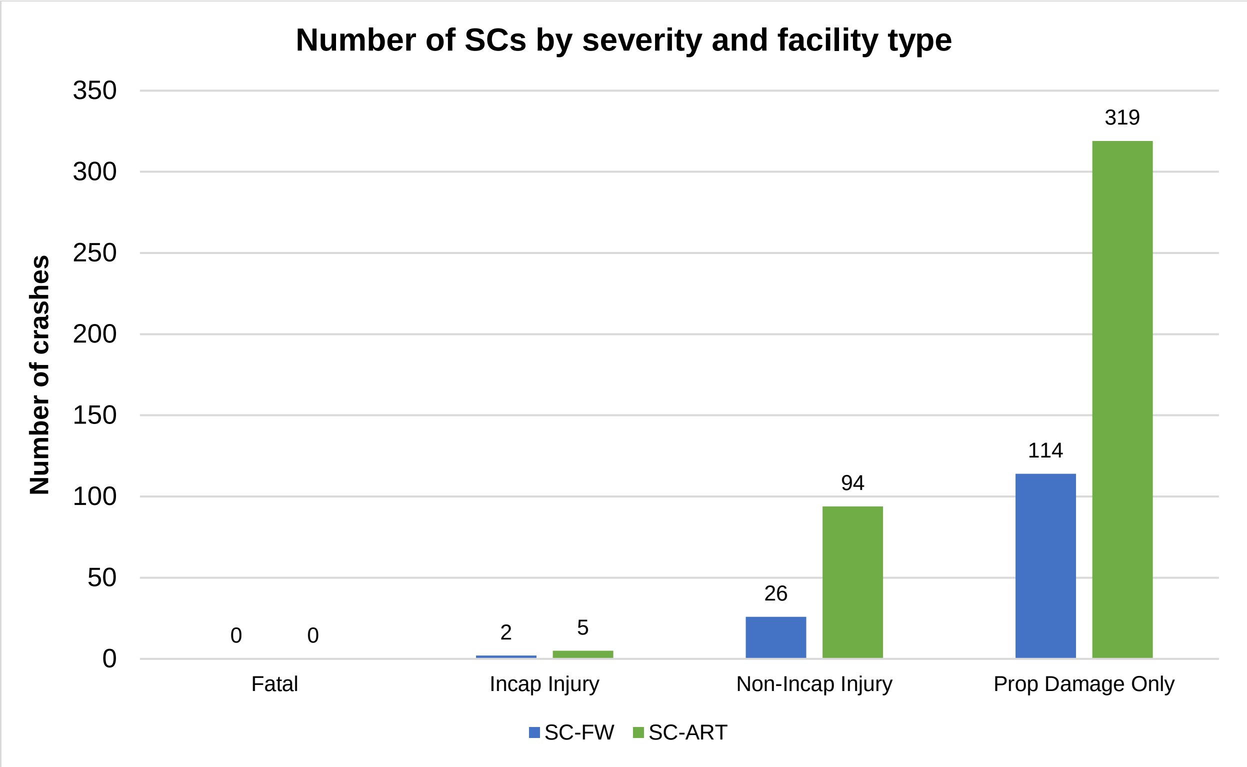


Figure C-27 Number of SCs by severity and facility type.

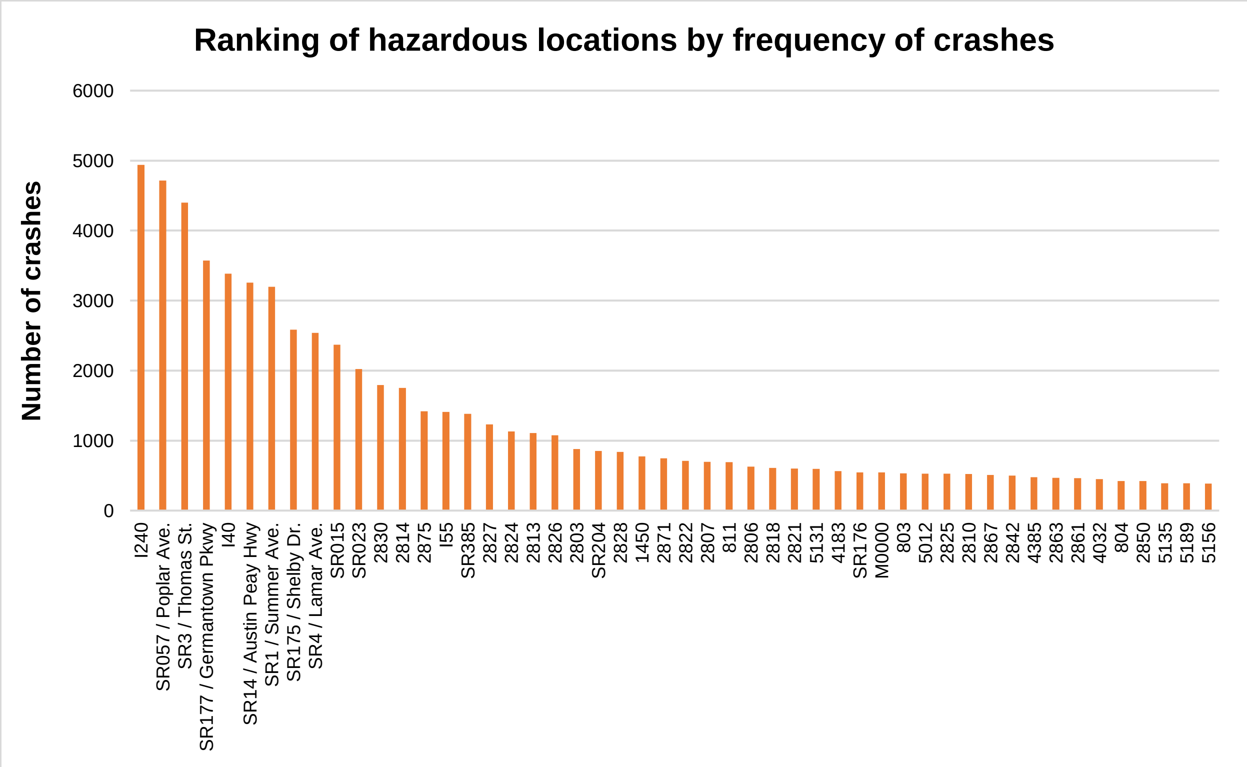


Figure C-28 Ranking of hazardous locations by frequency of crashes (top 50).

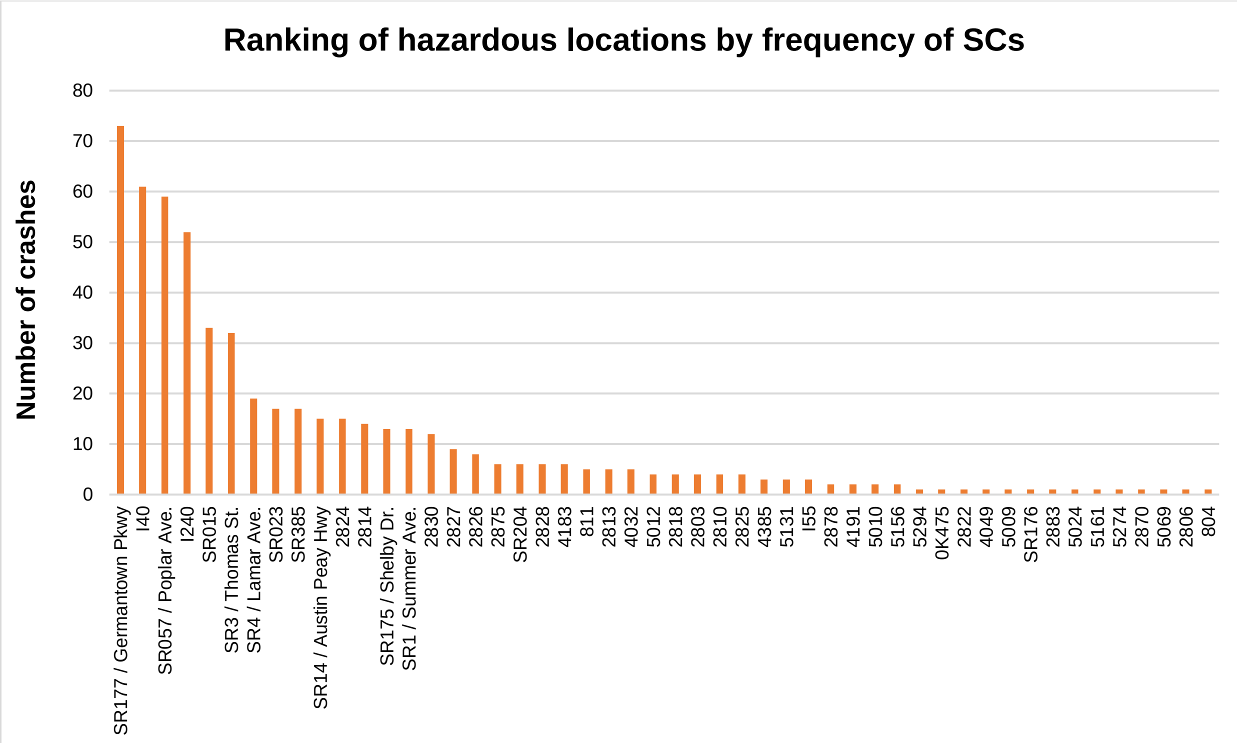


Figure C-29 Ranking of hazardous locations by frequency of SCs (top 50).

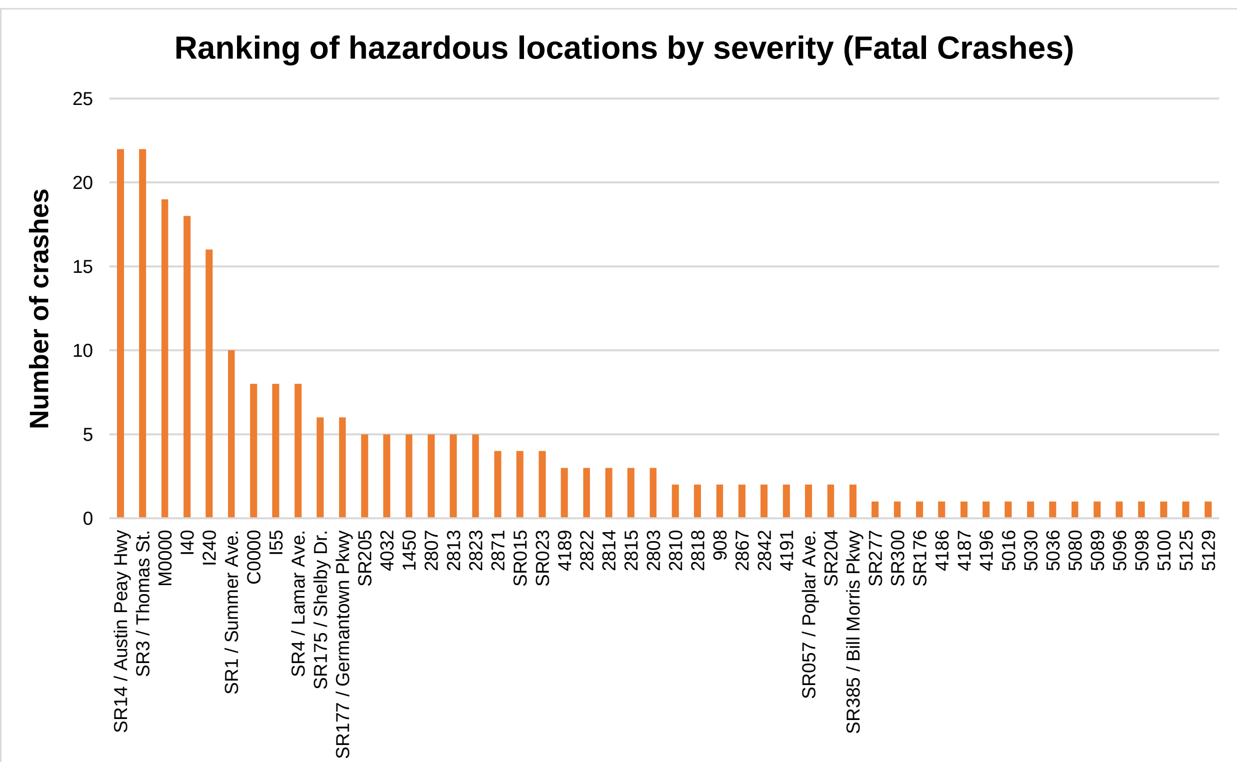


Figure C-30 Ranking of hazardous locations by severity - fatal crashes (top 50).

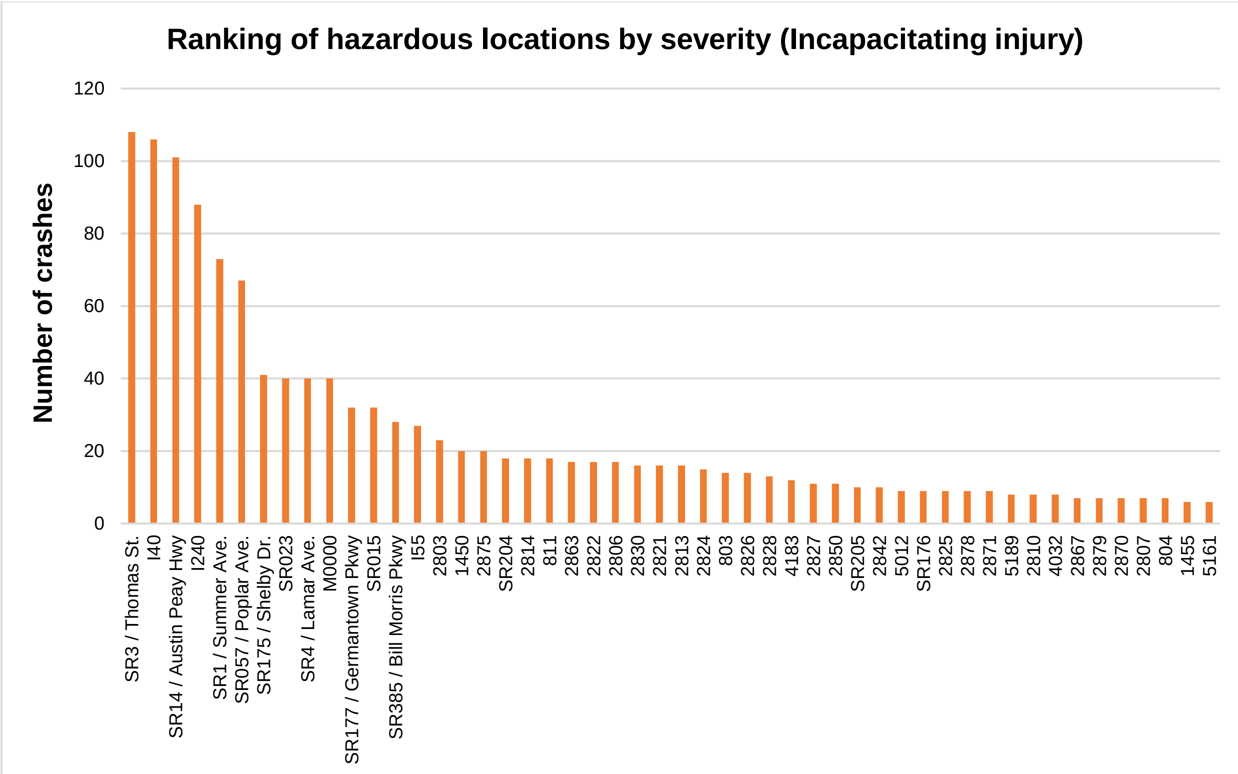


Figure C-31 Ranking of hazardous locations by severity- incapacitating injury (top 50).