Incorporating Time-Dependent Data for Proactive Safety Management

**Raul Pineda-Mendez, Qiming Guo,
Noshin Ahmad, Mario A. Romero, Andrew P. Tarko**

AUTHORS

Raul Pineda-Mendez
Qiming Guo
Noshin Ahmad
Graduate Research Assistants
Lyles School of Civil Engineering
Purdue University

Mario A. Romero, PhD
Research Scientist and Program Manager for the Center for Road Safety
Purdue University

Andrew P. Tarko, PhD
Professor of Civil Engineering and Director of Center for Road Safety
Lyles School of Civil Engineering
Purdue University
(765) 494-5027
tarko@purdue.edu
Corresponding Author

JOINT TRANSPORTATION RESEARCH PROGRAM

NOTICE

COVER IMAGE

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA/IN/JTRP-2024/01	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Incorporating Time-Dependent Data to Proactive Safety Management		**5. Report Date** December 2023	
		6. Performing Organization Code	
7. Author(s) Raul Pineda-Mendez, Qiming Guo, Noshin Ahmad, Mario A. Romero, and Andrew P. Tarko		**8. Performing Organization Report No.** FHWA/IN/JTRP-2024/01	
9. Performing Organization Name and Address Joint Transportation Research Program Hall for Discovery and Learning Research (DLR), Suite 204 207 S. Martin Jischke Drive West Lafayette, IN 47907		**10. Work Unit No.**	
		11. Contract or Grant No. SPR-4540	
12. Sponsoring Agency Name and Address Indiana Department of Transportation (SPR) State Office Building 100 North Senate Avenue Indianapolis, IN 46204		**13. Type of Report and Period Covered** Final Report	
		14. Sponsoring Agency Code	
15. Supplementary Notes Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.			

16. Abstract

This study proposed a risk-based safety management framework to supplement the current crash-based safety management system. The proposed tool considers time-dependent factors (e.g., hourly traffic, speed features, weather conditions, signal controls) to help justify operational measures for safety improvements (e.g., variable message signs, variable speed limits, warnings). These selected temporal factors subsequently were included in the developed sequential logit models; and those models, applied hour by hour, were then used to estimate the crash probability and severity level. Two typical roadway elements, rural freeway segments and signalized intersections, were also included in the analysis. The obtained crash risk profiles can be used to predict the expected number of crashes in periods when the operational safety countermeasures are expected to be active based on certain triggering conditions (e.g., traffic, weather, nighttime). These results, together with crash modification factors, may be used in the benefit and cost analysis process to justify the application of specific countermeasures.

17. Key Words safety management, operational countermeasures, temporal conditions, crash probability, risk profile, logit regression		**18. Distribution Statement** No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.	
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 62 including appendices	**22. Price**

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized

EXECUTIVE SUMMARY

Introduction

Traditional safety management uses crash data aggregated in multi-year periods to evaluate safety performance at segments and intersections. The candidate roads for treatment and proposed countermeasures for application currently are selected based on crash statistics and crash modification factors (CMFs). The expected safety performance is currently estimated as a function of the road's average annual daily traffic (AADT) and major geometric features. The high-resolution data sets now available make it possible to perform disaggregate safety analysis. Specifically, statistical models can be developed to capture the effect of short-term conditions on crash risk within a certain time period (e.g., 1 hour). This approach advances the quality of safety prediction by considering previously excluded time-dependent conditions (e.g., speed, weather, and traffic control) and their interactions with the geometric features of a road. This study established a time-dependent proactive safety management framework for this purpose. The risk-based safety management tools proposed and developed in this study can be used to supplement an existing crash count-based safety management system (SMS).

Findings

There were three main outcomes from this study.

1. Achieving a feasible and accurate time-dependent crash risk estimation framework. Although some time-dependent factors (e.g., precipitation and segment speeds) were identified as having effects on roadway safety performance in past research, this study established comprehensive model sets for two typical roadway elements (rural freeway segments and signalized intersections). The proposed models took advantage of all the available time-dependent data; and the risk profiles they produced show promising changing trends compared to the observed crash occurrence.
2. Identifying significant time-dependent and fixed condition factors. Several time-dependent factors, such as hourly volume (traffic), ice (weather), standard deviation of segment speed (speed), and coordination offset (traffic control), were found to affect crash probability and conditional severe crash probability, thus validating the idea of using disaggregated

data in safety management. The fixed condition factors, mainly the geometry settings, were identified as significant in the models as well.
3. Creating an example prototype tool that illustrated the feasibility of the new approach and facilitated its implementation into existing safety management practices. This tool applied user-decided triggering conditions (e.g., traffic and weather) that activated the intended operational countermeasure (e.g., reduced speed limit and warning). It then estimated the expected number and severity of crashes, which can be used with CMFs to support the benefit-cost analysis process. The obtained time-dependent crash risk models and the safety benefits of several time-dependent safety countermeasures, such as variable speed limits and optimized coordination time plans, were tested and verified in this study.

Implementation

Implementation of the proposed risk-based safety management framework included three components.

1. Enquiring and accessing data sources. There are several data sources outside the Indiana Department of Transportation (INDOT) SMS database, such as hourly weather conditions, INRIX speed data, and, for signalized intersections, detector/phase data maintained by the Traffic Management Center.
2. Processing data. The raw data should be pre-processed, which includes formatting, filtering, and imputing, and then processing with the risk models to estimate the risk.
3. Applying the processed data in the current SMS supplemented with risk-based elements.
   a. The *crash-based* SMS tool, the Safety Needs Identification Program (SNIP), identified candidate road segments and intersections.
   b. The *risk-based* SMS identified conditions that temporarily cause elevated crash risk. These conditions may be used as triggers of operational interventions.
   c. The provided prototype *risk-based* tool estimated the number of crashes on roads in periods with potential operational interventions where such interventions have yet to be applied.
   d. The *crash-based* SMS tool, the Roadway Hazard Analysis Tool (RoadHAT), performed a benefit/cost analysis of the considered operational countermeasures to justify their implementation.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

State and local transportation agencies supervise the performance of roads in their jurisdictions. The Highway Safety Manual (HSM) published by the American Association of State Highway Transportation Officials (AASHTO), is among the most widely used road safety publications and provides knowledge and analytical tools for monitoring, evaluating, and improving the safety performance of roads (AASHTO, 2010). The Indiana Department of Transportation (INDOT) followed the principles described in the HSM to develop and implement its statewide safety management system (SMS) (Tarko et al., 2019).

1.1 Indiana's Safety Management System

The primary objective of the Indiana SMS, as it is for other states as well, is reducing the number and severity of road traffic crashes. This objective is achieved by identifying high-crash roads, determining causes, considering alternative countermeasures, selecting the most promising ones for implementation, and then evaluating a solution's effectiveness after its implementation. A successful SMS requires adequate communication and coordination across the various engineering and decision-making units involved in traffic safety, improving analytical tools for problem identification, and providing decision-support tools for policymakers and managers to enable the effective use of the resources allocated to safety management. The implemented tools that support INDOT's SMS decision-making process include the Safety Needs Identification Program (SNIP), the Roadway Hazard Analysis Tool (RoadHAT), the Crash/Conflict Diagram Builder (CDB), and the Hazard Elimination Program (HEP).

HEP is a crucial element of INDOT's SMS. HEP focuses on road improvements and provides analytical tools for identifying safety problems and their solutions. HEP's objectives are identifying high-crash locations, conducting safety reviews to find the causes of crashes and corresponding road deficiencies, suggesting appropriate countermeasures, grouping countermeasures to form projects, determining the economic feasibility of projects, and conducting evaluations of the implemented safety projects to provide feedback into the program. The core components of the HEP process are shown in Figure 1.1 (Tarko & Kanodia, 2004).

Identifying high-crash locations (HCLs): The first step of the HEP process involves selecting sites with safety problems from a pool of thousands of candidates, a small amount of which is typically available at the road network level. Such data include the type of location (e.g., signalized intersection, rural freeway segment), basic geometric characteristics, traffic volume, and crash records. A priority list of sites for further investigation is prepared using evidence-based criteria.

Identifying HCLs involves evaluating thousands of candidates with safety problems, which then are
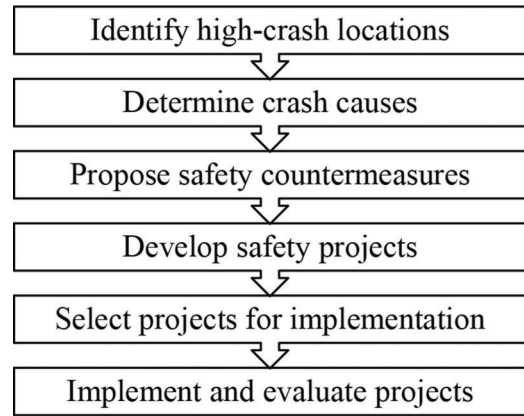
Figure 1.1 Core components of INDOT's Hazard Elimination Program.

included on a priority list of sites. A systematic process for this purpose was developed by INDOT and the Center for Road Safety (CRS) at Purdue University. The first step of this process consists of classifying the location into one of the current road categories (e.g., rural freeway segments, state-local intersections, and ramps). Following the classification, the required crash, traffic, and roadway data are collected. Then, a set of locally calibrated safety performance functions (SPFs) are used to estimate the typical crash frequency at various injury severity levels, the results of which are used to evaluate two safety performance measures: the index of crash frequency (ICF) and the index of crash cost (ICC). Equation 1.1 and Equation 1.2 show a simplified form of these two safety indexes. Full versions and example calculations of the ICF and ICC formulas are available in the Indiana SMS guidelines (Tarko et al., 2016). A priority list of sites that need improvement ultimately is prepared from this evaluation.

$$I_{CC} = \frac{Observed\ Crash\ Cost - Expected\ Crash\ Cost}{\sqrt{\begin{array}{l} Observed\ Crash\ Cost\ Variance \\ + Expected\ Crash\ Cost\ Variance \end{array}}} \quad \text{(Eq. 1.1)}$$

$$I_{CF} = \frac{\begin{array}{l} Observed\ Crash\ Frequency \\ - Expected\ Crash\ Frequency \end{array}}{\sqrt{\begin{array}{l} Observed\ Crash\ Frequency\ Variance \\ + Expected\ Crash\ Frequency\ Variance \end{array}}} \quad \text{(Eq. 1.2)}$$

Determining causes: After selecting the high-crash locations, safety reviews are conducted to determine the potential causes of the high-crash risk at these locations. Historical crash records are analyzed to identify the predominant crash patterns and to determine their probable causes. On-site visits, including safety checks and engineering studies, then are conducted to confirm potential causes and to identify additional safety deficiencies.

Determining countermeasures: After the safety review is conducted, the next step is to prepare a list of

potential crash causes and a set of safety countermeasures addressing these causes. For this purpose, the Federal Highway Administration (FHWA) maintains a repository of proven safety countermeasures (FHWA, n.d.). In addition, INDOT's safety management guidelines include an inventory of safety countermeasures to tackle specific crash causes (Tarko et al., 2016).

Developing safety projects: Various potential safety countermeasures then are grouped to form safety projects that are expected to be economically and technically feasible and effective in improving safety performance.

Selecting projects for implementation: The next step includes selecting safety projects based on the priority ranking obtained in the economic evaluation. Restricted budgets demand achieving the most significant overall safety benefits from the choices made. From potential projects for multiple HCLs, the selected projects are expected to be highly cost-effective while maximizing the overall safety benefit.

Implementing and evaluating projects: The effectiveness of an implemented safety project in reducing the number and severity of crashes is considered in the final step. These results can be used to update the CMFs and to determine whether the safety project produced a statistically significant reduction in the number of crashes at various injury severity levels.

The identification and ranking of HCLs is critical for the effectiveness of a SMS. If the locations with serious safety problems are not identified during the identification phase, they are not considered again in the current HEP cycle. To use resources efficiently, only the most promising HCLs should be selected for safety reviews.

1.2 A Concept of Risk-Based Safety Management

As discussed in the last section, a crash-based SMS methodically searches for HCLs to identify hazards. For this purpose, it uses crash count models. These models, named safety performance functions (SPFs), are the primary analytical component in current safety engineering practice. Crash count models estimate the expected numbers of crashes at specific injury severity levels for each road of an analyzed road network. A typical count-based crash model includes the annual average daily traffic (AADT) and the major geometric features of each road, such as the number of lanes, number of legs, traffic signal presence, and functional classification, among others (AASHTO, 2010). The expected annual crash frequencies at various injury severity levels are then compared to the corresponding number of reported crashes, usually in periods of 3 to 5 years. An excessive number of crashes above the expected number and the corresponding excessive cost of crashes thus points out a potential safety problem, triggering additional investigation (a road safety audit) to identify the needed road or control improvements.

Past researchers identified a potentially strong hour-to-hour variation of crash risk (Dutta & Fontaine, 2019; Mohammadnazar et al., 2021; Yuan et al., 2021).

The time-related factors confirmed included weather conditions (Naik et al., 2016), operating speed (Fitzpatrick et al., 2017), presence of police enforcement (Tarko, 2009), road closures, and temporal changes in geometry due to road construction (Chen & Tarko, 2012), among others. These studies ultimately uncovered the major weakness of the crash count models, namely, providing crash count estimates only in periods of 3 to 5 years. This inability to grasp the short-term temporal variability of crash risk precludes consideration of operational countermeasures (e.g., variable speed limits or warning messages) about dangerous driving conditions ahead, such as icy pavements or slow-moving traffic. However, this omission in the current safety management process can be addressed thanks to the increased availability of high-resolution data and improved information technology.

In summary, short-term risk-based safety management components should be added to the current crash-based management practices to identify locations and periods when operational countermeasures might be justified to reduce the crash risk. Comprehensive safety management should use both time-independent geometric features and time-dependent hourly volumes, operating speeds, and weather conditions. The potential safety benefits could include better insight into the crash risk and preventing crashes by real-time monitoring of traffic and environmental conditions in order to proactively apply operational countermeasures if such needs are detected.

The emergence of connected and autonomous vehicles (CAVs) further justifies evaluating crash risk in short intervals and responding with mitigating measures if needed. The increased presence of CAVs is anticipated to significantly affect many aspects of our transportation systems, including traffic safety, user accessibility, and urban planning (Fagnant & Kockelman, 2015; Meyer et al., 2017; Talebpour & Mahmassani, 2016). Once CAVs are fully implemented, they will collect a vast amount of travel and safety-related information that should eventually become available for engineering use. Furthermore, these vehicles would comply and adequately respond to locally broadcast messages. An additional potential side effect of CAVs could be their positive influence on the human drivers around them.

Creating analytical tools that capture and process data into useful information is imperative. For traffic safety management in particular, the crash count models currently used for safety management should be supplemented with risk-based models capable of predicting crash probability in short intervals. Implementing such tools will allow engineers to monitor the crash risk to maintain an acceptable level with operational means such as variable speed limits, advanced signal warnings, in-vehicle messages, safety-based rerouting, and modification of the navigation strategies of CAVs.

This risk-based approach to SMS can complement and supplement current crash-based SMSs. Crash-

based monitoring of roads can identify HCLs based on aggregate data in long periods, while risk-based SMS can provide a more insightful picture of temporal risk fluctuation at individual locations. Improving safety at such locations could be accomplished by applying operational measures. For example, an analytical tool could be used to identify the hours during which an elevated risk of crash exists and the conditions that cause this high risk. These events are the result of specific combinations of geometry and time-dependent factors; therefore, risk estimated in short intervals may enrich the body of knowledge about interactions between the geometric elements of roads, such as curves or ramp intersections, and temporal conditions such as weather. These insights may lead to better geometric design and improvement decisions.

On the other hand, risk-based SMS also can improve the identification and performance evaluation of safety countermeasures. Risk-based SMS identifies conditions that cause temporarily elevated crash risk that can be used as triggers for conducting operational interventions. Estimation of the number of crashes on roads during periods when potential operational interventions are active can be accomplished by determining the volume of crash probabilities in periods when these interventions are activated. Finally, the crash-based SMS tools can be applied to perform a benefit/cost analysis of considered operational countermeasures to justify their implementation.

1.3 Research Scope

This project determined the data needed and the research method utilized to obtain risk models for two selected example types of roads: signalized intersections and rural freeway segments. The first objective of this project was to enable the use of the available time-dependent data sources to support safety management by identifying and justifying investments that include traffic control, advisory information, and other operational countermeasures. To address this need, an hourly crash risk equation was developed that includes relevant time-related conditions applicable to a large-scale system. The second objective was achieved by integrating the proposed risk-based SMS with the existing crash-based system. This need was addressed by developing a prototype tool that supports time-dependent analysis, which can become a component of the integrated safety management process. This work was conducted via a pilot study that demonstrates the method for selecting two types of roads: rural freeway segments and signalized intersections. The concept of risk-based SMS was developed to apply these equations within the existing framework of decision-making, including selecting high-risk conditions, identifying countermeasures, and evaluating their economic feasibility.

This study is a continuation and expansion of SPR-4302, which identified promising emerging data sources of satisfactory quality (Tarko et al., 2021). The current project refined the crash risk equations developed in SPR-4302 for two road categories to make them applicable to the Indiana road network level and connects the crash risk with the static and temporary factors represented by available time-dependent data.

The remainder of this report is organized as follows. Chapter 2 presents the implementation of risk-based SMS for rural interstate freeway segments and includes a summary of the data description, methodology, sample characterization, results, and implications. Chapter 3 illustrates the similar implementation of risk-based SMS for signalized intersections. Chapter 4 focuses on implementation of a novel tool for risk-based SMS, explaining the concepts of data processing and showing various illustrative examples of using the tool. Finally, Chapter 5 summarizes the work conducted, emphasizing its implications to safety management practices in Indiana, and describes the additional research needed to move forward with implementation.

1.4 Literature Review

The variation of crash risk across space, time, and even data aggregation levels was identified (Dutta & Fontaine, 2019; Mohammadnazar et al., 2021; Yuan et al., 2021). From their findings, the idea of risk-based safety management evolved, which uses both static geometric features and time-dependent factors to evaluate the change in crash risk over time. In addition, as high-resolution data collection techniques advanced in the last two decades, time-dependent data have become available in the last two decades. Many researchers have estimated the probability and severity of crashes at a disaggregate level and an abundance of so-called real-time crash prediction models have been introduced (Ahmed & Abdel-Aty, 2011; Lee et al., 2003; Oh et al., 2001; Xu et al., 2012; Yu et al., 2020).

Meanwhile, the methods to identify and assess crash risk factors have evolved. For instance, a Bayesian model implemented by a probabilistic neural network was used to identify hazardous traffic conditions, where the standard deviation of 5-minute speeds was associated with higher crash risk (Oh et al., 2001, 2005). Other studies applied case-control logistic regression models to link real-time traffic flow characteristics to crash probability under different scenarios and collision types (Abdel-Aty et al., 2004; Pande & Abdel-Aty, 2006; Zheng et al., 2010). Another common approach involves taking high-risk conditions as a binary classification problem. Multiple recent studies applied more machine learning techniques, such as support vector machine (Wang et al., 2019; Xiao & Liu, 2012), back propagation neural network (Cheng et al., 2010), extreme machine learning (Li et al., 2017), convolutional neural network (Yu et al., 2020), and deep learning (Huang et al., 2020).

The above methods have shown promise in capturing crash risk changes over disaggregated time scales. Nevertheless, there are four well-known limitations when applying these models in a large-scale highway

network, thereby preventing their use in risk-based safety management. First, dependence on the heavily instrumented road segments (Hossain et al., 2019), which are the primary data source for real-time crash prediction models. Probe vehicle data such as INRIX may be a more promising data source for systematic safety management since it covers all the U.S. interstate freeways and most state highways (Sharma et al., 2017). Second, the relationship between crash injury severity and the time-dependent factors is under-investigated. The lack of research is due to the rarity of severe crashes on instrumented road segments. In traditional count-based analysis, the factors that could lead to severe crashes differ significantly from those affecting crash frequency, so the safety effects of time-dependent factors are expected to differ at multiple injury severity levels. Third, most of these models focused on instantaneous traffic dynamics like traffic volume and speed but failed to include weather conditions. Considering the proven significant impacts of weather conditions on driving behaviors and traffic safety (Yu & Abdel-Aty, 2014), these factors should be included in crash risk prediction models to avoid a biased estimation of the speed or volume effects. Lastly, there is a direct connection between crash occurrences and changes in traffic dynamics. Therefore, carefully defined offsets for the pre-crash period are needed to avoid the potential endogeneity issues caused by inaccurate recorded crash times.

Past researchers demonstrated the potential of estimating crash risk in the near-term using safety-related data collected over short periods. Several authors linked the short-term crash risk with traffic characteristics, including traffic volume, operating speed, average speed, speed variation, and speed difference between consecutive road segments (Roshandel et al., 2015; Tarko et al., 2019). Some researchers also were able to estimate the safety effect of weather characteristics (Ahmed et al., 2012).

The relationship between speed, weather, traffic, and pavement conditions and the probability of crash was investigated in the JTRP research project SPR-4302, *Use of Emerging and Extraordinary Data Sources as Means to Improve Traffic Safety* (Tarko et al., 2021). Such relationships will help estimate the short-term risk of crash based on temporary conditions. Aggregating the risk over more extended periods can yield the expected number of crashes proactively and can reveal temporary factors occurring on the road that contribute to crashes. Within a new paradigm of safety management, high numbers of crashes on roads can be explained by frequent periods experiencing high-risk conditions.

2. RURAL FREEWAY SEGMENTS

A case study of implementing a risk-based SMS on Indiana's rural freeways is presented in this chapter. The connection between time-dependent factors, crash probability, and injury severity is assessed for a sample of freeway segments with available data. The estimated impacts, including static and time-dependent elements, are meant to be the foundation of analytical tools for a system-wide risk-based SMS. This chapter focuses only on the development of the crash probability and injury severity models. Chapter 4 describes their application in specific safety management tasks.

2.1 Data Description

Multiple factors are known to affect traffic safety performance. Some examples include driver characteristics, vehicle features and operation, and roadway attributes. In addition, time-dependent factors that reflect certain environmental conditions also should be considered in risk-based SMS (e.g., congestion, adverse weather, poor lighting, and seasonal behavior such as summer travel).

This section discusses the available data regarding fixed and time-dependent CRFs to enhance the decision-making process behind long-term infrastructure safety improvements. Special attention is placed on time-dependent factors previously omitted in safety management. This section is focused on data preparation for statistical analysis and risk-based SMS. A more in-depth description of the data sources can be found in the final report for JTRP SPR-4302 (Tarko et al., 2021).

Data from five types of CRFs were collected. Crash records were extracted from the Automated Reporting Information Exchange System (ARIES). The operating travel speeds at the segment level were assembled from the National Performance Management Research Data Set (NPMRDS). Hourly traffic volumes and vehicle classifications were obtained from INDOT's Traffic Count Database System (TCDS). Weather conditions were accessed via the Indiana State Climate Office (INClimate) at Purdue University. Roadway characteristics were obtained from INDOT's Road Network Data (RND) supplemented with Google Earth's historical imagery.

Data from other CRFs were also pursued and their potential benefits were described. However, these data sources are not ready for systemwide implementation. Examples of such data include police citations and road construction work orders.

2.1.1 Crash Records

Police crash reports were accessed from the ARIES database, which includes detailed information about the crash, road site, involved people, and vehicles. The geographical coordinates provided were used to assign individual crashes to segments for analysis. The reported crash time was used as well to determine the time-dependent factors before the stated crash hour. In total, there were 2,091 crashes assigned to the selected roads for 5 years (2014 through 2018).

Based on the crash injury severity scale, 85.8% of the crashes were classified as property damage only (PDO), 6.6% as a potential or minor injury, and 7.6% as fatal or incapacitating.

2.1.2 Speed Data

Operating travel speeds were obtained from NPMRDS. Established in 2013 and later updated in 2017, NPMRDS condenses the travel times along all the road segments in the National Highway System (NHS). The NPMRDS segments are systematically defined between two consecutive ramps. The travel times for all vehicles, passenger cars, and trucks are reported separately every 5 minutes. This study used the travel times for all the vehicles to calculate the 5-minute speeds, which then were aggregated to one hour, and multiple statistics then were created to characterize the distribution of speed within each hour for each segment. The final set of speed characteristics for statistical analysis included the average hourly speed, the temporal standard deviation of speed, and the speed trend.

2.1.3 Traffic Counts

There are 70 permanent traffic count stations on Indiana's interstate freeways. All the selected segments have nearby stations. However, to implement a systematic risk-based safety evaluation, short-term traffic volume prediction models were developed for different types of highways. These models estimate the hourly volumes using system-wide predictors, such as the NPMRDS's probe density, operating speed characteristics, weather conditions, roadway geometry, and time indicators. Additional modeling details and an assessment of the results of this instrumental model are available in Tarko et al., 2021. These models allow for predicting hourly volumes on selected roads without counting stations. An effort to estimate hourly traffic volumes on low-classified roads was also part of this study, and the results are presented in Appendix A.

2.1.4 Weather Data

INClimate archives weather observations recorded throughout Indiana and its neighboring states. Two datasets, Parameter-elevation Relationships on Independent Slopes Model (PRISM) and Multi-sensor Precipitation Estimates (MPE), are provided by INClimate. PRISM is a gridded data set with virtual weather stations evenly spaced every 2.5 miles that contains information about the distribution of daily ground temperatures. These data are previously spatially interpolated with observed temperatures using statistical parameterization algorithms accounting for terrain influences. The MPE data set provides hourly liquid precipitation amounts. Like PRISM, MPE has virtual stations evenly spaced every 2.5 miles. MPE's data combines observational precipitation data with derived estimates from the Doppler radar network and satellites to offer a gridded characterization of precipitation. This network of high-density gridded weather stations permits deriving hourly weather conditions on selected segments. From the 5,724 virtual gridded weather stations in Indiana, hourly precipitation and temperature data are obtained using interpolation. Details about the interpolation process are presented in a previous publication (Tarko et al., 2021). The weather data are collected with ground stations, satellites, and aerial radar and then extrapolated and integrated accounting for elevation, topography, rain shadows, temperature inversions, and coastal effects. The available gridded data include hourly precipitation and temperature daily statistics with additional pieces of information possible.

2.1.5 Road Inventory

The existing aggregate geometry and traffic data from INDOT's road inventory data sets were supplemented with other road characteristics gathered from Google Earth's historical imagery. The collected data included cross-sectional elements, horizontal alignment, pavement, roadside elements, signage, and road lighting. A comparison of the available data in the sample and at the system level was made for the system-wide implementation of risk-based SMS. Among the CRFs available at the system level, the factors that were identified as needing more data collection included the presence of artificial lighting, type of road surface (asphalt or concrete), and offset distance to a barrier. While other variables are available in INDOT's inventory, additional preparation is required before they are suitable for estimating crash risk. These variables include overpass presence, barrier type and location in relation to the road, outside shoulder width, number of barrier-ends facing the travel direction, horizontal curvature, presence, and type of ramps, and posted speed limit.

2.1.6 Additional Data Preparation

Information from various data sources was linked to produce a data set used for statistical analysis. Once the data linking process was complete, the resulting modeling data set was formed by two types of observations: crashes and non-crashes. A 1:30 ratio between the two types of records was enforced via sampling non-crashes based on the number of crashes. Since crash observations are uncommon compared to non-crashes, this sampling was needed to estimate the effects of the contributing factors on the hourly crash risk and injury severity. The offset of the fitted model was later adjusted to represent the original conditions before the sampling.

The different data sources were linked and integrated as the input for the risk-based SMS tool. Two data sets were created: (1) time-independent data, such as road

characteristics, and (2) time-dependent data, such as weather information and speed. The time-dependent data was preprocessed to provide the information in a one-hour resolution. If needed, this resolution can be adjusted to 1 week, month, and year. However, the larger the aggregation level used, the closer the model becomes count-based rather than risk-based.

2.2 Analysis Methods

Three models were used to estimate the safety effects of the time-dependent and fixed predictors (sequential binary logit, multilevel logit, and mixed logit), all of which are variations of standard logistic regression. A brief description of each model is provided in this section as well as a discussion of their advantages and limitations. Lastly, a comparison strategy to balance the performance of all three types of models is introduced.

Due to the rarity of crashes when measured in one-hour periods, it is more reasonable to model the probability of a crash as a binary response (i.e., whether the crash happened or not). Therefore, when there are N (N > 1) crashes within one hour on the same road segment, only the first crash was used for analysis. It was assumed that the following crashes might be partially, if not totally, influenced by the first incident. This rare case accounts for less than 0.5% of observations and is outside of the scope of this study.

2.2.1 Sequential Binary Logit

The sequential binary logit model is a helpful regression approach for discrete dependent variables. For dichotomous (i.e., having two possible values) outcomes, this method estimates the effect of independent predictors on the probability of a safety event. In our case, an initial model estimates the probability of the hourly crash occurrence, while a second model estimates the effects on severe injury given crash occurrence. The logit function is the LN of the odds ratio and is represented as a function of the independent predictions in the following form.

$$Y_i = logit(P_i) = \ln\left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \sum_{k=1}^{K} \beta_k X_k \quad \text{(Eq. 2.1)}$$

$$P_i = \frac{\exp(\beta_0 + \sum_{k=1}^{K} \beta_{k,i} X_{k,i})}{1 + \exp(\beta_0 + \sum_{k=1}^{K} \beta_{k,i} X_{k,i}))} \quad \text{(Eq. 2.2)}$$

where β_0 is the intercept constant and the $\beta_1, \ldots, \beta_K$ are the unknown parameters corresponding with the explanatory variables ($X_k = 1, \ldots, X_k$). These unknown parameters are usually calculated using maximum likelihood methods. Once estimated, these parameters can be used to calculate the probability that the outcome takes the value of 1 as a function of covariates.

2.2.2 Multilevel Logit

Multilevel models are used where data are hierarchically organized. In this study, a two-level organizational model is proposed. The first level represents the safety effects of the time-dependent predictors, while the second upper level focuses on the effects of the roadway characteristics. The random effects, estimated as the variance components, are the model parameters believed to vary between the higher-level units. In contrast, the fixed effects are the estimates that are modeled to not vary between the higher-level units.

2.2.3 Mixed Logit

The hourly crash probability at various injury severity levels is estimated as a function of the static roadway characteristics and time-dependent factors. To do so, a sequential binary mixed logit approach is used. This method fits two consecutive models: (1) a model of the hourly crash probability with a crash vs. no-crash binary response and (2), a model of the probability of severe crashes (injury or fatal) conditioned on crash occurrence. The latter model uses a binary severe outcome vs. not severe outcome.

The mixed logit (or random-parameters logit) model addresses several weaknesses of the traditional logit model by allowing the parameter values to vary across observations according to some pre-specified distribution. The conventional binary logit and mixed logit models are illustrated in Equation 2.3 and Equation 2.4.

$$P_n = \frac{\exp(X\beta)}{1 + \exp(X\beta)} \quad \text{(Eq. 2.3)}$$

$$P_n^{mixed} = \int P_n f(\beta|\varphi)d\beta \quad \text{(Eq. 2.4)}$$

where P is the probability of crash (or severe outcome if considering the injury severity), **X** are the contributing factors, and β are the estimated parameters, $f(\beta|\varphi)$ is the density function of β with φ referring to a vector of parameters of that density function (e.g., for normal distribution, $\varphi = (\mu, \sigma^2)$). The selection of significant variables is based on the overall goodness of fit of the model in terms of their AIC value and individual t-test statistics.

The SBML model was chosen over the ordered model, the most common modeling approach for an ordinal response such as crash injury severity. However, the main drawback of ordered models is that the parameter estimates, and significant explanatory variables are the same at all injury severity levels. Even though the parameter estimates of each explanatory variable could be different across different crash severity levels in a generalized ordered logit model, the set of significant explanatory variables is still assumed to be the same across different injury severity levels.

2.2.4 Comparison Strategy

The three models described previously were compared based on their log-likelihood. Specifically, the Akaike Information Criterion (AIC) and the Likelihood Ration Test (LRT) were used to compare the models. The lower the AIC, the better the use of the information and therefore the more preferred model. In terms of the LRT, two nested models, such as binary logit vs. multilevel logit and binary logit vs. mixed logit, were compared based on the ratio of their likelihoods, which followed a chi-square distribution with degrees of freedom equal to the number of additional parameters.

2.3 Empirical Setting

A sample of 133 one-way miles of rural freeways was selected for this study. The sample was comprised of approximately 5% of the total one-way mileage of rural interstates. These segments were chosen based on their data availability, mainly for observed short-term traffic volumes. The spatial distribution of the sample is presented in Figure 2.1. Out of the 133 miles of selected road segments, there were 25 miles on I-64, 23 miles on I-65, 39 miles on I-69, 40 miles on I-70, and 6 miles on I-74. There were 2,091 crashes assigned to the selected road sections between 2014 and 2018.

In past studies that used disaggregated traffic data, the analysis segment lengths were determined mainly by the spacing of the loop detectors. The chosen analysis segment lengths varied from 0.50 miles on I-880 in the San Francisco Bay area of California (Xu et al., 2013), 0.40 miles on an urban expressway in Shanghai (Yu et al., 2020), and 0.95 miles on I-235 in Des Moines, Iowa (Huang et al., 2020). In this current study, however, the selected road sections were divided into 532 segments with a fixed length of 0.25 miles. This segmentation was chosen for several reasons. First, this segmentation enabled the ability to reflect sporadic changes in roadway characteristics, which can be diminished with a larger segmentation. Second, on monotonous highways, it has been found that drivers can see as far as 0.28 miles under clear weather conditions. Lastly, this segmentation permits a quick transfer of the results to Indiana's current safety management practice.

2.4 Results and Discussion

Several factors were found to affect hourly crash risk and severity. In short, hourly crash risk is increased by hourly volume, AADT by vehicle type, horizontal curves, barriers, lower speed limits, the standard deviation of speed, and the interaction of low-intensity rain and freezing temperatures. Additional factors that increase short-term crash risk include auxiliary lanes, downtrend speeds, and congestion while overpassing roads, average speed, and uptrend speeds were found to enhance safety. The conditional probability of a severe
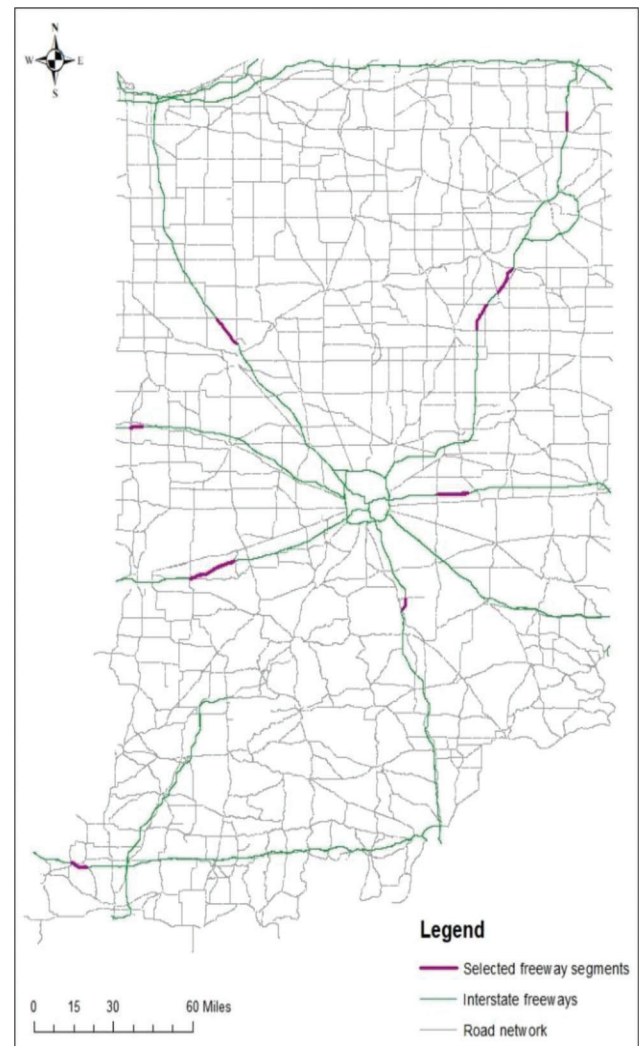
Figure 2.1 Sample road segments used in the crash risk analysis of rural freeways.

outcome is increased by mild curves, average speed, the standard deviation of speed, downtrend speeds, and intermediate congestion while lighting and lower speed limits were found to reduce crash severity.

2.4.1 Effects on the Crash Probability

Table 2.1 shows the parameter estimates, standard errors, and the AIC values of the hourly crash probability models. In addition to the random parameters (mixed logit) model, the fixed and random effects (multilevel logit) models were fitted to compare their performance. The random effects model produced the lowest AIC, followed by the random parameters, and fixed effects models. The random effects model's performance may have been due to its capacity to reflect the hierarchical data structure.

Hourly crash risk is increased by hourly volume, AADT by vehicle type, horizontal curves, barriers, lower speed limits, the standard deviation of speed, and the interaction of low-intensity rain and freezing

TABLE 2.1
Parameter estimates of the hourly probability of crash models

Effect	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Intercept	2.048	0.832	1.845	1.168	2.293	0.922
Hourly Traffic Volume (1,000 veh/h)	0.306	0.096	0.338	0.098	0.305	0.099
AADT Cars (1,000 veh/day)	0.009	0.005	–	–	0.009	0.005
AADT Trucks (1,000 veh/day)	0.056	0.009	0.075	0.018	0.060	0.010
Overpassing Road	-0.193	0.079	–	–	-0.210	0.082
Moderate Curve (5.5–13.9 degrees)	0.132	0.078	–	–	0.138	0.081
Sharp Curve (14 degrees or more)	0.232	0.074	–	–	0.249	0.077
Proportion with Median Cable Barrier	0.196	0.089	–	–	0.206	0.090
Proportion with Median Guardrail	-0.720	0.239	-1.035	0.444	-0.749	0.249
Proportion with Roadside Guardrail	0.713	0.090	0.657	0.228	0.740	0.092
Median Barrier Offset < 30 ft	-0.123	0.051	–	–	-0.122	0.053
Proportion with Concrete Pavement	0.512	0.085	0.465	0.203	0.515	0.092
Proportion with Entering Auxiliary Lane	0.559	0.154	–	–	0.545	0.173
Proportion with Exiting Auxiliary Lane	1.465	0.222	1.443	0.656	1.488	0.238
Speed Limit Reduced by 5 mph	0.304	0.087	–	–	0.306	0.092
Average Roadside Shoulder Width (ft)	0.118	0.034	0.140	0.078	0.122	0.037
Light Rain (precipitation < 0.098 in)	0.250	0.090	0.214	0.094	0.248	0.094
Freezing Temperature (T ≤ 32F°)	0.246	0.063	0.235	0.065	0.220	0.068
Ice Conditions	0.549	0.145	0.544	0.154	0.624	0.155
Average Hourly Travel Speed (mph)	-0.141	0.011	-0.146	0.012	-0.146	0.012
SD of Hourly Travel Speed (mph)	0.058	0.010	0.058	0.011	0.054	0.013
Hourly Speed Trend	-0.262	0.037	-0.275	0.039	-0.391	0.059
SD of Speed Trend	–	–	–	–	0.556	0.127
Downtrend Speed (slope < -5/60)	0.194	0.057	0.209	0.059	0.116	0.065
Intermediate Traffic	-4.976	0.786	-4.879	0.841	-4.619	0.900
Congested Traffic	-5.289	0.850	-5.097	0.919	-5.679	0.940
Speed Under Intermediate Traffic (mph)	0.104	0.013	0.102	0.014	0.097	0.015
Speed Under Congested Traffic (mph)	0.095	0.024	0.082	0.027	0.106	0.027
Friday	0.148	0.071	0.138	0.071	0.142	0.073
Weekend	0.117	0.054	0.156	0.068	0.124	0.056
Year = 2014	-0.535	0.064	-0.568	0.063	-0.844	0.189
SD of Year = 2014	–	–	–	–	0.842	0.266
Year = 2015	-0.176	0.069	-0.190	0.069	-0.186	0.072
Year = 2017	0.115	0.069	–	–	0.121	0.071
06:00 AM–11:59 AM	0.541	0.088	0.504	0.091	0.535	0.090
12:00 PM–17:59 PM	0.600	0.110	0.531	0.113	0.602	0.115
18:00 PM–23:59 PM	0.462	0.086	0.431	0.088	0.464	0.089
Covariance Estimate for Segment ID	–	–	1.087	0.106	–	–
AIC (smaller is better)	15,743.84		14,781.16		15,740	

temperatures. These findings concur with past researchers (Dutta & Fontaine, 2019; Roshandel et al., 2015; Zou & Tarko, 2016). Additional factors that were found to increase short-term crash risk included auxiliary lanes, downtrend speeds, and congestion while overpassing roads, average speed, and uptrend speeds were found to enhance safety.

Most of the time-dependent factors remained significant after adding random effects (multilevel logit) and random parameters (mixed logit). However, some static roadway characteristics were no longer statistically significant (i.e., passenger cars' AADT, horizontal curvature, median cable barrier presence, median barrier offset, entering auxiliary lane, and speed limit). Therefore, their estimates from the fixed effects model should be used carefully to reflect local roadway conditions from the selected sample.

Two parameters (the hourly speed trend and the 2014 indicator) were found to follow a normal distribution. For the 2014 indicator, 84% of the observations had a negative estimate (lower crash risk) while 16% had a positive parameter estimate. For the hourly speed trend, 76% of the records had a negative parameter estimate while 24% had a positive estimate (higher crash risk). These random parameters may reflect local conditions (e.g., road construction or temporal variations such as forming and dispersion of queues).

2.4.2 Effects on Injury Severity

Compared to the crash probability model, fewer predictors were found to affect injury severity. Table 2.2 presents the parameter estimates, standard errors, t-test statistics, and p-values of the binary logit model

TABLE 2.2
Parameter estimates of the conditional probability of severe crash model

Effect	Estimate	Std. Error	t value	Pr. > |t|
Intercept	-2.6924	0.1993	-13.51	<.0001
Proportion with Median Guardrail	1.3215	0.4725	2.80	0.0052
Mild Curve (3.5–5.4 degrees)	0.5946	0.3202	1.86	0.0635
Speed Limit Reduced by 5 mph	-0.6134	0.2552	-2.40	0.0163
Segment Proportion with Lighting	-1.1242	0.5530	-2.03	0.0422
Temperature (F)	0.0056	0.0030	1.88	0.0604
Ice Conditions	-0.7939	0.3183	-2.49	0.0127
SD of Hourly Travel Speed (mph)	0.1057	0.0176	6.01	<.0001
Downtrend Speed (slope < -5/60)	0.3794	0.1372	2.76	0.0058
Intermediate Traffic	-1.4939	0.9153	-1.63	0.1028
Speed Under Intermediate Traffic (mph)	0.0312	0.0172	1.81	0.0704
Friday	-0.2896	0.1900	-1.52	0.1277

for the probability of a severe outcome. Both the random effects and the random parameters were tested and were found to be not significant. Therefore, for this study, the fixed effects model was preferred. However, a larger sample may help obtain significant random parameters to create a third model to separate fatal and incapacitating crashes from non-incapacitating crashes.

The conditional probability of a severe outcome increased with the presence of a median guardrail (typical around bridges), mild curves, average speed, standard deviation of hourly speed, downtrend speeds, and intermediate congestion while lighting and lower speed limits reduced crash severity. The presence of icy conditions (the combination of precipitation and near-freezing temperatures) had a negative parameter estimate (reduced severity), which may be due to risk compensation as drivers adjust their operating speeds to account for inclement weather conditions. The parameter estimates of the hourly temperature and Friday indicators were not included since they may reflect local conditions in the sample.

2.4.3 Risk-Based SMS Application

The concept of risk-based safety management holds real promise. If disaggregate safety analysis can be used to supplement crash-based safety management by estimating the short-term crash risk performance of a target road segment, the estimated impacts, including the static and time-dependent elements, potentially can be valuable for multiple safety management tasks. While the presence of random parameters was confirmed, the application of such models is expensive, and their practical interpretation is cumbersome. Therefore, fixed effects models were selected for the prototype tool in this study.

This model calculates the hourly probability of a crash on a given segment using a logit form as described in Equation 2.5 and Equation 2.6.

$$P_i = \frac{e^{\beta_0 + \beta_1 X_{1,i} + \cdots + \beta_r X_{r,i}}}{1 + e^{\beta_0 + \beta_1 X_{1,i} + \cdots + \beta_r X_{r,i}}} \qquad \text{(Eq. 2.5)}$$

where P_i is the hourly probability of crash on segment-hour i, β are a set of estimated coefficients, and X are a set of independent predictor variables.

$$\beta_0 + \beta_1 X_{1,i} + \cdots + \beta_r X_{r,i} =$$

$$-\mathbf{3.9856} + \mathbf{0.299484} * volume1{,}000 + \mathbf{0.010069}$$

$$* aadt1{,}000_{-car} + \mathbf{0.057019} * aadt1{,}000_{-truck} - \mathbf{0.162261} * Overpass$$

$$+ \mathbf{0.145457} * curve_{-CD} + \mathbf{0.225796} * curve_{-EF} + \mathbf{0.190715} * BCAM\_p$$

$$-\mathbf{0.645816} * BGUM_{-P} + \mathbf{0.681674} * BGUR_{-P} - \mathbf{0.12285} * off_{-near}$$

$$-\mathbf{0.153395} * medhazard + \mathbf{0.083025} * roadhazard + \mathbf{0.518069}$$

$$* PAVCO\_p + \mathbf{0.550125} * ramp_{-in\_p} + \mathbf{1.495065} * ramp_{-out\_p} + \mathbf{0.296026}$$

$$* slim\_65 + \mathbf{0.122417} * SHRI\_w + \mathbf{0.255362} * light_{-rain} + \mathbf{0.24604}$$

$$* freeze + \mathbf{0.544991} * rainfreeze + \mathbf{0.058592} * speed\_StdDev$$

$$-\mathbf{0.261321} * beta\_speed + \mathbf{0.196682} * downtrend - \mathbf{0.140161}$$

$$* speed\_Mean + \mathbf{0.102853} * intermspeed + \mathbf{0.093825} * congspeed$$

$$-\mathbf{4.926306} * intermediate - \mathbf{5.220979} * congested + \mathbf{0.133184}$$

$$* friday + \mathbf{0.14076} * sunday - \mathbf{0.535518} * year\_2014 - \mathbf{0.176977}$$

$$* year\_2015 + \mathbf{0.115983} * year\_2017 + \mathbf{0.544993} * morning$$

$$+ \mathbf{0.604009} * afternoon + \mathbf{0.464139} * evening$$

(Eq. 2.6)

where,
    volume1,000 is the hourly traffic volume in 1,000s veh/h,
    aadt1,000_car is the passenger vehicle's AADT in 1,000s veh/day,
    aadt1,000_truck is the combined truck's AADT in 1,000s veh/day,
    Overpass is 1 if there is an overpassing road and 0 otherwise,
    Curve_CD is the proportion of segment length with moderate curve (5.5–13.9 degrees),
    Curve_EF is the proportion of segment length with sharp curve (14 or more degrees),
    BCAM_p is the proportion of segment length with median cable barrier,
    BGUM_p is the proportion of segment length with median guardrail,

BGUR_p is the proportion of segment length with roadside guardrail,

off_near is 1 if the barrier offset is 11 ft or less and 0 otherwise,

medhazard is the number of median hazards (i.e., beginning of barriers),

roadhazard is the number of roadside hazards,

PAVCO_p is the proportion of segment length with concrete pavement,

ramp_in_p is the proportion of segment length with auxiliary lane corresponding to entering ramp,

ramp_out_p is the proportion of segment length with auxiliary lane corresponding to exiting ramp,

slim65 is 1 if the maximum posted speed limit is 65 mph 0 otherwise,

SHRI_w is the average roadside shoulder width in ft,

light_rain is 1 if there is precipitation of 0.1 in/h,

freeze is 1 if the air temperature is 32F° or less and 0 otherwise,

rainfreeze is 1 if there is any precipitation under freezing temperatures and 0 otherwise, *speed_StdDev* is the standard deviation of travel speeds in mph,

beta_speed is the temporal speed trend,

downtrend is 1 if speed is reducing at a rate equal or higher than 5 mph/h,

speed_Mean is the average operating travel speed in mph,

intermspeed is the interaction of travel speed and intermediate traffic indicator,

congspeed is the interaction of travel speed and congested traffic indicator,

intermediate is 1 if the congestion index *(9)* is between 0.1 and 0.5 and 0 otherwise,

congested is 1 if the congestion index is higher than 0.5,

Friday/Sunday are day of the week indicators,

year_2014/year_2015/year_2017 are year indicators, and

morning/afternoon/evening are time of day indicators.

Similarly, the second model, which uses the same logit form (Equation 2.1), calculates the hourly probability of a severe outcome (injury or fatal) on a given segment conditioned on crash occurrence. This model is shown in Equation 2.7.

$$\beta_0 + \beta_1 X_{1,i} + \cdots + \beta_r X_{r,i} =$$
$$-2.692599 + 1.321856 * BGUM\_p + 0.594525 * curve\_B$$
$$-0.613392 * slim\_65 - 1.124153 * Light\_p + 0.005573 * temp$$
$$-0.793718 * rainfreeze + 0.105718 * speed\_StdDev + 0.379398$$
$$* downtrend - 1.493624 * intermediate + 0.031159 * intermspeed$$
$$-0.28966 * friday$$

(Eq. 2.7)

where,

Curve_B is the proportion of segment length with a mild curve (3.5–5.4 degrees),

Light_p is the proportion of segment length with artificial lighting,

temp is the air temperature in F, and
all other variables as described before.

3. SIGNALIZED INTERSECTIONS

The case study of implementing a risk-based SMS on Indiana's signalized intersections is presented in this chapter. Three typical crash-generating scenarios at signalized intersections were identified: same-direction (SD) crashes, opposite-direction (OD) crashes, and right-angle (RA) crashes. The time-dependent and fixed factors safety effects can be estimated with separate models. Due to the lack of available data, however, these analysis results will not be implemented system-wide. The potential implementation suggestion and steps are presented here with several in-sample simulation examples.

3.1 Data Preparation

The time-dependent safety performance at signalized intersections can be affected by various factors, as analyzed in JTRP Project SPR-4302 (Tarko et al., 2021), where the detailed description of the data sources were presented. Although the analysis of signalized intersections shares some common data sources as the analysis of rural freeway segments, some of the data preparation processes are different. This chapter consists of three main sections: general data description, directional volume data preparation, and signal setting data preparation.

3.1.1 General Data Description

In general, the data preparation for risk-based safety performance analysis on signalized intersections involved six major data sources: crash, weather, speed, detector data, signal settings, and geometry.

3.1.1.1 Crash. The crash data in this analysis were accessed from the ARIES database, which is where the police reports for each crash are available. Apart from the typical collected crash information, this study also labelled the directional movements of the involved vehicles in the crash based on crash narratives or crash diagrams. Figure 3.1 shows a typical crash diagram from a police report. It could be inferred from the diagram that Unit2 was traveling southbound and attempted to make the left turn when it hit Unit1 (traveling northbound) and then had a secondary collision with Unit3. To simplify the problem, this study only considered the first collision manner and the major involved traffic participants of the collision. The directional movements of the example crash in Figure 3.1 were recorded as Unit1-NT (northbound through) and Unit2-SL (southbound left turn). These directional movement data were later used to link the corresponding volume, speed, and signal settings.
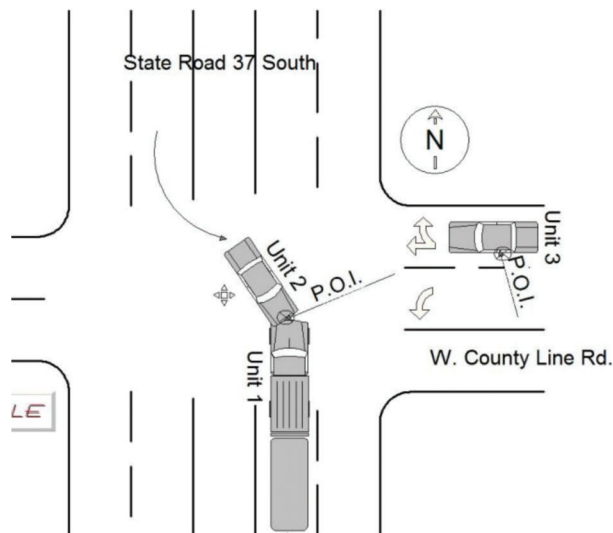
Figure 3.1 Typical crash diagram (OD collision example).

3.1.1.2 Weather. The weather data for signalized intersections utilize the same sources and interpolating methods as described for rural freeway segments.

3.1.1.3 Speed. Different from rural freeway segment data, the speed data for signalized intersections comes from INRIX, which provides better coverage of the road segments and is precise to the minute level. These GPS probe vehicle-based speed data are collected every 60 seconds and the average length of the data collecting segments is 0.5 mile. Because the INRIX speed data are collected from GPS probe vehicles or cell phones, there are certain time periods where there are no observations. The INRIX speed data table includes the location (INRIX segment ID), date, epoch, length, speed, and data quality index. The disadvantage of the INRIX speed data is that it only updates the segmentations every six months, which therefore produced three different shapefiles in the studied year of 2018. The segment lengths among the different shapefiles were slightly different, but this study assumed that the speeds from those segments were comparable.

3.1.1.4 Detector data. The research team received the raw detector data and the detector feature table from INDOT. Since INDOT connects its signalized intersections data to their system gradually during the project period, not all of the data were available at the same time. There were raw detector logging data for about 300 intersections during the year 2018. The detector feature table that recorded the location and operating status of the detectors was also provided to the research team. Connecting the logging data and the detector feature table yielded 130 intersections with complete detector data. The detailed data preparation steps are described in Section 3.1.2.

3.1.1.5 Signal setting. The signal settings of the targeted intersections were collected with help from the Indianapolis Traffic Management Center. Among the 130 signalized intersections with complete detector data, there were 115 intersections with credible signal setting data. The other intersections were removed or upgraded after the year 2018. The detailed data preparation steps are described in Section 3.1.3.

3.1.1.6 Geometry. The geometry data were collected from INDOT's road inventory data sets supplemented with other road characteristics gathered from Google Earth's historical imagery. The manually collected data included lane settings (e.g., lane number, left turn only lane, exclusive right turn lane), intersection scale (e.g., distance between stop lines, skewness), and approaching segment characteristics (e.g., close access points, flashing yellow signs, speed limits, crosswalks).

3.1.2 Volume Data Preparation

The volume data preparation followed the procedure shown in Figure 3.2. The raw logging data recorded all the operations of the detectors precisely within 0.1 second, including phase on/off, gap out, green splits, etc. The detector on and off were used to determine the vehicle counts for each detector.

One major challenge of using detector data was that the loop detector may not have performed well across the entire year 2018. Many of the detectors were broken or were outputting unreasonable values during the studied period. To overcome this problem, a missing data imputation program was developed.

To predict hourly vehicle counts, three kinds of time series models were tested: (1) the seasonal trend decomposing using the LOESS model (Stlm); (2) the Harmonic regression model (HamReg); and (3) the Trigonometric seasonality, Box-Cox transformation, ARMA errors, and Trend and Seasonal components model (TBATS). One example of their prediction across 1 week is illustrated in Figure 3.3. Because all three models performed well in capturing the volume trends across time, the most calculation-efficient method, Stlm, was selected for application on all the detector data. Figure 3.4 shows the hourly volume profile across 5 weeks where the dashed line is the predicted missing volume.

In the last step, the detectors' hourly vehicle counts were transferred into directional volumes based on the detectors' feature table and manual checking from the Google Map historical images. Figure 3.5 shows one typical example of the settings of detectors at signalized intersections. The advanced detectors on the upstream usually counted all the through and left turn volume while the detector at the stop lines counted the left turn or through movements. Because right turn traffic is seldom controlled by signals, there were very few specific detectors for right turn traffic. To compare the volumes among the intersections consistently, the directional volume only considered left turn traffic and through/right turn traffic combined. For each movement (e.g., ET represents eastbound through/right turn
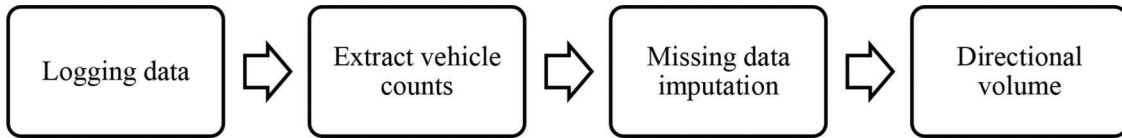
Figure 3.2   Volume data preparation steps.
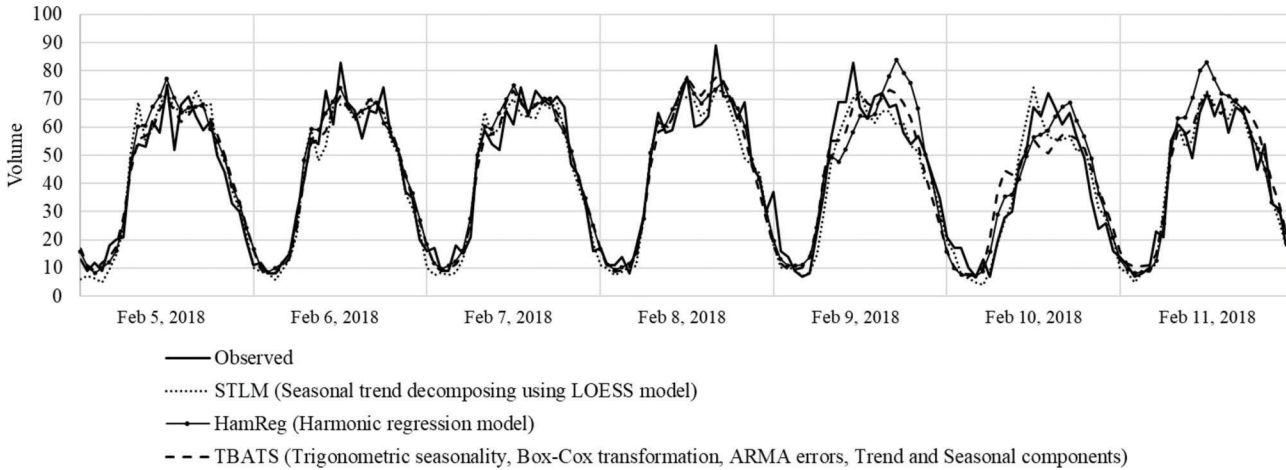
Figure 3.3   Predicted time series models vs. observed vehicle counts across 1 week.
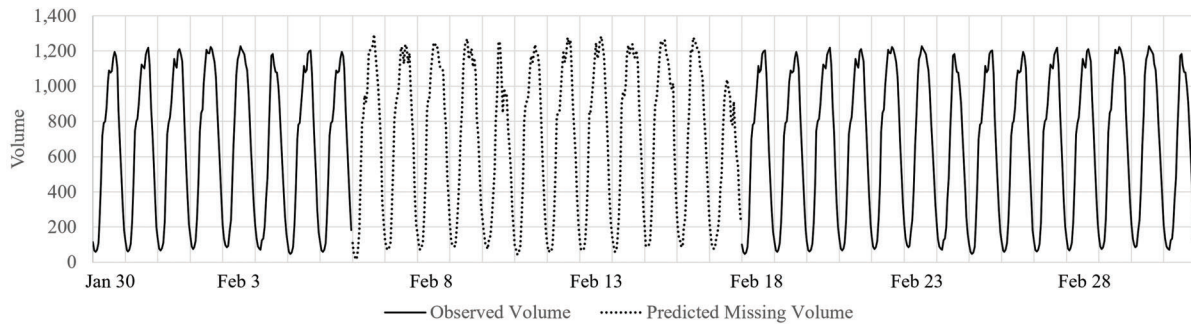
Figure 3.4   Missing volume prediction example.

traffic), the directional volumes were obtained by summing up the vehicle counts from all the corresponding detectors.

3.1.3 Signal Setting Data Preparation

The obtained signal settings were saved in Microsoft Excel files in the following sheets: (1) unit and configuration, (2) phase & overlaps, (3) coordination, (4) preempt, (5) TOD, and (6) detector options, which provided abundant signal settings information. Two signal settings were used in this study: (1) local phase settings and (2) coordination settings.

The local phase settings were obtained mainly from controller timing data. Table 3.1 shows one typical example of the controller timing data. The first row is the phase number following the general application setting of the dual-ring control scheme (i.e., phases 2

and 6 are the through movements on the major street, phases 1 and 5 are the left turn movements on the major street, and phases 4 and 8 are the through movements on the side street). The other rows present the settings of the corresponding phase. For example, the second row is the minimum green time for each phase. The two zeros for phases 3 and 7 imply that the example intersection in Table 3.1 does not have protected phases for the side street's left turn movements.

The coordination settings are inferred from the coordination patterns and day plans. Figure 3.6 and Figure 3.7 show one example of the coordination settings at signalized intersections. In this example, Figure 3.6 shows three coordination patterns with the same cycle length (90 seconds) but different offsets (79, 41, 78) and different phase splits. Figure 3.8 shows the day plans of the targeted intersection with day plan 1 for weekdays and day plan 2 for weekends. For each
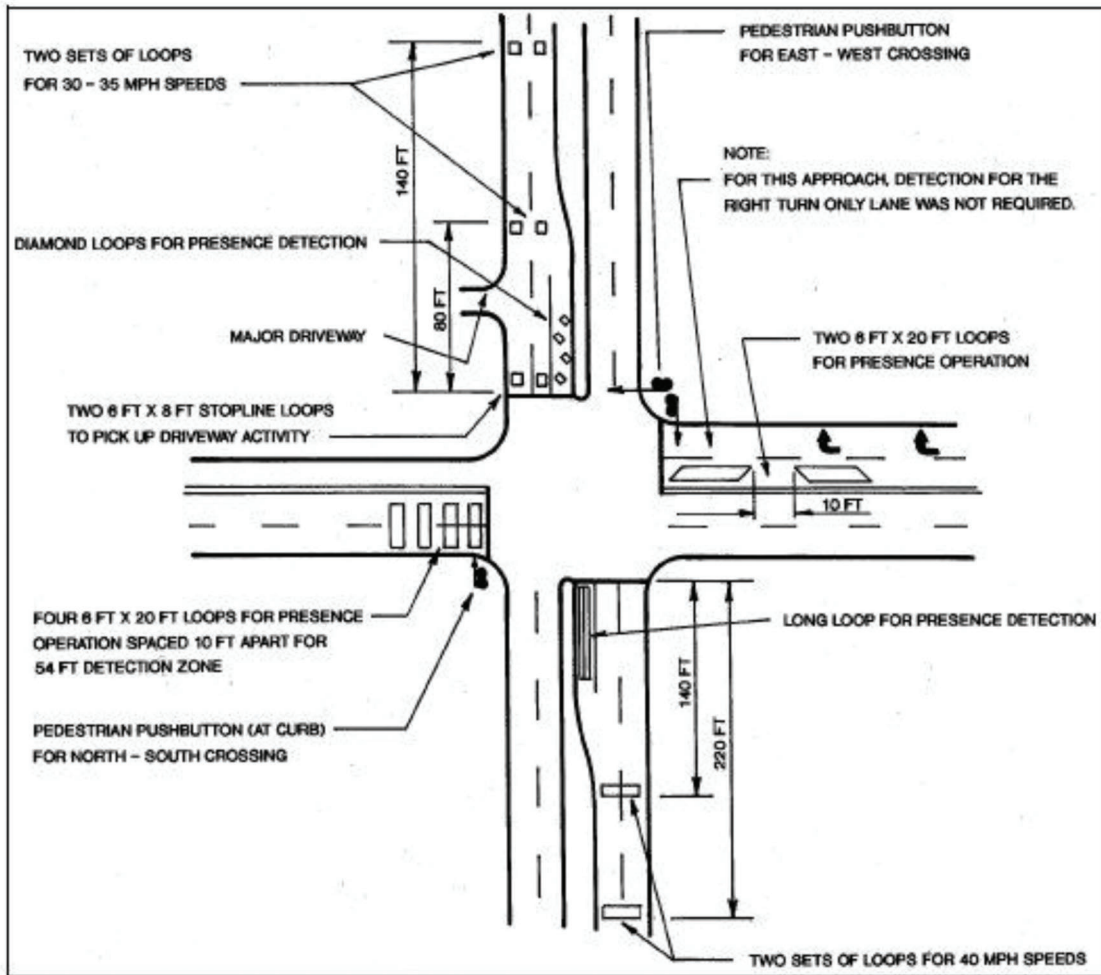
Figure 3.5 Typical settings of detectors at signalized intersections.

day plan, the coordination patterns were performed differently according to the day plan settings. For example, during the weekdays, pattern 2 took effect at 6:00 AM from free mode (pattern 99 or 0) and then switched to pattern 1 at 9:00 AM; it switched to pattern 3 at 15:00 PM and then to pattern 1 at 18:00 PM; and finally, it switched to the free mode at 22:00 PM until the next day. All the coordination settings were coded and assigned to the crash and non-crash samples according to their occurrence hours.

3.2 Analysis Methods

The analysis of signalized intersections in this study was more complex considering the diversity of the crash generating scenarios compared to rural freeway segments. The crashes related to signalized intersections were first classified into three typical scenarios and then analyzed following a similar strategy to the one for rural freeway segments.

To estimate the safety effects of time-dependent and fixed predictors, variations of logistic regression models were tested; but considering the application of

the models, only the sequential binary logit models with fixed effects were adopted, which are discussed in this chapter. The statistical settings of the sequential binary logit models are presented in Section 2.2.1.

3.2.1 Crash Generating Scenarios

Unlike crashes on segments, crashes that occur at signalized intersections can vary across each other on maneuver, causality, and involved traffic participants. For example, an RA crash that involves two vehicles from different segments is very different from a rear-end crash on the same approaching segment to the intersection. When considering the exposure factor of crash occurrence, usually traffic volume, an RA crash analysis should use the two involved traffic volumes, while the rear-end crash should use only the volume from the one segment.

To discompose the variation of crashes at the signalized intersections while maintaining enough crash samples for analysis, three typical crash generating scenarios were proposed in this study based on the basic logic of a dual-ring signal control scheme (Figure 3.8).

Seven types of crashes were identified through this control logic, and they were grouped into three general categories: SD crashes, OD crashes, and RA crashes.

TABLE 3.1
Typical controller timing data

Phase	1	2	3	4	5	6	7	8
Min Green	7	10	0	7	7	10	0	7
Bike Min Green	0	0	0	0	0	0	0	0
Cond Service Min Green	0	0	0	0	0	0	0	0
Delay Green	0	0	0	0	0	0	0	0
Walk	0	8	0	9	0	8	0	9
Walk2	0	0	0	0	0	0	0	0
Walk Max	0	0	0	0	0	0	0	0
Ped Clear	0	25	0	30	0	25	0	30
Ped Clear 2	0	0	0	0	0	0	0	0
Ped Clear Max	0	0	0	0	0	0	0	0
Veh Ext	3.0	3.0	0.0	3.0	3.0	3.0	0.0	3.0
Veh Ext2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Max 1	20	40	0	20	20	40	0	20
Max 2	20	60	0	40	20	60	0	40
Max 3	0	0	0	0	0	0	0	0
Dynamic Max	0	0	0	0	0	0	0	0
Dynamic Max Step	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Yellow	3.6	3.6	0.0	3.6	3.6	3.6	3.0	3.6
Red Clear	2.0	1.8	0.0	1.8	2.0	1.8	0.0	1.8
Red Max	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Red Revert	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Actuations Before	0	0	0	0	0	0	0	0
Added Initial	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Max Initial	0	0	0	0	0	0	0	0
Time Before Reduce	0	0	0	0	0	0	0	0
Cars Before Reduce	0	0	0	0	0	0	0	0
Reduce By	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Time to Reduce	0	0	0	0	0	0	0	0
Min Gap	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Same-direction collision scenarios

- SDS Same phase: 1-1, 2-2, 3-3, 4-4, 5-5, 6-6, 7-7, 8-8
- SDC Concurrent phases: 1-6, 2-5, 3-8, 4-7

Opposite-direction collision scenario

- ODC Consecutive phases: 1-2, 5-6, 3-4, 7-8
- ODS Same phase: 2-6, 4-8

Right-angle collision scenarios

- RAN Cross-barrier with no phase skipped: 4-1, 4-5, 8-1, 8-5, 2-3, 2-7, 6-3, 6-7
- RAL Cross-barrier with left-turn phase skipped: 2-4, 2-8, 6-4, 6-8
- RAT Cross-barrier with through phase skipped: 1-3, 1-7, 5-3, 5-7

The typical movements of the seven collision scenarios are shown in Figure 3.9. The proposed three general crash generating scenarios represent three typical collision scenarios at signalized intersections.

- The SD crash scenario represents the most prevailing type of crash at signalized intersections while rear-end and sideswipe crashes were more prevalent at the approaching segments. The occurrence of such crashes is expected to be mostly affected by the traffic volume, dynamic queueing status, coordination, and operating speed on the segment.
- The OD crashes scenario includes both ODC and ODS, but according to the distribution of crash records, most of the ODCs crashes between the through traffic and the OD's left turn traffic. The occurrence of such crashes is expected to be mostly affected by signal settings (protected or permissive left turn) and exposure (through traffic and left turn traffic).
- The right-angle crashes represent crashes between the main street traffic and the side street traffic. Although

Figure 3.6 Example of coordination settings and patterns.

Day Plan 1

Month of Year						Days of Week							Days of Month										
J	F	M	A	M	J	S	M	T	W	T	F	S	1	2	3	4	5	6	7	8	9	10	
X	X	X	X	X	X	.	X	X	X	X	X	.	X	X	X	X	X	X	X	X	X	X	
J	A	S	O	N	D								11	12	13	14	15	16	17	18	19	20	
X	X	X	X	X	X								X	X	X	X	X	X	X	X	X	X	
													21	22	23	24	25	26	27	28	29	30	31
													X	X	X	X	X	X	X	X	X	X	X

Day Plan 2

Month of Year						Days of Week							Days of Month										
J	F	M	A	M	J	S	M	T	W	T	F	S	1	2	3	4	5	6	7	8	9	10	
X	X	X	X	X	X	X	.	.	.	.	.	X	X	X	X	X	X	X	X	X	X	X	
J	A	S	O	N	D								11	12	13	14	15	16	17	18	19	20	
X	X	X	X	X	X								X	X	X	X	X	X	X	X	X	X	
													21	22	23	24	25	26	27	28	29	30	31
													X	X	X	X	X	X	X	X	X	X	X

Day Plans

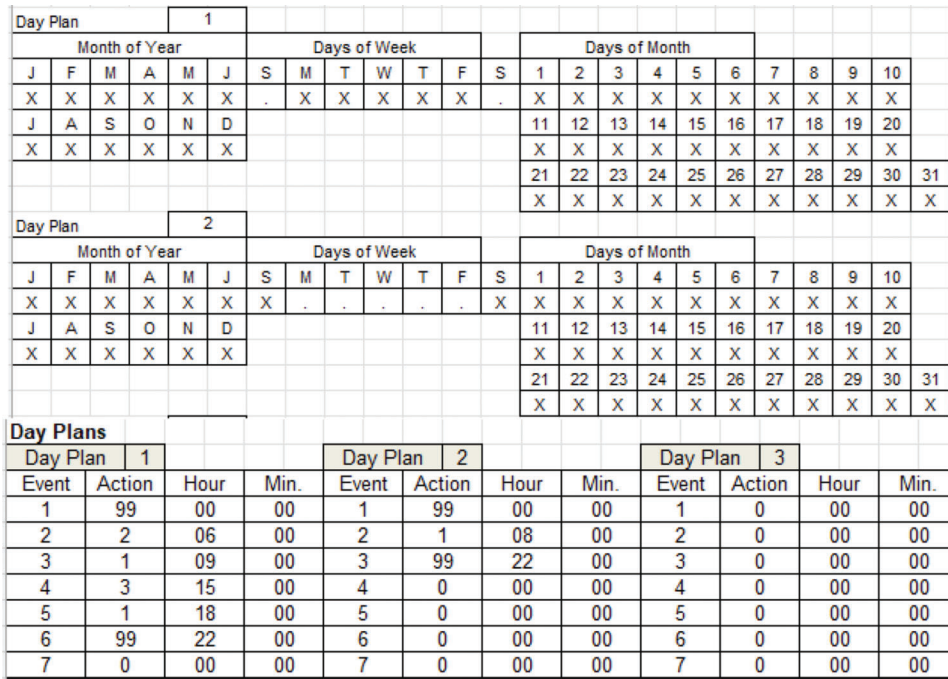| Day Plan 1				Day Plan 2				Day Plan 3			
Event	Action	Hour	Min.	Event	Action	Hour	Min.	Event	Action	Hour	Min.
1	99	00	00	1	99	00	00	1	0	00	00
2	2	06	00	2	1	08	00	2	0	00	00
3	1	09	00	3	99	22	00	3	0	00	00
4	3	15	00	4	0	00	00	4	0	00	00
5	1	18	00	5	0	00	00	5	0	00	00
6	99	22	00	6	0	00	00	6	0	00	00
7	0	00	00	7	0	00	00	7	0	00	00

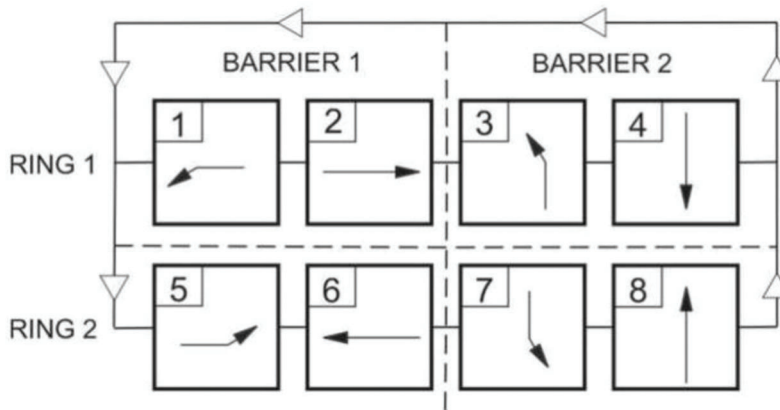**Figure 3.7** Example of coordination day plans.

Figure 3.8 Dual-ring signal control scheme.

RA crash-generating scenarios have the most possible conflict pairs, their real occurrence possibility is relatively low because all the conflicting scenarios are prohibited by signal controls. According to crash records, such crashes happen only when drivers violate the traffic rules or during the signal switching periods.

3.2.2 Empirical Setting

The selection of sample signalized intersections was limited to data availability, which was mostly affected by the detector data for the hourly traffic volume and the signal setting data, depending on whether the targeted intersections were connected to the INDOT traffic signal database. While the research team received the raw detector data for approximately 300 intersections in the year 2018, only 130 of them contained complete detector feature tables that included information about the detectors' locations and settings. Out of the 130 intersections, the research team obtained 115 intersections' signal settings with the help of the Indianapolis Traffic Management Center. Figure 3.10 shows the spatial distribution of the 115 selected signalized intersections. These intersections are distributed across the entire state of Indiana and some of them are in the same corridor, which enabled the analysis of the coordination effects on safety.

One major challenge for disaggregated level analysis was the highly imbalanced crash and non-crash ratios (i.e., there were far more hours without crashes than hours with crashes). There were 1,001 crashes (SD: 783; OD: 146; RA: 72) identified on the analyzed 115 intersections during the year 2018, but there were (115*365*24 − 1,001) = 1,006,399 non-crash hours.
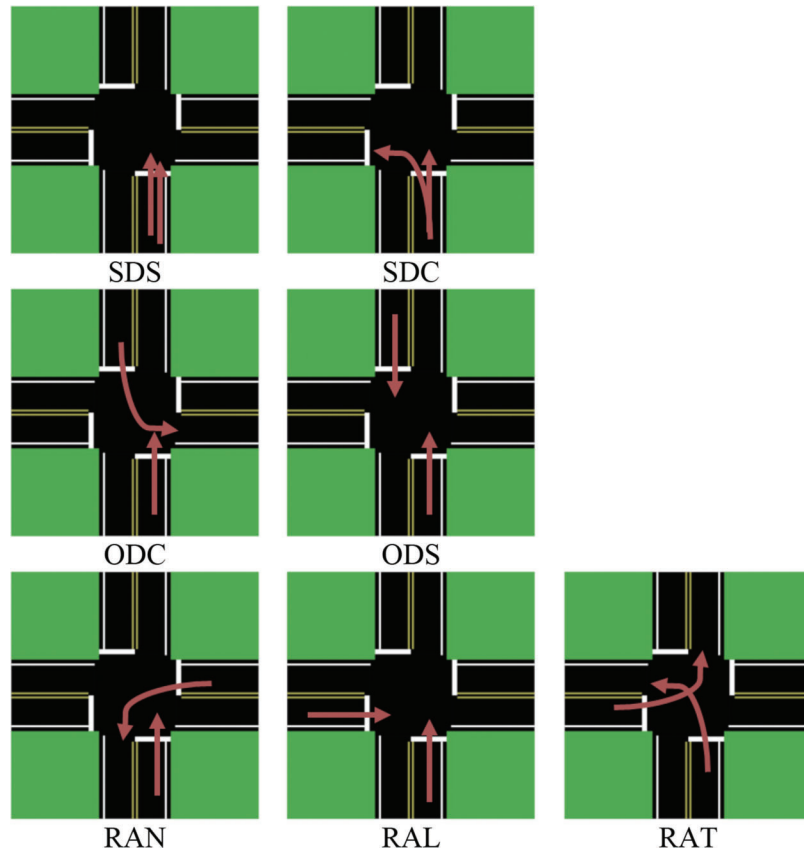
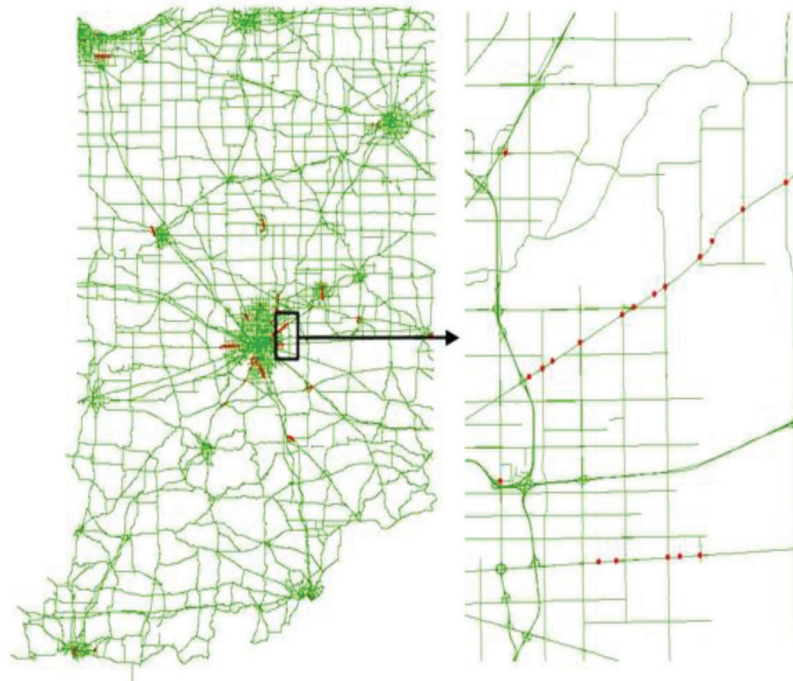**Figure 3.9** Typical crash generating scenarios.

Figure 3.10 Spatial distribution of sample signalized intersections used for analysis.

Such a skewed crash and non-crash ratio prohibited normal estimation of the safety effects. To overcome this problem, a fixed crash and non-crash ratio (1:30) was adopted when sampling the non-crash hours, and correction of the estimated intercept was performed when the real crash probability was calculated.

3.3 Results and Discussion

As discussed in the analysis methods, three types of crashes were modeled separately using sequential logit models. Both time-dependent and fixed factors were identified as significantly influencing the probability and severity of crashes. Considering the differences among the three types of crashes, their safety effects are presented separately. Although there were over 100 variables tested for each model during the modeling process, few of them turned out to be statistically significant. The discussion of the models below focuses on the significant variables.

3.3.1 Safety Effects on Same Direction Collision

Table 3.2 and Table 3.3 show the estimates, standard error, z.value, and p.value of the same direction crash probability model and conditional probability model. The highly significant variables are shown in bold text in the tables, and some of the merely significant variables with p.values lower than 0.15 are discussed. There was one fixed data source (the geometry settings), and four time-dependent data sources (weather conditions, traffic volume, segment speed features, and signal settings) were expected to affect the hourly crash probability and severity. The discussion of the models will follow these five data sources.

3.3.1.1 Weather conditions. The variable *LightRain,* which was defined as hourly precipitation greater than 0 but lower than 2.5 mm, was found to increase the hourly probability of SD crashes (mostly rear-end crashes), which is intuitive since light precipitation can make the pavement slippery and drivers may not be aware of the risk. Although the research team attempted to include temperature-related features (cold weather, ice condition) and other precipitation characteristics (heavy rain, accumulated rain) in the model, they were not found to be significant, nor did they outperform *LightRain*. This result may have been due to light rain conditions being the most common situation and more significant weather conditions are too rare to become significant.

3.3.1.2 Volume-related variables. There were three volume-related variables, *logVol*, *LTVolRate*, and *logVolSide*, in the probability model and one volume-related variable, *Vol*, in the conditional severity model, which turned out to be significant. *Vol* was the sqrt of the product between the through and left turn volumes on the approaching segment to the intersection; *logVol*

TABLE 3.3
Parameter estimates of the conditional probability of same-direction severe crash model

Variable	Estimate	std.error	z.value	p.value
Intercept	-1.366	0.272	-5.018	0.000
Vol	**-0.003**	0.001	-3.063	0.002
SpStd	**0.076**	0.038	2.000	0.045
Curved	0.639	0.424	1.505	0.132
RushHour	-0.390	0.254	-1.535	0.125

Note: Estimates of the significant variables are bolded.

TABLE 3.2
Parameter estimates of the hourly probability of same-direction crash model

Variable	Estimate	std.error	z.value	p.value
Intercept	**-15.213**	0.527	-28.883	0.000
LightRain	**0.342**	0.127	2.706	0.007
logVol	**0.634**	0.071	8.919	0.000
LTVolRate (0.2-0.8)	**-0.685**	0.226	-3.033	0.002
LTVolRate (0-0.2)	-0.047	0.215	-0.221	0.825
logVolSide	**0.785**	0.062	12.685	0.000
SpStd	**0.027**	0.015	1.821	0.069
LargeCycle	**1.038**	0.179	5.812	0.000
Coordination (No)	**0.393**	0.176	2.238	0.025
Coordination (Free)	0.017	0.095	0.173	0.863
Coordination (EarlyArr)	**-1.104**	0.469	-2.353	0.019
Coordination (LateArr)	0.229	0.266	0.859	0.390
GreTimeSplit	**5.740**	0.347	16.520	0.000
LaneNum(2)	**1.320**	0.165	7.979	0.000
LaneNum(3)	**1.654**	0.157	10.528	0.000
LaneNum(4)	**1.699**	0.187	9.086	0.000
ExcluRightTurn	**-0.224**	0.110	-2.046	0.041
Curved	**0.789**	0.194	4.079	0.000

Note: Estimates of the significant variables are bolded.

was the logarithm of *Vol*; *LTVolRate* was the rate between the left turn volume and the through volume (ranged from 0 to 1); and *logVolSide* was the logarithm of the total volume on the side street.

As expected, the hourly traffic volume (*logVol*) served as the exposure factor (i.e., the larger the volume, the higher probability of crashes). Compared to the raw volume, the logarithm of the volume performed better. The volume on the side street (*logVolSide*) also had a positive effect on crash probability because the side street volume could have reflected the scale as well as the traffic intensity of the entire intersection (i.e., the busier the intersection, the higher probability of the occurrence of rear-end crashes). The *LTVolRate* reflected the intensity of the left turn traffic, according to the model, compared to segments with large left turn volumes (*LTVolRate* $\geq$ 0.8). Segments with reasonable left turn volumes (*LTVolRate* between 0.2 and 0.8) were significantly safer, but this effect was not significant when there were insignificant left turn volumes (*LTVolRate* < 0.2). The effect of the volume (*Vol*) on the conditional severity model was negative, which was expected since the conditional severity model estimated the probability of severe crashes (fatal and injured) given the occurrence of a crash. When the traffic volume was high, the operating speed on the segment was low, inducing more crashes that were mostly PDO crashes.

3.3.1.3 Speed features. The variable *SpStd*, defined as the standard deviation of the speeds on the segment across one hour (the speed data were precise to the minute in the raw speed data), was found to increase both the probability and the severity of SD crashes. This speed feature reflects the frequent and significant changes of speed across minutes, which is a good indicator of the disturbance of the traffic flow. This disturbance of traffic flow can induce more crashes that are also at a higher severity level. Any engineering countermeasures that could mitigate such disturbances would bring safety benefits.

It should be noted that to avoid the potential endogeneity of the speed data (the change of speed is due to the occurrence of crash), the extracted speed hour is the exact hour before the crash. For example, if the crash happened at 8:24 AM, then the extract speed is from 7:24 AM to 8:24 AM, so it did not include the time after the crash occurrence. The variable *SpStd* reflects the status of the traffic flow before crashes.

3.3.1.4 Signal settings. There are three significant signal setting variables in the crash probability model while there was nothing significant in the conditional severity model. The two local signal-related variables, *LargeCycle* and *GreTimeSplit*, were both positive while the coordination related variable, *Coordination*, varied across different conditions.

The variable *LargeCycle* is the dummy variable reflecting the local signal cycle (sum of minimum green time) greater than 50 seconds and the variable *GreTimeSplit* is the proportion of minimum green time for the two phases (the left turn and through movements) of the segment. The local signal cycle reflects the scale and the traffic intensity of the intersection, and the green time split reflects the traffic demand intensity of the targeted segment. Both variables are positively related to the crash probability.

The variable *Coordination* is a categorical variable comparing five different coordination scenarios.

1. *No:* there is no coordination on the targeted segment (usually the side street case).
2. *Free:* the coordination is set as the Free mode (usually the main street during night).
3. *EarlyArr:* the coordination is on; and according to the reference travel time and relative offsets between intersections, the traffic fleet of the coordinated upstream intersection arrives at the targeted intersection slightly earlier (0–5 seconds) than the start of the green light (Figure 3.11).
4. *LateArr:* the coordination is on; and according to the reference travel time and relative offsets between intersections, the traffic fleet of the coordinated upstream intersection arrives at the targeted intersection slightly later (0–5 seconds) than the start of the green light (Figure 3.12).
5. *Reference:* the coordination is on, but the reference travel time and the relative offsets are far from each other (the absolute value of their difference is larger than 5 seconds).

Note: the reference travel time was obtained as the ratio between the measured distance between consecutive intersections and the posted speed limits on the segments.

The relative offsets among intersections are the major parameters that control the quality of coordination. Theoretically, in terms of traffic efficiency, if the relative offset is close to the reference travel time, the green band will be larger, and more vehicles will take advantage of the coordinated phase. However, the safety effects of coordination are not consistent according to the obtained SD crash probability model. When there is coordination (on the main street), compared to the reference (the not well coordinated case), *EarlyArr* is significantly safer while there is no significant difference for the other two cases (*LateArr* and *Free*). This could be explained by the clearance of the coordinated fleet. The SD crashes, mostly rear-end crashes, are prone to happen when traffic lights turn yellow, and drivers experience the yellow light dilemma. During their hesitation, they may be hit by vehicles from the rear. However, for the coordinated traffic fleet, if all the vehicles from the upstream intersection pass the targeted intersection before the end of green, such yellow light dilemma situations tend to be less frequent, which is exactly what happens for the *EarlyArr* scenario. On the other hand, the effect of *Coordination(No)* also is possible, indicating that under the same other conditions (e.g., volume), side streets without coordination are significantly more dangerous,
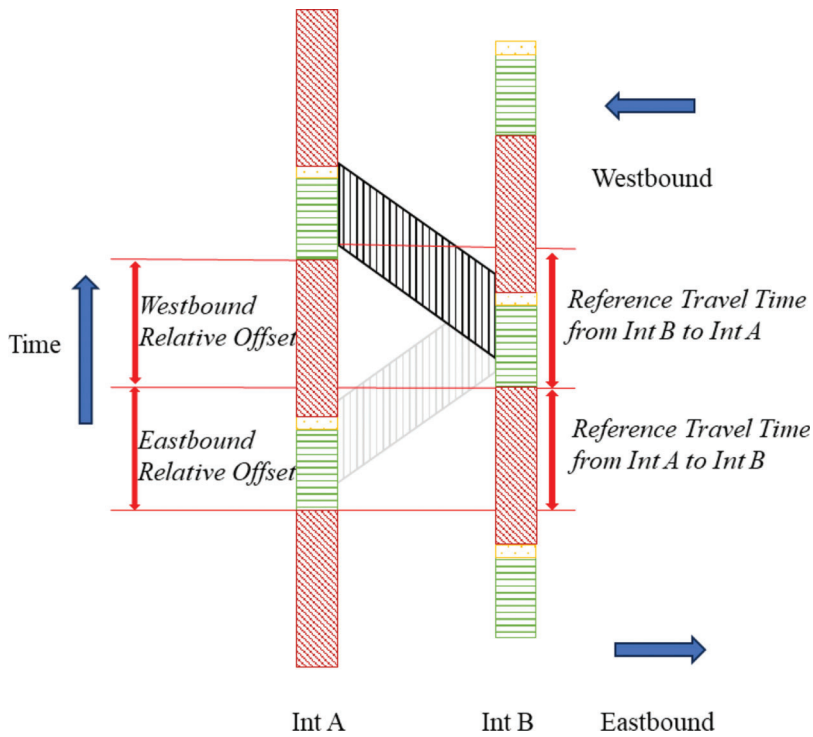
Figure 3.11 Coordination (westbound: early arrival example).
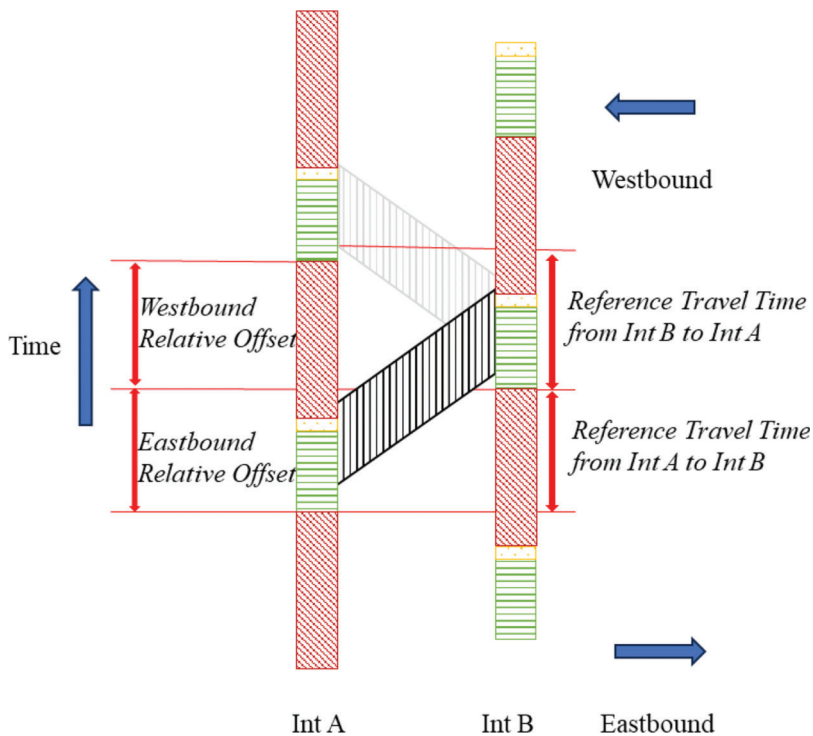
Figure 3.12 Coordination (eastbound: late arrival example).

which proves the safety effects of coordination in reducing SD crashes.

3.3.1.5 Geometry settings. Several geometry settings (*LaneNum*, *ExcluRightTurn*, and *Curved*) were identified as significantly affecting the SD crash probability and severity. *LaneNum* is the number of lanes for the through and right turn traffic, which ranged from 1 to 4 in the sample data. The one lane cases are used as references. Compared to one lane

cases, more lanes introduce higher SD crash probability, but the effects of *LaneNum(3)* and *LaneNum(4)* are very close. More lanes indicate more lane change possibilities, and thus more sideswipe crashes; therefore, even the volume factor was included in the model as the effects of the number of lanes are still significant.

The variable *ExcluRightTurn* is a dummy variable reflecting whether or not there is an exclusive right turn lane or turning bay. A negative sign of its estimate indicates this geometry setting could reduce crash probability. The existence of an exclusive right turn lane can direct the right turn traffic earlier and thus reduce the possibility of lane changing collisions.

The variable *Curved* is a dummy variable indicating whether the approaching segment to the intersection is curved. This variable turned out to be positively related to both crash probabilities (highly significant) and severity (merely significant). When a segment is curved, the sight distance can be affected, making it more difficult for drivers to change lanes and thereby inducing more SD collisions.

3.3.2 Safety Effects on Opposite Direction Collision

Table 3.4 shows the estimation results of the hourly probability of OD crashes and Table 3.5 shows the estimation results of the conditional probability of OD severe crashes. Since some of the variables in the OD crash models are discussed in the previous crash models, the discussion for those variables will be brief.

It should be noted that out of the 146 OD crashes, there were only four crashes involving through movements from opposite directions, while the remaining crashes were through movement collisions with OD left turn movements. To simplify the problem, when there were two through movements involved, the one with the lower hourly traffic volume was treated as the left turn movement.

3.3.1.6 Weather conditions. The variable *LightRain* was found to significantly increase the probability of

TABLE 3.4
Parameter estimates of the hourly probability of OD crash model

Variable	Estimate	std.error	z.value	p.value
Intercept	**-9.740**	0.847	-11.501	0.000
LightRain	**0.530**	0.262	2.024	0.043
logVol	**0.380**	0.132	2.884	0.004
HighImbal	**-0.805**	0.410	-1.963	0.050
logVolSide	**0.373**	0.107	3.499	0.000
SpStd	**0.109**	0.033	3.335	0.001
LaneNumber (3,4)	**0.908**	0.254	3.582	0.000
FlashYellow	**1.463**	0.197	7.431	0.000
LeftGreSplit	**3.833**	1.140	3.363	0.001
SmallCycle	**0.841**	0.399	2.106	0.035
Curved	**0.517**	0.232	2.226	0.026

Note: Estimates of the significant variables are bolded.

TABLE 3.5
Parameter estimates of the conditional probability of OD severe crash model

Variable	Estimate	std.error	z.value	p.value
Intercept	**-2.961**	1.716	-1.726	0.084
LTVolRate(0–0.2)	**-3.046**	0.899	-3.389	0.001
VolSide	**-0.001**	0.001	-2.159	0.031
Speed	**0.119**	0.036	3.287	0.001
LargeCycle	**-2.865**	0.996	-2.876	0.004
LaneNumber (2, 3, 4)	**2.235**	1.188	1.881	0.060

Note: Estimates of the significant variables are bolded.

OD crashes, which was expected since precipitation makes roadways slippery and impedes vehicle control.

3.3.1.7 Volume-related variables. In the crash probability model, *logVol* and *logVolSide* were found to increase crash risk as expected. The interpretations of these two variables were similar to those in SD crash models, but the variable *logVol* was defined as the logarithm of sqrt (Through Volume * Left Turn Volume), which reflects in general the intensity of both traffic movements involved. The variable *HighImbal* is a binary indicator of whether the through movement's volume is 40 times greater than the OD left turn volume. A negative sign implies that when there is very little left turn volume in the opposite direction, the crash probability is much smaller.

In the conditional crash severity model, *LTVolRate(0–0.2)* and *VolSide* were found to reduce the probability of severe crashes where the variable *LTVolRate(0–0.2)* represents the cases when the ratio between the left turn volume and the opposite direction through volume is between 0 and 2. These two variables are difficult to interpret.

3.3.1.8 Speed features. In the hourly crash probability model, the *SpStd* variable was found to be positively related to crash risk, which is consistent with the findings in the SD models. In the conditional crash severity model, the variable *Speed,* which represent the average operating speed on the segment, was found to increase the crash severity level, which is intuitive since the severity of the collisions are mainly dependent on the kinetic energy when they conflict with each other.

3.3.1.9 Signal settings. There are two significant signal settings, *LeftGreSplit* and *SmallCycle* in the OD crash probability model and one variable *LargeCycle* in the conditional crash severity model. The variable *LeftGreSplit* reflects the proportion of the minimum green time for the involved left turn movements. The larger this proportion is, the larger the left turn traffic demand and therefore a higher probability of crashes. The two binary indicators of signal cycle length were difficult to interpret.

3.3.1.10 Geometry settings. In the hourly crash probability model, the *LaneNumb(3,4)* indicating the 3-lane and 4-lane cases, was found to increase the probability of OD crashes, which is intuitive. The more lanes on the through movement, the more through vehicle fleets and left-turn vehicles that will need to cross. The number of lanes on the through movement also may increase the crash severity as the variable *LaneNumb(2,3,4)* is positive and significant.

The other geometry-related settings that increase OD crash probability include *FlashYellow*, indicator of the existence of a flashing yellow yield sign, and *Curved*. The OD crash probability for the latter one was intuitive while the former one, *FlashYellow*, was contrary to expectations. One explanation is that flashing yellow yield signs are usually set at intersections where there are many left turn traffic demands, thus causing an endogeneity problem. The effect of *FlashYellow* does not reflect its safety benefit but rather the excessive left turn demands at such locations.

3.3.3 Safety Effects on Right-Angle Collision

Table 3.6 shows the estimation results of the hourly probability of RA crashes and Table 3.7 shows the estimation results of the conditional probability of RA severe crashes. Compared to SD and OD crashes, there were fewer RA crashes and therefore less significant safety effects were identified.

Like the previous two models, the variable that served as the exposure factor, *logVol*, also was highly significant and positively related to the probability of RA crashes. The other significant variable in the crash probability model was *SpeedMain* (the average speed on the main street), which reduced the RA crash probability. This could be explained in the way that higher speeds on the main street are an indication of less traffic on the side street. The variable *ExcluRightTurn* was found to improve safety, although it was found to be merely significant. The variable *Cycle* was found to be negatively related to RA crash probability.

In the conditional crash severity model, two geometry settings, *HighSkew* and *FlashYellow* were found to significantly affect RA crash severity. When the intersection was highly skewed (greater than 60 degrees), the crash tended to be more severe; when there were flashing yellow yield signs, the crashes tended to be less severe.

TABLE 3.6
Parameter estimates of the hourly probability of right-angle crash model

Variable	Estimate	std.error	z.value	p.Value
Intercept	**-4.532**	1.090	-4.157	0.000
logVol	**0.558**	0.157	3.564	0.000
SpeedMain	**-0.048**	0.016	-3.054	0.002
ExcluRightTurn	-0.385	0.285	-1.350	0.177
Cycle	**-0.007**	0.003	-1.890	0.059

Note: Estimates of the significant variables are bolded.

TABLE 3.7
Parameter estimates of the conditional probability of right-angle severe crash model

Variable	Estimate	std.error	z.value	p.value
Intercept	**-0.988**	0.406	-2.435	0.015
HighSkew	**1.179**	0.563	2.093	0.036
FlashYellow	-1.206	0.713	-1.691	0.091

Note: Estimates of the significant variables are bolded.

3.3.4 Risk-Based SMS Application-Risk Profile

Although the implementation of the signalized intersection crash risk models is not feasible now due to limited data, the in-sample simulations did provide insight to the potential future implementation of the research results. There can be two major applications of the obtained risk-based safety models: (1) estimation of the time-dependent changes of crash risk over time; and (2) simulation of the potential safety benefits when certain countermeasures take place.

3.3.4.1 Risk profile. The major outcomes of the first application were sets of risk profiles. For each typical four-leg intersection, there are four approaching segments to the intersection, and therefore four different time-sensitive SD crash risk profiles. Similarly, the OD crash risk model estimated the probability of crashes between a given direction and the OD left turn traffic. There also were four separate crash risk profiles. The RA crash risk model estimated the crash probability between a given direction and any cross-street traffic; but to make the outputs consistent and avoid double counting of crash risks, "Direction-RA Crash" was defined as the crash risk between the direction's traffic and its clockwise directions' traffic. For example, the "E-RA Crash" is the crash risk between the eastbound traffic and the southbound traffic. The total probability of crash at a certain intersection should be the sum of all 12 crash probabilities (3 types of crashes * 4 directions).

Figure 3.13, Figure 3.14, and Figure 3.15 show examples of SD, OD, and RA crash risk changes across a 1-week period. The sample intersection is US 36 at Dan Jones Rd where the main street is east-west US 36. The sample period was June 1–7 (Friday through Thursday) in 2018. The obtained crash risks followed the general changing trends of the traffic volume. For the weekdays, the crash risks were higher; and for one day, the evening peak hours were more dangerous. In this example, the probability of SD and OD crashes were higher for the eastbound and westbound directions, which was where the main street of the case study area is located.

Figure 3.16 shows the crash risk (three types) profile across the 1-week period. SD crashes were the major source of crash risk, which was consistent with the real distributions of these three types of crashes. While SD crashes had the lowest probability of severe crashes
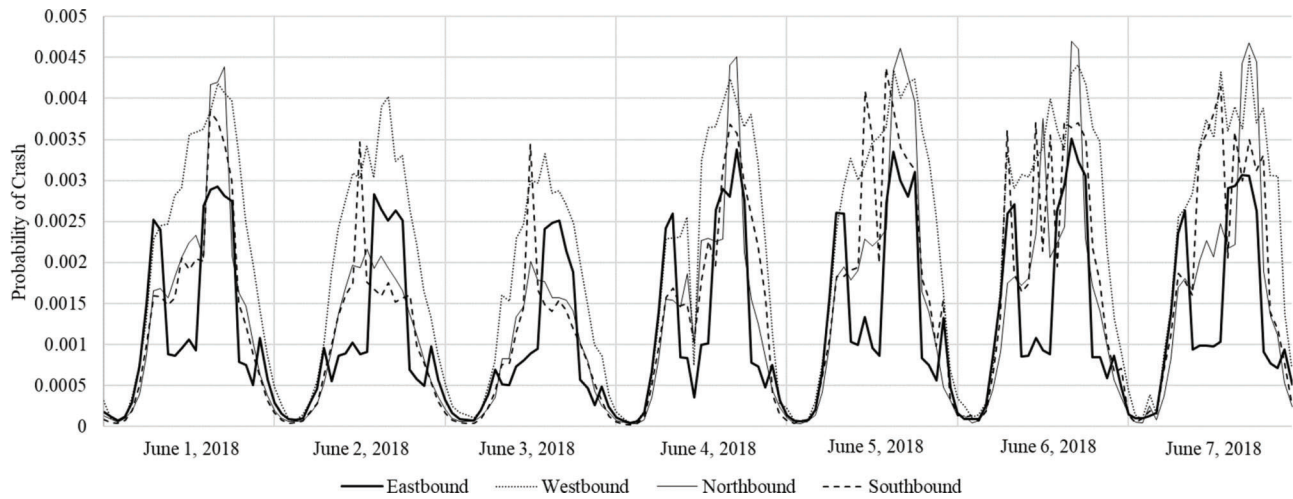
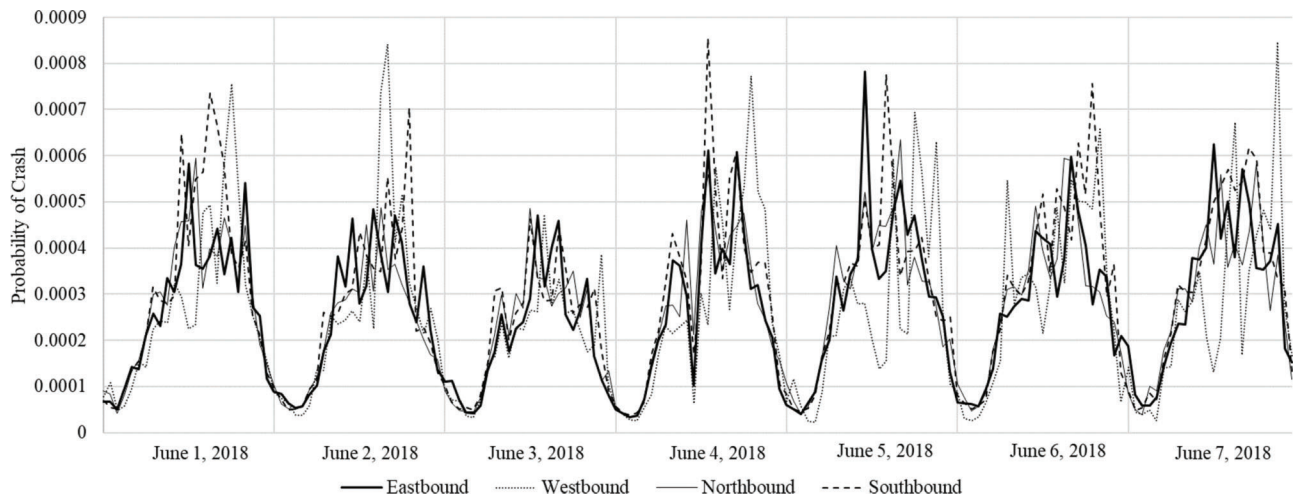**Figure 3.13** Same direction crash risk profile at intersection US 36 at Dan Jones Rd (example days).

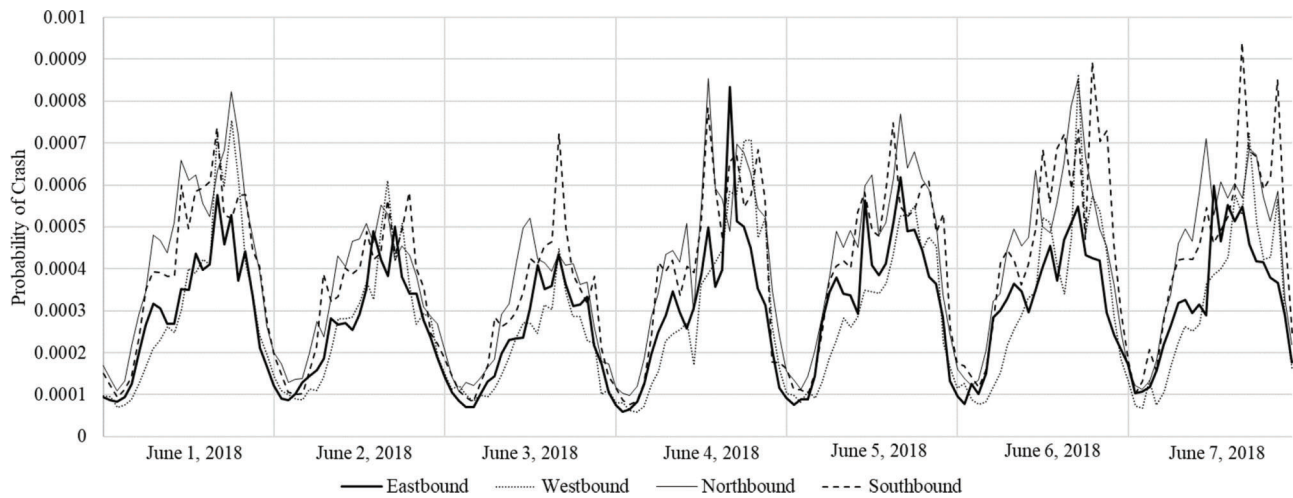**Figure 3.14** OD crash risk profile at intersection US 36 at Dan Jones Rd (example days).

Figure 3.15 Right angle crash risk profile at intersection US 36 at Dan Jones Rd (example 7 days).
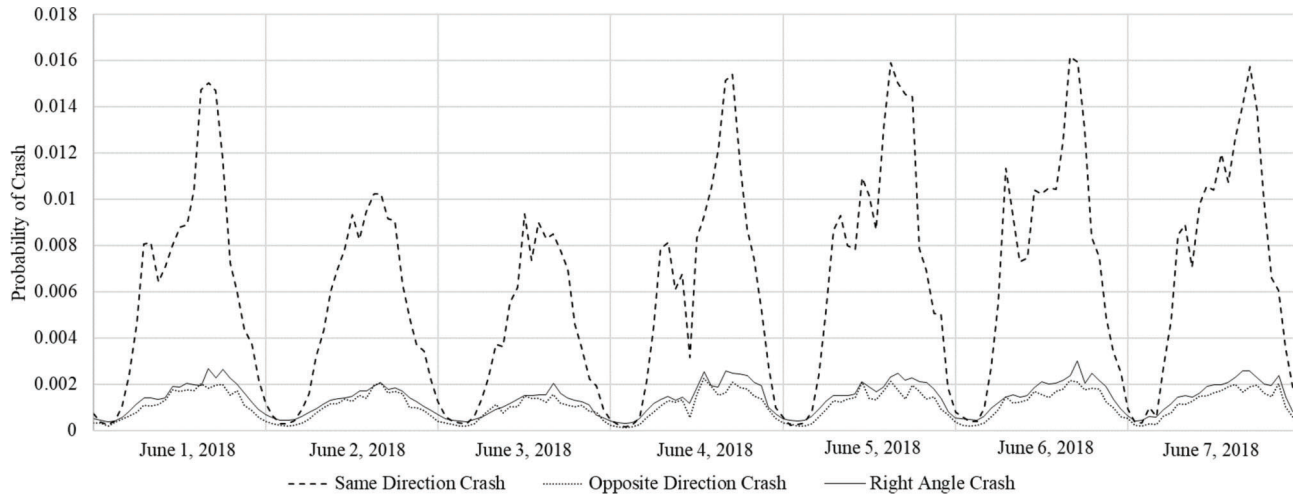
Figure 3.16 Crash risk profiles at intersection (US 36 at Dan Jones Rd) from June 1st to June 7th, 2018.
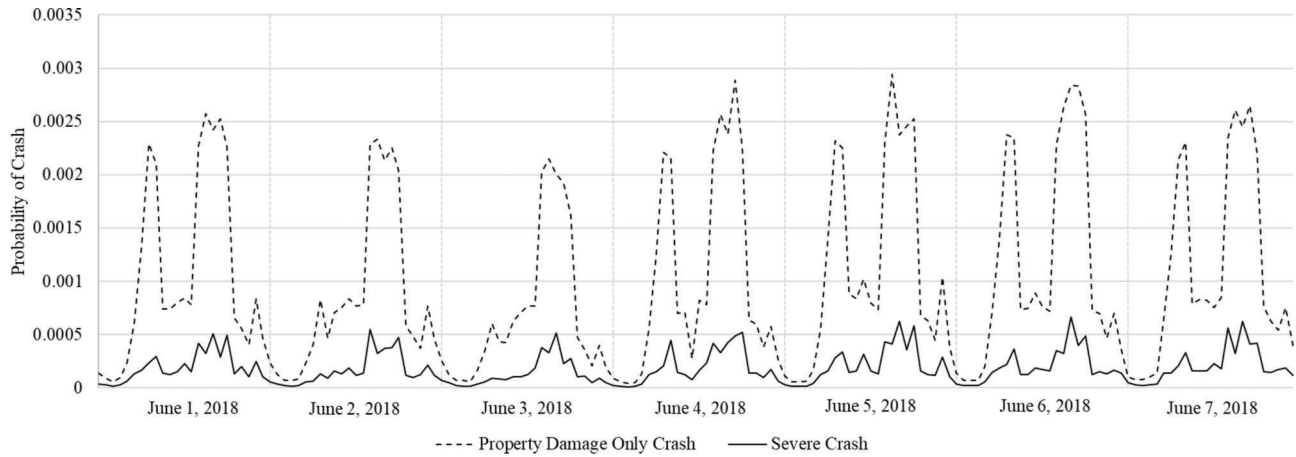

Figure 3.17 Decomposed same direction PDO/injury crash risk profile for Eastbound segment at intersection US 36 at Dan Jones Rd.
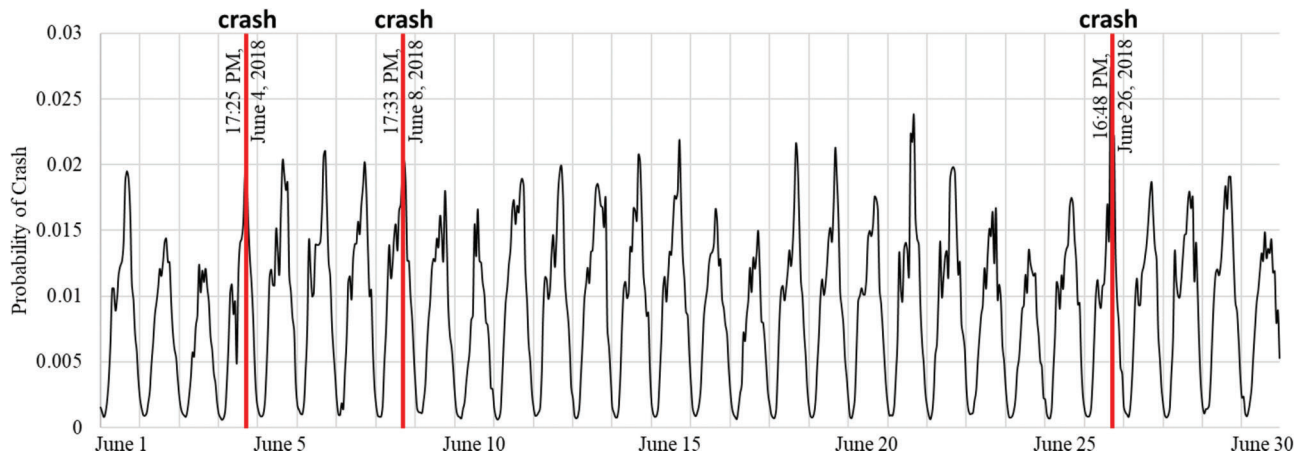

Figure 3.18 Predicted crash risk and observed crashes at intersection (US 36 at Dan Jones Rd) in June 2018.

(13.8%) compared to OD (37.0%) and RA (29.2%) crashes, they were more prevalent. The sample application of the conditional severe crash model is shown in Figure 3.17, where the eastbound SD PDO and severe crash risk profiles are compared.

Figure 3.18 shows the predicted total crash risk profile across the entire month of June 2018 aligned with the observed real occurrence of crashes at the sample intersection. The three crashes occurred around the predicted high crash risk periods, but the risk profile did not give obvious indication before a crash occurred (i.e., it is difficult to "predict" real crash occurrence). What the model provides is a prediction of the high-risk time periods.

3.3.4.2 Countermeasure example. The other potential application of the risk-based safety model is a simulation of the changes in safety performance based on the time-dependent factors. There are several significant time-dependent variables in the obtained models, but few of them can be controlled by traffic engineers. One of these variables is the offsets between coordinated intersections. It was identified in the SD crash probability model that when a traffic fleet arrived at the downstream intersection slightly earlier than the relative offset, the safety performance of that approaching segment improved. The offset effect is illustrated below with three coordinated sample intersections on US 36. Figure 3.19 shows the locations of the three targeted intersections and Table 3.8 shows their original coordination settings. The three intersections are consecutive signalized intersections across the main street US 36,

and they share the same coordination cycle of 120 seconds. However, there are three different offsets and green time split plans depending on the weekdays or weekends. The original offset settings did not fulfill the safest case of "*EarlyArr*" on either direction (eastbound or westbound).

Since traffic engineers were not aware of the difference in safety performance for different offset settings, the original offset settings were probably set solely based on traffic efficiency considerations. However, based on this study, there are several candidate coordination plans that could maintain similar high efficiency as well as good safety performance. In the example below, we simulated the potential changes of crash risks if all the coordinated hours (06:00–22:00) could fulfill the *EarlyArr* conditions for the westbound traffic. The westbound traffic was selected because it had the largest volume and there was no way to fulfill the *EarlyArr* conditions for both directions (the sum of the relative offsets between two intersections should always equal the cycle length).

The simulation results for the two westbound approaching segments (from Avon Marketplace to Dan Jones Rd and from Dan Jones Rd to Beechwood Dr) are shown in Figure 3.20 and Figure 3.21. Compared to the original coordination offset settings, if the offsets could be optimized during all the coordinated hours, the corresponding SD crash risk could drop significantly. Considering that most of the crashes are SD crashes and that the westbound segment is on the main street with the largest volume, the safety improvement at the intersection level could be considerable.
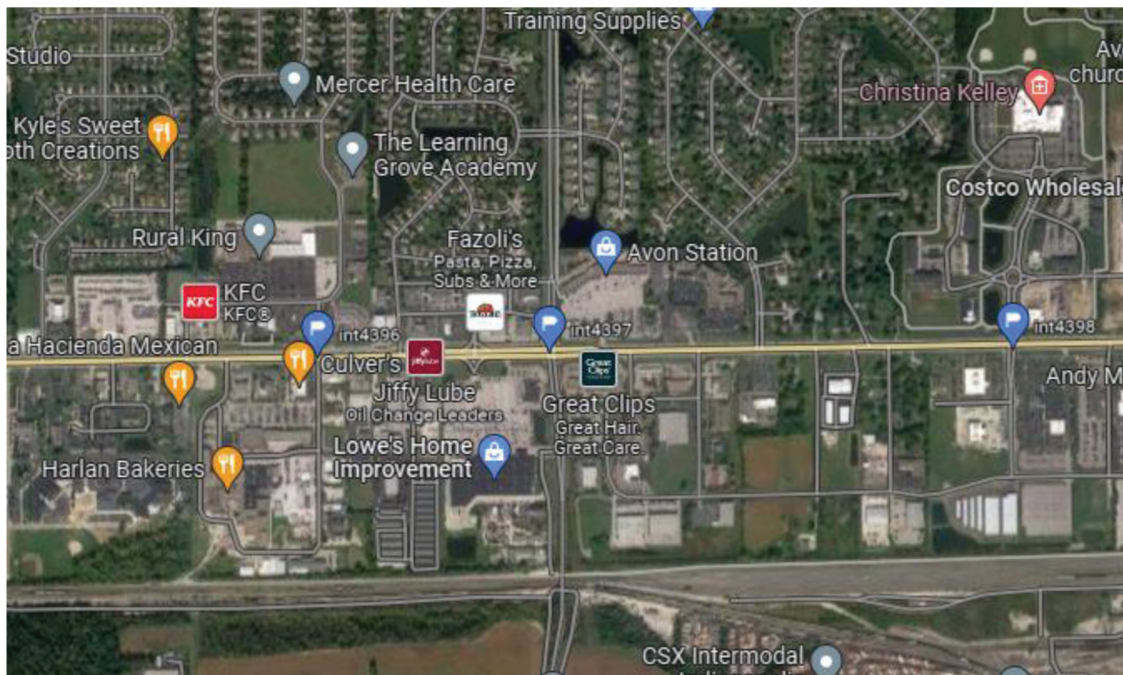

Figure 3.19 Intersection locations (example of three coordinated intersections on US 36).

TABLE 3.8
Original coordination settings (example of three coordinated intersections on US 36)

| Main Street	US 36		
Side Street	Beechwood Dr	Dan Jones Rd	Avon Marketplace
Coordination Cycle		120 seconds	
Offset1	27	59	11
Offset2	53	104	60
Offset3	88	14	88
Eastbound Relative Offset1	–	32	72
Eastbound Relative Offset2	–	51	76
Eastbound Relative Offset3	–	46	74
Westbound Relative Offset1	88	48	–
Westbound Relative Offset2	69	44	–
Westbound Relative Offset3	74	46	–
Eastbound Reference Travel Time	–	25	50
Westbound Reference Travel Time	25	50	–
Offset1 Take Effect Time		Weekday 06:00–9:00; Weekend 06:00–8:00	
Offset2 Take Effect Time		Weekday 9:00–14:00, 19:00–22:00	
		Weekend 8:00–14:00, 19:00–22:00	
Offset3 Take Effect Time		Weekday 14:00–19:00; Weekend 14:00–9:00	
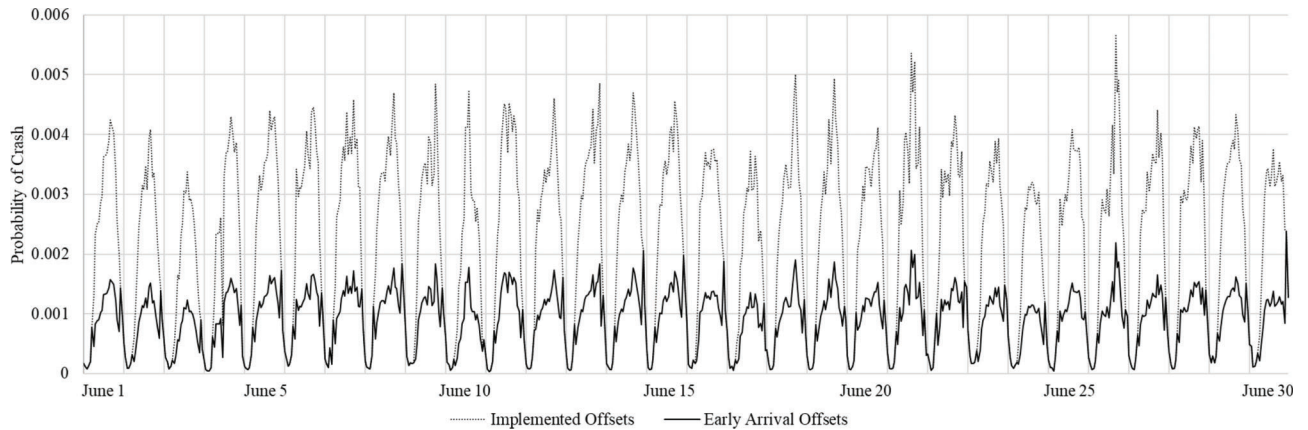
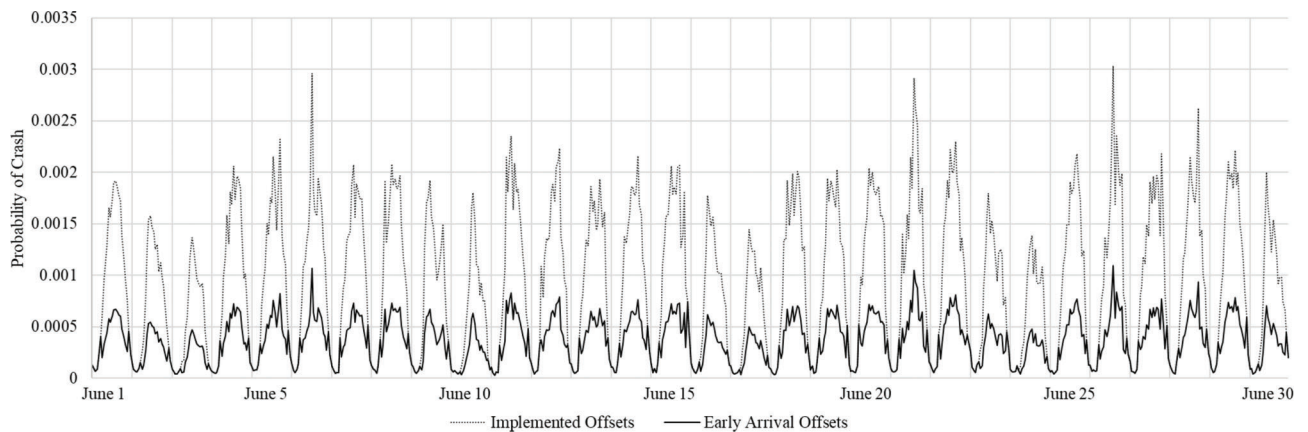**Figure 3.20** Simulation results for westbound segment at US 36 at Dan Jones Rd.

Figure 3.21 Simulation results for westbound segment at US 36 at Beechwood Dr.

4. IMPLEMENTATION

The implementation of the proposed risk-based safety management framework includes three components.

1. *Enquiring and accessing data sources.* There are several data sources outside INDOT's safety management system database, such as hourly weather conditions, INRIX speed data, and, for signalized intersections, detector/phase data maintained by the Traffic Management Center.
2. *Processing data.* The raw data should be pre-processed including formatting, filtering, and imputing, and then processed with risk models to estimate the risk (probability of crash and severe crash).
3. *Applying the processed data in the current SMS supplemented with risk-based elements.*
   a. The *crash-based* SMS tool, SNIP, provides candidate road segments and intersections.
   b. The *risk-based* SMS identifies conditions that cause temporarily elevated crash risk. These conditions may be used as triggers of the operational intervention.
   c. The provided prototype *risk-based* tool estimates the number of crashes on roads in periods with potential operational interventions; no interventions applied yet.
   d. The *crash-based* SMS tool, RoadHAT, is used to perform a benefit/cost analysis of the considered operational countermeasures.

4.1 Prototype Tool for Risk-Based Safety Management

A risk-based SMS application was created to identify risk-contributing factors (Figure 4.1). This tool allows the user to select road segments and evaluate the crash risk over a user-selected period. The program calculates and displays the hourly crash risk for the selected

segments during a user-specified period. One of its critical components for analysis is a ranking of the risk-contributing factors. The user can filter the selection period to investigate further the effect of different factors on the crash risk. This allows the user to propose safety interventions that might not be revealed using the traditional approach because some effects might be diluted when using long aggregation periods. The introduction and detailed step-by-step implementation of the risk-based SMS prototype program are included in Appendix B. Risk-Based SMS Application RMT.

The current SMS identifies HCLs. The next step involves identifying high-risk conditions that can be affected with a target safety countermeasure. One year of hourly data on selected rural freeways was acquired, processed, and analyzed. Then, the crash risk is estimated using Equation 2.5 through Equation 2.7. High-risk conditions are assumed to happen at the 50th percentile, which may be adjusted by the end user based on their criteria and experience. Another criterion to filter data for analysis is high-speed conditions. The 50th percentile of average travel speed is used here as well. The summary statistics of the 2018 sample are presented in Table 4.1.

4.2 Input Data Handling

Data on fixed and time-dependent risk factors are essential for using the proposed risk-based SMS tool. The different data required can be classified into three types: (1) data that are available in-house at INDOT; (2) data that are available to INDOT through Purdue CRS; and (3) data not currently available that must be collected

Figure 4.1 Screenshot of prototype risk-based SMS tool.

before using the tool. Table 4.2 presents a classification of the specific data sources in these three cases. Crash records are available in-house to INDOT via the ARIES data set. CRS has developed various analytical tools (e.g., CLIP) to refine and assign individual crashes to target road sites. Operating speeds are available in-house and via Purdue CRS, depending on the data set name. NPMRDS is available in-house while INRIX speed data are available to Purdue CRS. Data preparation programs in SAS and R have been developed to use these two data sources. Traffic counts are available in-house to INDOT. Road characteristics are partially available to INDOT in-house. Additional variables may require manual data collection by the end user via online repositories such as Google Earth. Weather conditions can be accessed via the INClimate group and NOAA's online data query tools. Purdue CRS has developed additional programs to prepare large amounts of weather data for statistical analysis of traffic safety.

4.3 Illustrative Examples

4.3.1 Risk Profiles

The models presented in Table 2.1 and Table 2.2 can be used to supplement crash-based safety management by estimating the short-term crash risk performance of a target road segment. This estimation may improve the current safety management tasks. Specific applications are risk-based network screening, operational counter-measure evaluation, and crash risk visualization tools for safety audits.

An example visualization tool that may benefit safety audits is the risk profile. Using historical crash data, the expected hourly risk and observed crashes can be plotted against time. Two example risk profiles were developed and are presented in Figure 4.2 and Figure 4.3. These profiles use data from a section of I-65 northbound in Bartholomew County. The 0.25-mile segment has an auxiliary entering lane, a wide-ditched median, a median cable barrier with a relatively small offset, and a roadside guardrail on the right shoulder.

Figure 4.3 depicts data from January 2014 when seven crashes occurred. A cyclic baseline risk was perceived when the crash risk trended (solid blue line). This base risk was then altered by rapid changes in travel speeds or weather events. For example, four out of the seven crashes occurred during adverse weather conditions, specifically, two on January 2, one on January 6, and one on January 16. Figure 4.3 presents the risk profile using data from April 2017 when four crashes occurred. It is worth mentioning that the fourth crash was excluded from the analysis as it was a secondary crash.

Risk profiles and other summary statistics were used to develop the safety dashboards available in the risk-based SMS tool. Such dashboards can be a valuable tool for safety management when preparing safety audits or evaluating operational countermeasures. Using the calibrated parameter estimates (e.g., the downward speed trend or a reduction in the speed limit and the amount of time a given operational counter-measure is active), can produce the expected safety benefit. This estimation then can be compared to the actual benefit once the real-time data are available.

4.3.2 Countermeasure Evaluation

4.3.2.1 Variable speed limits. This section presents an example of the risk-based SMS and its tools for implementing variable speed limits (VSLs) on rural freeways. Available data sources are presented as well

TABLE 4.1
Distribution of crash risk and travel speed in 2018

Statistic	Speed (mph)	Hourly Crash Risk (%)
Mean	65.01	0.0089
Std. Deviation	65.58	0.0055
Mode	65.92	0.0023
100% Max	77.17	2.0371
99%	67.67	0.0738
95%	66.92	0.0170
90%	66.58	0.0124
75% Q3	66.08	0.0083
50% Median	65.58	0.0055
25% Q1	64.83	0.0036
10%	63.67	0.0025
5%	61.25	0.0021
1%	54.75	0.0014

TABLE 4.2
Accessibility of input data for prototype risk-based SMS tool

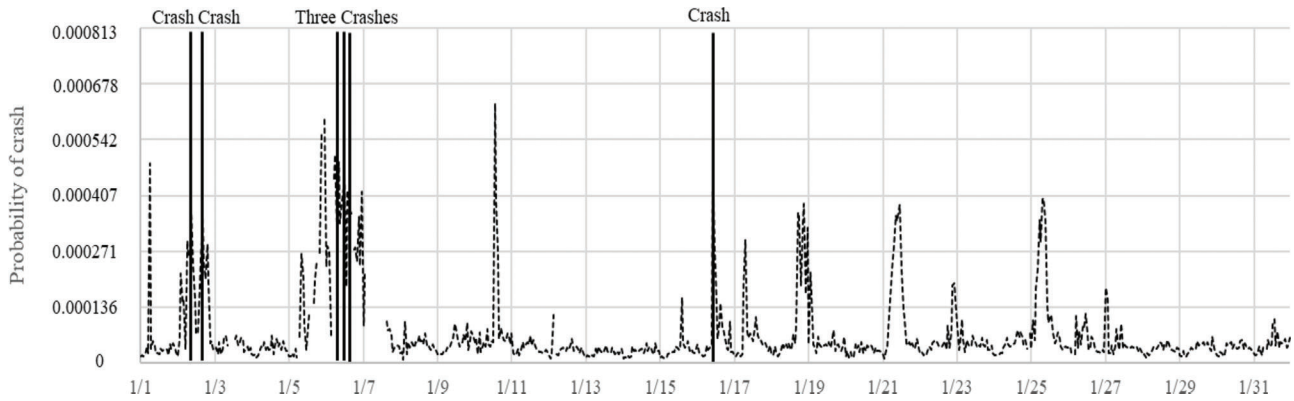| Input Data Type	Source	Accessibility
Crash Records	Automated Reporting Information Exchange System-(ARIES)	In-house, Purdue CRS
Operating Speeds	National Performance Management Research Data Set (NPMRDS); INRIX Inc	In-house, Purdue CRS
Traffic Counts	Traffic Count Database System (TCDS); High-Resolution Detector Data (HRDD)	In-house
Road Characteristics	Road Network Data (RND); Google Earth Imagery	In-house, Collect
Weather Conditions	Indiana State Climate Office (INCLimate); National Oceanic and Atmospheric Administration (NOAA)	Purdue CRS

Figure 4.2 Crash risk profile for I-65 northbound near Exit 76 B in January 2014.
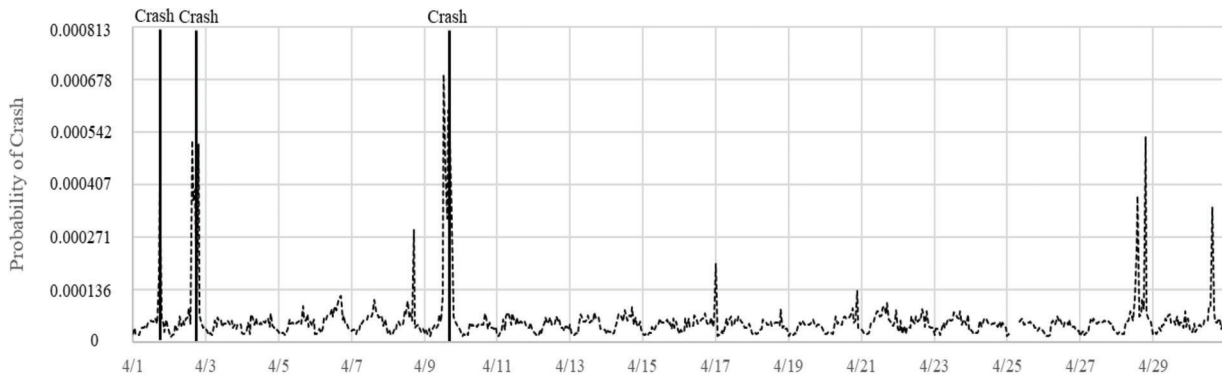

Figure 4.3 Crash risk profile for I-65 northbound near Exit 76 B in April 2017.

as the components of the prototype application tool and a sample of high-crash roads with high-risk conditions. Finally, the estimated effectiveness of VSLs is demonstrated and discussed.

Table 4.3 presents the descriptive statistics of the crash risk factors under high-risk, high-speed conditions. In addition, the relation to the population mean (the baseline) is shown. Regarding exposure variables, the short-term traffic volume and AADT values were higher than the population under the selected conditions. High-risk, high-speed observations tend to happen in segments with moderate to sharp curves, higher proportions of cable median barriers, outside guardrails, concrete pavements, and ramps. In addition, such observations present a lower number of overpasses and median guardrails. Regarding time-dependent factors, high-risk high-speed observations tend to happen under light rain conditions with growing queues. Weekends and daytime are overrepresented under these conditions.

To evaluate the implementation of VSLs, a subset of high-risk high-speed conditions was used. This subset accounts for 18.84% of the 2018 data (878,217 observations). The effectiveness of VSLs on rural highways was recently evaluated (Avelar et al., 2021; El Esawey et al., 2022). They produced two CMFs that

match the grouping of injury severity levels used in this study. A 29% reduction in PDO crashes (CMF = 0.71) was reported by Avelar et al. (2021) while a larger 32% reduction in KABC crashes (CMF = 0.68) was reported by El Esawey et al. (2022).

Using the subset of short-term risk estimates and the two high-quality CMFs, the expected number of crashes before and after the implementation of VSLs could be accurately estimated. The premise that VSLs are most effective under high-risk high-speed conditions is based on existing research on posted speed limits that suggest their maximum safety effect under non-congested conditions close to free flow (Tarko et al., 2019). The expected number of PDO crashes before the implementation of VSLs was 67.5, and the expected number of KABC crashes was 7.3. After the implementation of VSLs, 47.91 PDO and 4.9 KABC crashes are expected.

In practice, the end user of the risk-based SMS tool would have to input the data needed to calculate the hourly risks, select the target risk level, and identify the conditions for which the target countermeasure is most effective. The tool estimates the risk reduction factor for the selected segment by running both scenarios with and without countermeasures. This result is used in the current SMS system tools

TABLE 4.3
Descriptive statistics of risk factors under high-risk high-speed observations (N = 878,217)

Variable	Mean	Std. Dev.	Minimum	Maximum	Compared to Baseline
Hourly Volume (veh/day)	1.0204	0.4223	0.0150	3.1610	Higher
AADT Passenger Cars (veh/day)	24.6584	6.1124	8.2390	39.6390	Higher
AADT Heavy Trucks (veh/day)	10.7800	2.6585	4.9130	15.5000	Higher
Overpass	0.1010	0.3014	0.0000	1.0000	Lower
Moderate Curve (5.5–13.9 degrees)	0.1234	0.3289	0.0000	1.0000	Higher
Sharp Curve (14.0–28 + degrees)	0.1519	0.3590	0.0000	1.0000	Higher
Cable Median Barrier	0.8856	0.2729	0.0000	1.0000	Higher
Guardrail Median Barrier	0.0336	0.1117	0.0000	1.0000	Lower
Guardrail Roadside Barrier	0.1942	0.3085	0.0000	1.0000	Higher
Offset ≤ 11 ft	0.4838	0.4997	0.0000	1.0000	Higher
Median Hazard	0.1562	0.3756	0.0000	2.0000	Lower
Roadside Hazard	0.2628	0.4575	0.0000	2.0000	Higher
Concrete Pavement	0.1686	0.3512	0.0000	1.0000	Higher
Entering Ramp	0.0360	0.1511	0.0000	1.0000	Higher
Exiting Ramp	0.0291	0.1106	0.0000	0.7584	Higher
Speed Limit = 65 mph	0.1926	0.3944	0.0000	1.0000	Higher
Outside Shoulder Width (ft)	11.4541	0.7220	9.1250	13.9800	Higher
Light Rain ≤ 0.1 inch/h	0.0840	0.2774	0.0000	1.0000	Higher
Temperature ≤ 32F°	0.1539	0.3608	0.0000	1.0000	Lower
Icy Conditions	0.0102	0.1007	0.0000	1.0000	Lower
Std. Dev. of Speed	1.2389	0.7006	0.0000	12.2526	Lower
Speed Trend	-0.0309	0.1592	-3.2093	2.0074	Lower
Growing Queue	0.3409	0.4740	0.0000	1.0000	Higher
Average Speed (mph)	66.1246	0.4628	65.5833	74.0000	Higher
Friday	0.1777	0.3823	0.0000	1.0000	Higher
Sunday	0.1754	0.3803	0.0000	1.0000	Higher
Morning	0.2893	0.4534	0.0000	1.0000	Higher
Afternoon	0.4343	0.4957	0.0000	1.0000	Higher
Evening	0.2448	0.4299	0.0000	1.0000	Lower

Note: Light rain is defined as precipitation lower than 2.5 mm or 0.098 inch per hour.

(RoadHAT 4D) to evaluate projects and perform economic analysis.

4.4 Compatibility with Existing Safety Management Tools

Since the crash risk assessment interval is small relative to the accident frequency, it is safe to use the annual cumulative hourly risk of crash as the expected number of crashes per year.

This allows the program to make such an estimate possible with and without a countermeasure and thus calculate the RMF that can be associated with the CMF for that countermeasure in the selected segments taking into account the historical weather patterns in that particular region.

The resulting risk reduction factor can be used in the RoadHAT 4D Economic Evaluation Form 5 module as one of the countermeasures to be evaluated by copying the CMF.

4.5 Pedestrian Safety Management

Pedestrians are the most vulnerable road users as they tend to endure more severe injuries in any collision with a vehicle. While innovative safety countermeasures promise improved pedestrian safety, a careful analysis of local conditions is required before selecting proper corrective measures.

JTRP Project SPR-4437 (Ahmad et al., 2021) focused on first developing a methodology to identify roads and areas in Indiana where the frequency and severity of pedestrian collisions were higher than the acceptable level and then selecting effective counter-measures to mitigate or eliminate the safety-critical conditions. As part of SPR-4437, and particularly relevant to the risk-based SMS framework, two sets of models were developed to facilitate their road-focused analysis: (1) pedestrian crossing activity level models to fill the gap in pedestrian traffic data, and (2) crash probability and severity models to estimate the risk of pedestrian crashes around urban intersections in Indiana.

4.6 Implementation Roadmap

The primary benefits of this study are an assessment of the feasible means to implement a crash risk-based approach to the selection and implementation of infrastructure improvements and possible operational safety improvement actions using data collected over a much shorter time frame than is currently possible.

These factors can potentially reveal static and time-varying conditions that alter crash risk. To make the components of the new risk-based SMS applicable to the current Indiana safety management programs, three actions must happen.

1. Incorporate the risk-based SMS data sets into INDOT's Information Technology practice, including their storage, renewal, and management.
2. Develop data preparation components to produce the inputs required by the safety management tools and make them available to the end users.
3. Modify the existing road screening, safety projects scoping, and cost-effectiveness evaluation tools to bring new input into the decision-making process for safety infrastructure.

Fundamental knowledge that supports these three actions is presented in this report. There are two possible avenues for implementing the results. First, the current recurring service arrangement with the project owner if the implementation effort turns out to be limited. Second, a new project with a stated implementation scope and budget needs to be approved within the JTRP process. This project may be a follow-up effort or may partially overlap with the project proposed here to speed up the implementation.

5. CLOSURE

5.1 Summary of Findings

This study is the first comprehensive research attempt to analyze the disaggregated crash risk on rural freeway segments and signalized intersections. All the time-dependent potential crash-risk factors (hourly traffic, speed features, weather conditions, and signal controls) were tested in several sequential logit models that can estimate both the probability of crash and the conditional probability of severe crash.

For rural freeway segments, the hourly crash risk was found to increase depending on the hourly volume, AADT by vehicle type, horizontal curves, barriers, lower speed limits, the standard deviation of speed, the interaction of low-intensity rain and freezing temperatures, auxiliary lanes, downtrend speeds, and congestion. On the other hand, overpassing roads, average speed, and uptrend speed were found to enhance safety. The conditional probability of a severe outcome was increased by mild curves, the average speed, the standard deviation of speed, downtrend speeds, and intermediate congestion while lighting and lower speed limits were found to reduce crash severity.

For signalized intersections, crashes were divided into three types: same-direction (SD), opposite-direction (OD), and right-angle (RA) crashes based on different crash generating scenarios. The significant variables were diverse across the three models; but in general, light rain, the logarithm of the square root of the involved traffic volume product, and the standard deviation of speed at the involved intersection leg were found to increase crash risk. Coordination settings were found to affect SD crash probability. Other factors including curved intersection legs, the number of lanes, and exclusive right turn lanes were found to be significant but had diverse effects across the different crash types.

The obtained hourly probability models were used to generate crash risk profiles which display and compare the changes of crash risk over time and to estimate the potential safety benefits if certain time-dependent countermeasures, such as variable speed limits, take effect.

5.2 Implications for Safety Management Practice

The concept of risk-based safety management and its tools have the potential to supplement the existing crash-based safety management practice. Thus, the development of risk-based SMS will not replace the current crash-based SMS but instead will supplement it with features not covered by the existing SMS. No disruption, but rather gradual expansion of the existing system, is envisioned.

On the one hand, disaggregated safety analysis can estimate the short-term variation of crash risk on a specific road segment. The estimated impacts, including static and time-varying factors, can be used to quantify changes in the crash risk caused by the implemented safety countermeasures. On the other hand, the role of count-based crash models remains. Once an HCL road is identified as requiring additional safety analysis and potential intervention, the existing count-based tools, such as safety performance functions, can be used.

Further analysis of risk estimated in short intervals would allow the identification of periods with high-risk conditions and the presence of specific risk factors. Such an approach is expected to help obtain robust estimates of the actual effect of safety countermeasures, particularly those with an operational component (e.g., variable message signs, variable speed limits, in-vehicle messages, and other active traffic control devices). The proposed approach to estimating both the crash risk and the effects of safety countermeasures in short intervals supplemented with long-term safety analysis offers a more comprehensive approach than the one offered solely by crash count analysis.

5.3 Future Research Directions and Systemwide Implementation Suggestions

The future research of time-dependent safety management mainly will depend on the availability of new data as well as emerging data sources. For the given data, the research team exploited the statistical tools to investigate and establish the relationship between risk factors and crashes. If more data become available (for example, if more years of detector data becomes available), which is expected since INDOT is connecting the data from increasingly more signalized intersections to the Traffic Center database), there might be

more significant time-dependent factors for signalized intersections models since the crash samples for OD crashes (146) and RA crashes (72) are still very limited.

The other potential research direction is an analysis of the disaggregated crash risk using black-box machine learning methods, which could provide the ability to make better predictions possible but could lessen the interpretability of the model. The machine learning approach will not be able to produce safety improvement suggestions for engineers since all the input variables are masked in the neural network, but it could help identify high-risk time periods. The high-risk time periods also may have similar operational conditions that could be classified.

The risk-based safety management system presented here is a data demanding safety analysis tool that considers crash risk decomposed at the hourly level with various sources of time-dependent risky factors. The primary challenge of the systemwide implementation of such an analysis tool is data management, which requires two steps: (1) safe and timely connection between time-dependent risk factors and INDOT safety management system, and (2) automatic preprocessing of the data. In this current study, these two steps were accomplished by the research team with the sample data, but if INDOT desires a systemwide application of the analysis framework, agency-wide professional database management and maintenance efforts will be required.

REFERENCES

AASHTO. (2010). *Highway safety manual*. American Association of State Highway and Transportation Officials.

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, M. F., & Hsia, L. (2004). Predicting freeway crashes from loop detector data by matched case-control logistic regression. *Transportation Research Record*, *1897*(1), 88–95. https://doi.org/10.3141/1897-12

Ahmed, M. M., & Abdel-Aty, M. A. (2011). The viability of using automatic vehicle identification data for real-time crash prediction. *IEEE Transactions on Intelligent Transportation Systems*, *13*(2), 459–468.

Ahmed, M. M., Abdel-Aty, M., & Yu, R. (2012). Assessment of interaction of crash occurrence, mountainous freeway geometry, real-time weather, and traffic data. *Transportation Research Record*, *2280*(1), 51–59. https://doi.org/10.3141/2280-06

Ahmad, N. S., Pineda-Mendez, R., Alqahtani, F., Romero, M., Thomaz, J., & Tarko, A. P. (2021). *Effective design and operation of pedestrian crossings* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2021/32). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317438

Avelar, R. E., Park, E. S., Ashraf, S., Dixon, K. K., Li, M., & Dadashova, B. (2021, May). *Developing crash modification factors for variable speed limits* (Publication No. FHWA-HRT-21-053). Texas A&M Transportation Institute. https://rosap.ntl.bts.gov/view/dot/57267

Capparuccini, D. M., Faghri, A., Polus, A., & Suarez, R. E. (2008). Fluctuation and seasonality of hourly traffic and accuracy of design hourly volume estimates. *Transportation Research Record*, *2049*(1), 63–70.

Cascetta, E. (2009). *Transportation systems analysis: models and applications* (2nd ed.). Springer.

Chen, E., & Tarko, A. P. (2012). Analysis of crash frequency in work zones with focus on police enforcement. *Transportation Research Record: Journal of the Transportation Research Board*, *2280*(1), 127–134. https://doi.org/10.3141/2280-14

Cheng, X., Lin, W., Liu, E., & Gu, D. (2010). Highway traffic incident detection based on BPNN. *Procedia Engineering*, *7*, 482–489.

Dutta, N., & Fontaine, M. D. (2019). Improving freeway segment crash prediction models by including disaggregate speed data from different sources. *Accident Analysis and Prevention*, *132*, 105253. https://doi.org/10.1016/j.aap.2019.07.029

El Esawey, M., Sengupta, J., Babineau, J. E., & Takyi, E. (2022). Safety evaluation of variable speed limit system in British Columbia. *Journal of Transportation Safety & Security*, *14*(10), 1776–1797.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*, 167–181. https://doi.org/10.1016/j.tra.2015.04.003

FHWA. (n.d.). *Crash modification factors clearinghouse*. Retrieved July 26, 2017, from http://www.cmfclearinghouse.org/

Fitzpatrick, C. D., Rakasi, S., & Knodler, M. A., Jr. (2017). An investigation of the speeding-related crash designation through crash narrative reviews sampled via logistic regression. *Accident Analysis and Prevention*, *98*, 57–63. https://doi.org/10.1016/j.aap.2016.09.017

Hossain, M., Abdel-Aty, M., Quddus, M. A., Muromachi, Y., & Sadeek, S. N. (2019). Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. *Accident Analysis & Prevention*, *124*, 66–84.

Huang, T., Wang, S., & Sharma, A. (2020, February). Highway crash detection and risk estimation using deep learning. *Accident Analysis & Prevention*, *135*, 105392.

Lee, C., Hellinga, B., & Saccomanno, F. (2003). Real-time crash prediction model for the application to crash prevention in freeway traffic (Paper No. 03-2749). *Transportation Research Record*, *1840*(1), 67–77. https://doi.org/10.3141/1840-08

Li, L., Qu, X., Zhang, J., & Ran, B. (2017). Traffic incident detection based on extreme machine learning. *Journal of Applied Science and Engineering*, *20*(4), 409–416.

Lingras, P., Sharma, S. C., Osborne, P. D., & Kalyar, I. A. (2000). Traffic volume time-series analysis according to the type of road use. *Computer-Aided Civil and Infrastructure Engineering*, *15*(5), 365–373.

Martin, J.-L. (2002, September). Relationship between crash rate and hourly traffic flow on interurban motorways. *Accident Analysis & Prevention*, *34*(5), 619–629.

Meyer, J., Becker, H., Bösch, P. M., & Axhausen, K. W. (2017). Autonomous vehicles: The next jump in accessibilities? *Research in Transportation Economics*, *62*, 80–91. https://doi.org/10.1016/J.RETREC.2017.03.005

Mohammadnazar, A., Mahdinia, I., Ahmad, N., Khattak, A. J., & Liu, J. (2021). Understanding how relationships between crash frequency and correlates vary for multilane rural highways: Estimating geographically and temporally weighted regression models. *Accident Analysis & Prevention*, *157*, 106146.

Naik, B., Tung, L.-W., Zhao, S., & Khattak, A. J. (2016). Weather impacts on single-vehicle truck crash injury

severity. *Journal of Safety Research*, 58, 57–65. https://doi.org/10.1016/j.jsr.2016.06.005

Oh, C., Oh, J.-S., & Ritchie, S. G. (2005). Real-time hazardous traffic condition warning system: Framework and evaluation. *IEEE Transactions on Intelligent Transportation Systems*, 6(3), 265–272.

Oh, C., Oh, J.-S., Ritchie, S., & Chang, M. (2001). Real-time estimation of freeway accident likelihood. *Journal of Transportation Engineering*, 131, 358–363.

Pande, A., & Abdel-Aty, M. (2006). Assessment of freeway traffic parameters leading to lane-change related collisions. *Accident Analysis & Prevention*, 38(5), 936–948.

Roshandel, S., Zheng, Z., & Washington, S. (2015). Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis. *Accident Analysis & Prevention*, 79, 198–211.

Sekuła, P., Marković, N., Vander Laan, Z., & Sadabadi, K. F. (2018). Estimating historical hourly traffic volumes via machine learning and vehicle probe data: A Maryland case study. *Transportation Research Part C: Emerging Technologies*, 97, 147–158.

Sharma, A., Ahsani, V., & Rawat, S. (2017). *Evaluation of opportunities and challenges of using INRIX data for real-time performance monitoring and historical trend assessment* (NDOT Research Report SPR-P1(14) M007). Nebraska Department of Transportation.

Talebpour, A., & Mahmassani, H. (2016). Influence of connected and autonomous vehicles on traffic flow stability and throughput. *Transportation Research Part C: Emerging Technologies*, 71, 143–163. https://doi.org/10.1016/j.trc.2016.07.007

Tarko, A. (2009). Modeling drivers' speed selection as a trade-off behavior. *Accident Analysis and Prevention*, 41(3), 608–616. https://doi.org/10.1016/j.aap.2009.02.008

Tarko, A. P., Guo, Q., Pineda-Mendez, R., & Romero, M. A. (2021). *Using emerging and extraordinary data sources to improve traffic safety* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2021/04). https://doi.org/10.5703/1288284317283

Tarko, A. P., & Kanodia, M. (2004). *Hazard elimination program–Manual on improving safety of Indiana road intersections and sections; Volume 1: Research report and Volume 2: Guidelines for highway safety improvements in Indiana* (Joint Transportation Research Program No. FHWA/IN/JTRP-2003/19). https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1583&context=jtrp

Tarko, A. P., Pineda-Mendez, R., & Guo, Q. (2019). *Predicting the impact of changing speed limits on traffic safety and mobility on Indiana freeways* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2019/12). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284316922

Tarko, A. P., Romero, M., Thomaz, J., Ramos, J., Sultana, A., Pineda, R., & Chen, E. (2016). *Updating RoadHAT: Collision diagram builder and HSM elements* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2016/11). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284316334

Wang, L., Abdel-Aty, M., Lee, J., & Shi, Q. (2019). Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. *Accident Analysis & Prevention*, 122 (pp. 378–384).

Xiao, J., & Liu, Y. (2012). Traffic incident detection using multiple-kernel support vector machine. *Transportation Research Record*, 2324(1), 44–52. https://doi.org/10.3141/2324-06

Xu, C., Liu, P., Wang, W., & Li, Z. (2012). Evaluation of the impacts of traffic states on crash risks on freeways. *Accident Analysis and Prevention*, 47, 162–171. https://doi.org/10.1016/j.aap.2012.01.020

Xu, C., Tarko, A. P., Wang, W., & Liu, P. (2013). Predicting crash likelihood and severity on freeways with real-time loop detector data. *Accident Analysis & Prevention*, 57, 30–39.

Yi, Z., Liu, X. C., Markovic, N., & Phillips, J. (2021). Inferencing hourly traffic volume using data-driven machine learning and graph theory. *Computers, Environment and Urban Systems*, 85, 101548.

Yu, R., & Abdel-Aty, M. (2014). Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety Science*, 63, 50–56. https://doi.org/10.1016/j.ssci.2013.10.012

Yu, R., Wang, Y., Zou, Z., & Wang, L. (2020). Convolutional neural networks with refined loss functions for the real-time crash risk analysis. *Transportation Research Part C: Emerging Technologies*, 119(C), 102740.

Yuan, J., Abdel-Aty, M. A., Fu, J., Wu, Y., Yue, L., & Eluru, N. (2021). Developing safety performance functions for freeways at different aggregation levels using multi-state microscopic traffic detector data. *Accident Analysis & Prevention*, 151, 105984.

Zahedian, S., Sekuła, P., Nohekhan, A., & Vander Laan, Z. (2020). Estimating hourly traffic volumes using artificial neural network with additional inputs from automatic traffic recorders. *Transportation Research Record*, 2674(3), 272–282. https://doi.org/10.1177/0361198120910737

Zheng, Z., Ahn, S., & Monsere, C. M. (2010). Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis & Prevention*, 42(2), 626–636.

Zou, Y., & Tarko, A. P. (2016). *Performance assessment of road barriers in Indiana* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2016/12). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284316

APPENDICES

Appendix A. Traffic Volume Prediction Model

Appendix B. Risk-Based SMS Application RMT

APPENDIX A. TRAFFIC VOLUME PREDICTION MODEL

High-resolution data for traffic exposure is critical for the successful assessment of road safety. Tarko et al. (2021) found that hourly traffic volumes are critical to crash risk prediction. Other time-dependent crash risk factors include weather, operating speed, and seasonal variations. Although state highway agencies, including INDOT, routinely count traffic with permanent and coverage detector stations, only a small percentage of the road segments in the total road network are covered.

One solution is to input the missing data with predictive models. To address this need, this section describes an effort to evaluate the possibility of predicting hourly traffic volumes for different classes of road segments in Indiana using a practical method based on statistical regression enhanced with a simplified representation of traffic generation and routing through a road network. A regression model for hourly traffic volumes was developed. Among the significant predictors found in the model, there are temporal effects (time of day, day of week, season), road functional classification, land development that potentially feeds a road segment with traffic (travel propensity), and a route selection element (travel time excess index). The estimation results indicate that all the listed variables significantly affected the traffic volumes on the considered road segments. The proposed model provides a reasonable estimation of hourly traffic volumes where the system currently does not provide any data. Possible further improvements of the model are also discussed. Hourly traffic predictions can help highway agencies with system-wide analysis including safety management and traffic operations.

A.1 Past Research

Traffic exposure is the primary cash risk factor on any target road location. Martin (2002) reported a direct relationship between crash rates and severity and hourly traffic volumes on interurban roadways on different days of the week. On another note, Lingras et al. (2000) performed traffic volume time-series analysis based on different types of road use which should help travelers plan their trips efficiently, which in turn, should maximize the use of highway capacity. Capparuccini et al. (2008) found that design hourly volume estimates were less accurate for roads with greater variability in volume fluctuation, but the selection of one method of analysis over another for particular traffic pattern groups resulted in smaller error. Furthermore, Zahedian et al. (2020) incorporated permanent counts as a direct input to the model, thus accounting for spatiotemporal correlations between hourly link volumes.

A wide range of dynamic traffic assignment models exist that aim to update the matrix using time-varying traffic counts. In recent years, advancements in machine learning-based methods, as well as the availability of large-scale data sets, such as probe vehicle data, have provided the opportunity to approach the link flow estimation problem from different perspectives. Sekuła et al. (2018) introduced an ANN-based regression approach that estimates hourly traffic volume using multiple data sources, such as vehicle probe counts and speeds, weather stations, and road characteristics. Recently, Yi et al. (2021) used data-driven machine learning and graph theory to infer hourly traffic volumes including the effect of spatial dependency among different locations.

The traditional method for link flow estimations is formed on traffic assignment techniques. These techniques aim to compute link flows based on the O-D matrix and equilibrium assumptions (Cascetta, 2009).

A.2 Methodology

The proposed prediction of hourly traffic volumes is based on the trip characteristics, network connectivity, land use, and temporal effects applicable to safety management at the level of a single road element and one hour. The research scope was limited to selected classes of urban road segments. A log-normal regression model (Equation A.1) was selected for two reasons: (1) to avoid negative estimates of traffic volumes, and (2) to address the heteroscedasticity typically found in count observations. Furthermore, the model with the dependent variable in its observed form is connected with the other variables via a multiplicative form that was found appropriate for many count data models. An observation in this model applies to a road segment between intersections.

$$Vh, k = \exp\left(\beta_0 + \beta_\tau^P \{\ln (\Sigma_{\{\tau\}} P_k)\} + \Sigma \beta_{\tau,i} X_i + \varepsilon_k\right) \qquad \text{\textit{Equation A.1}}$$

where: $V_{h,k}$ = hourly traffic volume along segment k, ε_k = error term, β's are the model estimates and P_k = weighted propensity of segment k. The proposed regression-based approach introduces two new variables for roadway segments: *travel propensity* and *travel time excess index*, which reflect trip motivation and the routing elements omitted in the existing regression models. These variables consider the link connectivity, route choice and land use characteristics of the origins and destinations. An *O-D* matrix with a static traffic condition is considered for understanding work-related trip patterns among traffic analysis zones (TAZ), and consequently, computation of the travel propensity and travel time excess index of the segments of interest.
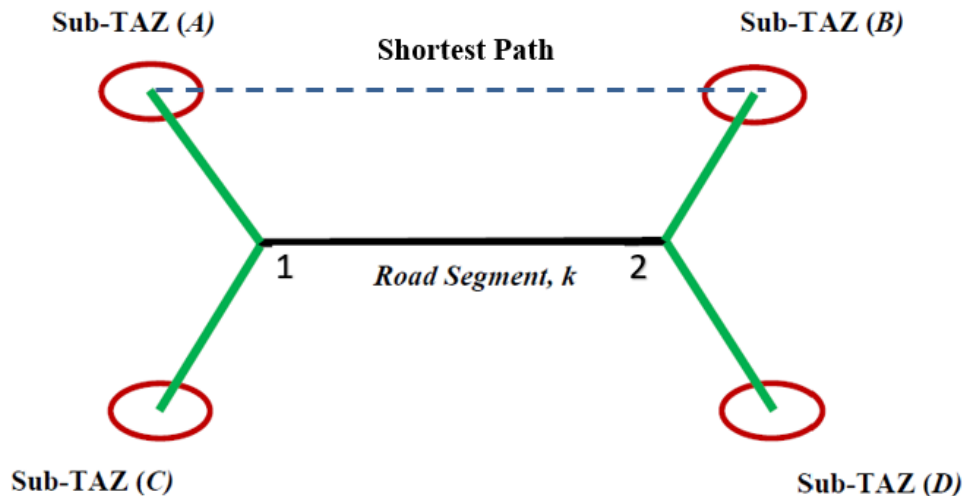

Figure A.1 Network routing via subject segment, k.

A-2

Typically, commuters prefer closely connected road segments that can be used to reach a wide range of destinations within a reasonable travel time range. Such characteristics of the segment should significantly affect the traffic volumes on that particular segment. In this study, the *travel propensity* of a road segment represents the potential traffic using a segment to travel between two groups of sub-TAZs distant from each other within a certain maximum travel time (e.g. 40 minutes). In Figure A.1, the road segment k connects sub-TAZs A and C to sub-TAZs B and D. The travel propensity within range τ (e.g., 40 minutes) from sub-TAZ i via segment k is expressed through the following equation:

$$p_{i,\tau,k} = \frac{r_i W_{i,\tau,k}}{W_{i,\tau}}$$ Equation A.2

where r_i = number of working residents in sub-TAZ i, $W_{i,\tau,k}$ = number of sub-TAZs reachable from sub-TAZ i via directional segment k within travel time range τ, and $W_{i,\tau}$ = number of sub-TAZs reachable using the *shortest paths* within travel time range τ from sub-TAZ i.

On the other hand, commuters prefer routes that offer short travel times. Therefore, a smaller number of trips will be made via segment k if the shortest travel time possible via this segment is longer than the travel along an alternative shortest path on the road network represented with the blue dashed line in Figure A.1. The travel time excess index represents the mitigating effect of a longer travel time via segment k than along alternative routes. In Figure A.1, the shortest path connecting sub-TAZ A and sub-TAZ B can be compared to the path connecting these two sub-TAZs through road segment k. The excess travel time is evaluated using this index. This variable serves as a proxy for route choice. This link-specific information is critical for the correct estimation of traffic volumes. Using the following Equation A.3, the variable can be computed:

$$\Delta = \frac{t_k - t_s}{t_s}$$ Equation A.3

where t_k = travel time on the *near-shortest path* (through the segment of interest) and t_s = travel time on the *shortest path* along the network.

Indeed, the application of the expected research outcome is practical in several ways. First, it does not involve any traffic count or trip information to compute the traffic predictions. Second, the input required for predicting traffic, such as land use and roadway characteristics information are easily available. Third, the computation of the two compound inputs explained in the remainder of the study (travel propensity and travel time excess index) are relatively simple, intuitive, and applicable to all cases. Fourth, traffic safety analysis on a network scale can be accomplished on local roads where no traffic volumes are currently available.

A.3 Data

A sample of 260 urban road segments with short-term traffic counts was prepared. The sample covers mainly lower classified roads (i.e., minor arterials, major collectors, minor collectors and

local roads) since past research (Tarko et al., 2021) tackled the prediction of hourly traffic volumes on freeways and other major roads on traffic. To exclude the COVID-19 pandemic effect, only traffic information from the year 2019 was analyzed. It also should be noted that this sample includes observations for all weekdays except Fridays for two reasons: (1) the proposed methodology is exclusively based on work-related trips, and (2) sufficient observations for weekends and Fridays were not available from INDOT's traffic count database system. Hourly traffic counts for these segments were used as the dependent variable in this model.

This study considers sub-TAZs as the origin and destination points to understand the work-related trip characteristics of commuters. The trip travel times were used to compute travel propensity and the travel time excess index for the sample road segments. First, to evaluate the trip travel times through the shortest paths along the network, a time-independent *O-D* Cost Matrix was created based on 6,188 sub-TAZs and the Indiana road network data set using ArcGIS Pro 2.8. For the purpose of this study, 40 minutes was considered the maximum commute time between sub-TAZ pairs acceptable to travelers when they decide their home and workplace locations. Longer travel times were assumed negligible and pairs of sub-TAZs with longer times of travel between them were not considered in this study. For instance, in A.1 27, the blue dashed line is regarded as the *shortest path* or *fastest path* along the network for sub-TAZ groups A and B. Furthermore, the *near-shortest* paths in this study refer to the shortest paths connecting pairs of sub-TAZ groups via the sample road segments along the road network. A separate but similar network routing procedure was carried out for examining these *near-shortest* paths where any redundant or longer than the shortest paths were eliminated. The specific boundary for the near-shortest paths was determined by a maximum allowable travel time excess index (i.e., 0.35). Therefore, only those near-shortest paths were considered that incurred travel times which were at most 35% of that incurred through the *shortest paths*.

Finally, the travel time information of the shortest paths and near-shortest paths was utilized in the calculation of weighted travel propensity and the mean travel time excess index of sample road segments. Summary statistics for each variable are shown in Table A.2. The weighted travel propensity was derived by applying a factor, *N,* to the total propensity value of a road segment where $N = (1-\Delta)/1,000$. Moreover, the mean travel time excess index for each segment was considered in the final estimation model. In this study, a travel time excess index higher than 0.35 indicated a "too slow" path and the corresponding travel propensity for that path was not included in the calculation of total travel propensity for the segment. It may be a concern that in the estimation sample only 3.8% of the urban road segments are minor collectors and only 4% are local roads, which might lead to reduced accuracy in the estimated parameters corresponding to these variables. This small number of observations is due to the low availability of data by INDOT.

Table A.1 Descriptive statistics of sample used in traffic volume prediction

Variable	Mean	Std. Dev.	Percent	Min	Max
Land Development and Routing					
Weighted travel propensity	264.4	99.76	–	4.27	524.57
Log of weighted travel propensity	3.86	1.22	–	1.45	6.26
Travel time excess index	0.16	0.04	–	0.06	0.25
Road Class					
Minor Arterial (4)	–	–	43.4	–	–
Major Collector (5)	–	–	48.8	–	–
Minor Collector (6)	–	–	3.80	–	–
Local Roads (7)	–	–	4.00	–	–
Season					
Fall	–	–	23.5	–	–
Winter	–	–	16.4	–	–
Spring	–	–	33.6	–	–
Summer	–	–	26.5	–	–
Day of Week					
Monday	–	–	18.9	–	–
Tuesday	–	–	45.3	–	–
Wednesday	–	–	27.4	–	–
Thursday	–	–	7.80	–	–

A.4 Results

The modeling results in Table A.2 show the parameter estimates and statistical significance of the 38 variables considered. All of the significant variables that improved the overall goodness of fit were retained in the final model. The average value of the dependent variable hourly traffic volume was 146.1 vehicles per hour and its standard deviation was 225.5 vehicles per hour, which indicated high variance. The weighted travel propensity was log-transformed before inclusion in the model. The R^2 value of 0.533 indicates that the log-normal regression model explains more than half of the variance in the sample. The adjusted R^2 was calculated as 0.530 with the following equation:

$$\bar{R}^2 = 1 - (1 - R^2)\frac{N - 1}{N - K} \qquad \text{\textit{Equation A.4}}$$

where N = the total number of observations and K = the number of parameters in the model. This measure adjusts for the number of predictors in the regression model. The constant term of 2.455 was statistically significant at the 99% confidence level. Under a correct sampling scheme, this term accounts for the joint effect of all the unobserved factors.

In terms of the specific effects of various factors, many categorical and continuous variables were considered. The model parameters are quite intuitive in terms of the general understanding of the corresponding traffic effects. For example, minor arterials serve as a reference case and as expected, major collectors, minor collectors and local roads tend to have lower hourly traffic volumes than the reference road as indicated by the negative parameters associated with these types of roads.

Travel propensity and the travel time excess index, which reflect the excess travel time along a segment above the minimum offered via the fastest path, also produced intuitive estimates and were found to be highly significant. The propensity of travel on a road segment was found to be positively correlated with the hourly traffic volumes on the segment at the statistical level of 99%. This finding is consistent with the expectation that a segment that offers a good connection between a large number of origin-destination pairs (travel potential) should be used by a larger number of travelers than in cases of lower travel potential. A high volume is a tangible manifestation of such a statistical connection. Again, the negative estimate of the higher travel time excess index refers to lower traffic flows through the links. Therefore, it properly reflects the expected discouraging effect of a longer trip on a driver's selection of a connection via a subject segment. The negative regression parameter associated with the travel time excess index confirmed this expectation. In addition, several temporal factors, such as time of day, day of week and seasonal variations are included in the model. Intuitively, the hour indicator shows a strong positive association with high traffic flows during the morning and evening peak hours. In the fall season, the traffic volume tended to be the highest followed by winter and then summer. The "back to school jump" could be one of the primary reasons why traffic is much higher during the fall. As for the days of the week, Mondays generally were the busiest. Moreover, the interaction variables include primarily the interaction among log of travel propensity and weekdays. These variables adjust the propensity values with regard to days of week variations. While the estimates can be challenging to interpret, they provide insights on how land use varies by the day of the week and their contributions to hourly traffic volume profile.

Table A.2 Parameter estimates of log-normal regression model of hourly traffic volume

Effect	Estimate	Std. Error	Wald χ^2 Statistic	Pr. > | χ^2 |
Intercept	2.455	0.313	61.72	< 0.001
Roadway Characteristics				
Minor arterial (4) – reference				
Major collector (5)	-0.557	0.024	534.51	<0.001
Minor collector (6)	-1.250	0.093	180.41	<0.001
Local roads (7)	-1.258	0.153	67.43	<0.001
Trip and Land-Use Characteristics				
Log of weighted travel propensity	0.289	0.016	326.30	<0.001
Travel time excess index	-2.103	0.327	41.45	<0.001
Temporal Factors				
Hour 0: 0:00–0:59 – reference				
Hour 1: 1:00–1:59	-0.500	0.574	0.76	0.384
Hour 2: 2:00–2:59	-0.623	0.629	0.98	0.323
Hour 3: 3:00–3:59	-0.554	0.597	0.86	0.354
Hour 4: 4:00–4:59	0.012	0.418	0.00	0.978
Hour 5: 5:00–5:59	0.956	0.319	8.99	0.003
Hour 6: 6:00–6:59	1.776	0.302	34.61	<0.001
Hour 7: 7:00–7:59	2.284	0.299	58.29	<0.001
Hour 8: 8:00–8:59	2.140	0.299	51.01	<0.001
Hour 9: 9:00–9:59	1.879	0.301	38.98	<0.001
Hour 10: 10:00–10:59	1.876	0.301	38.78	<0.001
Hour 11: 11:00–11:59	2.023	0.301	45.31	<0.001
Hour 12: 12:00–12:59	2.105	0.300	49.17	<0.001
Hour 13: 13:00–13:59	2.110	0.300	49.41	<0.001
Hour 14: 14:00–14:59	2.232	0.299	55.44	<0.001
Hour 15: 15:00–15:59	2.432	0.299	66.14	<0.001
Hour 16: 16:00–16:59	2.550	0.299	72.83	<0.001
Hour 17: 17:00–17:59	2.553	0.299	73.00	<0.001
Hour 18: 18:00–18:59	2.243	0.299	56.03	<0.001
Hour 19: 19:00–19:59	1.939	0.301	41.45	<0.001

Effect	Estimate	Std. Error	Wald χ^2 Statistic	Pr. > $\|\chi^2\|$
Hour 20: 20:00–20:59	1.669	0.304	30.25	<0.001
Hour 21: 21:00–21:59	1.344	0.309	18.94	<0.001
Hour 22: 22:00–22:59	0.922	0.323	8.16	0.004
Hour 23: 23:00–23:59	0.543	0.349	2.42	0.120
Season: Spring–reference				
Season: Summer	0.281	0.035	66.30	<0.001
Season: Fall	0.543	0.031	302.08	<0.001
Season: Winter	0.474	0.033	209.57	<0.001
Day: Monday–reference				
Tuesday	-0.146	0.087	2.86	0.091
Wednesday	-0.332	0.098	11.50	0.007
Thursday	-0.509	0.147	12.00	0.001
Interaction Effects				
Log propensity*Day: Mon reference				
Log propensity*Day: Tue	0.038	0.019	3.98	0.046
Log propensity*Day: Wed	0.089	0.021	18.14	<0.001
Log propensity*Day: Thu	0.141	0.029	22.55	<0.001
Number of Observations	6,294			
Adjusted R^2	0.53			
Residual Standard Error (veh/hour)	154.6			

To examine the estimation capability of the model, the observed and predicted values and the residuals were closely investigated. The calculated RSE was 154.6 vehicles per hour (Table A.2). Conversely, the mean of the observed hourly traffic counts was 146.1 vehicles per hour. One of the reasons for this difference could be that the travel propensity function does not by definition capture the effect of non-work-related trips and aberrations occur for not considering these trips. Also, since the trips were assumed to be from sub-TAZs to sub-TAZs, the traffic generated from the freeways and other major roads were not covered. To understand the prediction errors, the scatter plots and the residual plots were also examined.

In Figure A.2, different percentiles of the residuals for the sample as a function of hour are illustrated. It is clear that the fluctuations in the residuals for daytime traffic were comparatively higher than that of nighttime traffic.
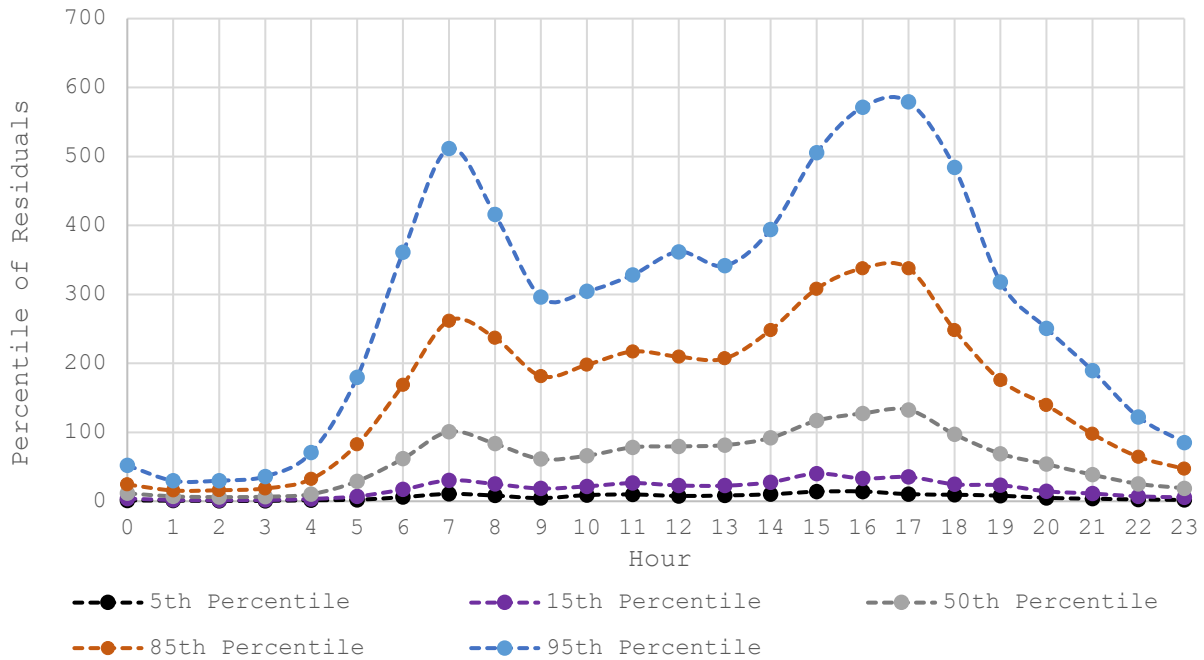
Figure A.2 Percentiles of residuals as a function of hour.

Furthermore, the residuals resulting from the regression model are shown in Figure A.3. The plot represents a "cone" shape, which indicates that the residuals were much more spread out as the fitted values became larger. This pattern is indicative of heteroskedasticity in the data set, which refers to non-constant variance. When the scatter of the errors varies depending on the value of one or more of the independent variables, the error terms are heteroskedastic. This situation can be remedied with a log-transformation of the dependent variable, which was demonstrated previously.
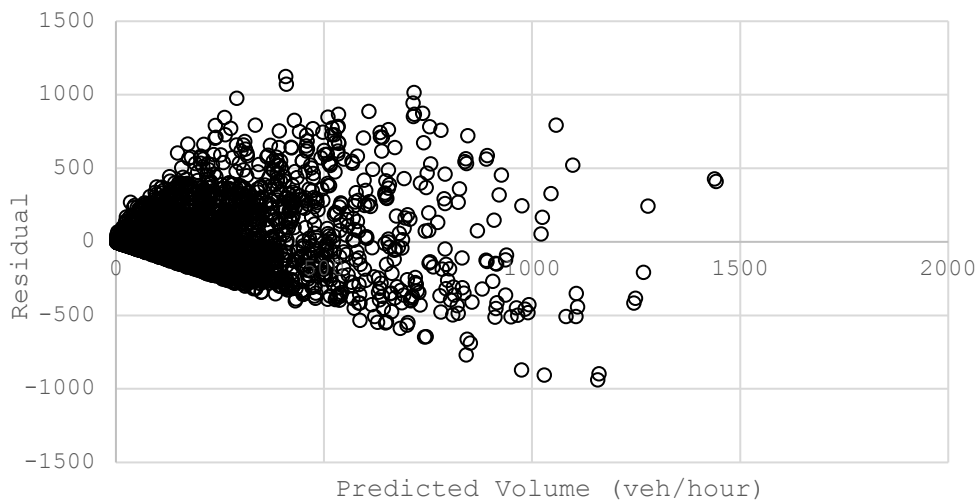
Figure A.3 Residuals plot.

A.5 Discussion

Two of the model variables were formulated to represent the traffic generation and exchange (travel propensity) and the path selection (travel time excess index). The proposed model for hourly traffic volume prediction follows the practice and meets the needs of INDOT and other agencies. The developed statistical model includes spatial and temporal variables, road characteristics, and the mentioned two variables of traffic generation and routing. Thus, the reported research addresses the conditions of a reasonable and relatively simple model for predicting hourly traffic volumes. In summary, promising groundwork for future efforts in traffic predictions were laid out in this study via regression analysis. In terms of the individual effects of the explanatory variables, higher travel propensity, as anticipated, was associated with higher traffic volumes on roadways. Also, intuitively, the travel time excess index was linked to lower traffic counts on segments. The model estimates also indicate that more vehicles are present in any time of day on higher road classes. As for seasonal fluctuations, fall has the highest traffic volume while fewer vehicles are present on roads in spring. The model also presents reasonable estimates regarding individual days of the week and the time of day.

However, the research results were affected by the following data limitations. Inadequate high-resolution data for lower class roads may have caused the sample to not fully reflect the population of these roads. The effect of traffic congestion was ignored for the sake of the method's convenience but may have contributed to the error magnitude. However, this effect could be much smaller when the method is applied to rural roads where congestion is less likely than on urban roads. The assumption was made that all the traffic observed was the result of travel activities between sub-TAZs and within 40 minutes of travel time. In reality, some traffic observed near interchanges and other intersections of interstate and US roads may have had some origin and destination traffic of long distance and was not related to work. Since the methodology here focused on work-related trips between sub-TAZs, it does not address the contribution of non-work-related trips.

There were several limitations of the sample data set chosen for the estimation models. The scope of this study did not include interstates, freeways, or principal arterials. Aside from that, the percentage of minor collectors and local roads was less than 10 percent of the sample size since the INDOT data set lacked coverage counts for these roads. Moreover, observations for weekends and Fridays were omitted due to insufficient representative data. To further develop the models, traffic counts from these days could be useful. The scope of the research did not include rural road segments due to the time constraints of the study. As mentioned earlier in this section, another limitation was time-independent network modeling, which does not consider traffic congestion. However, this comment also increases the prospect that an applied methodology may perform better when applied for rural roads where roads are rarely congested. Other potentially useful land-use variables such as vehicle ownership, business sales, floor size of establishments, etc. could be considered in future work. Disaggregate weather data also may better facilitate the estimation of hourly traffic volumes.

Addressing these drawbacks would be the first steps to further refine the proposed models. Although in the prediction of hourly traffic volumes, a log-normal regression was implemented,

alternatively, a negative binomial (NB) model should be considered. An NB count data model with an over-dispersion parameter may better explain the variance. Nevertheless, the research methodology provides greater understanding of commuter trip characteristics, route preferences, and the impacts of time-dependent conditions on traffic.

APPENDIX B. RISK-BASED SMS APPLICATION RMT

B.1 Risk-Based SMS Overview

RMT tool (named here Risk Management Tool, RMT) is a prototype computer application that supports the following operations:
- Selecting for evaluation the high-crash segments identified with SNIP.
- Visualizing the hourly crash risk on the selected freeway segments in several years.
- Applying conditions (called triggers) that activate operational countermeasures.
- Estimating the number of crashes during the subperiods with active countermeasures to evaluate the treatment's effectiveness (computational elements of RoadHAT 4D to be implemented in RMT).

B.1.1 Segment Selection

The RMT facilitates selecting road segments by clicking on the map or by selecting the corresponding record on the data table. Both the segments on the map and the record on the table are highlighted when selected. The program automatically loads the shape file of the road segments and the corresponding time dependent data such weather conditions, vehicles' speed, and traffic volume.

B.1.2 Results Visualization

The *selection* process is immediately followed by building and displaying a risk profile that can be aggregated by hour, day, or month. Such visualization is beneficial in presenting the results to decision-makers and to identify time patterns not detectable otherwise.

B.1.3 Conditions (Triggers)

The temporal profile can be filtered by applying by the user certain traffic and environmental (for example, weather) conditions called triggers. These triggers may be used to identify periods when a particular countermeasure is to be applied. Once the conditions (triggers) are selected, the user should refresh the results to see the updated results that include the estimated number of crashes affected by the countermeasure.

B.1.4 Crashes Affected by Countermeasures

Once the road segments are selected and the triggers of operational countermeasure are set, the program identifies the periods when the selected countermeasures are active, and it estimates the total number of crashes affected by the countermeasure in these periods. The number of crashes is also estimated by severity levels as defined in RoadHAT 4D. Estimation of injury crashes (KABC) and PDO crashes are done with the statistical models. Then, the KABC crashes are split between KA and BC crash categories based on the proportion of these crashes in rural freeways. The resulting estimates may be then used to run the BC analysis with RoadHAT 4D computational elements embedded in RMT (not implemented).

B.2 Installation

RMT is compatible with MS Windows 64-bit operating systems. The RMT files for installation are compressed and the user must unzip them in the folder of his or her choice prior to installing the software. The unzipped files are shown in Figure B.1.
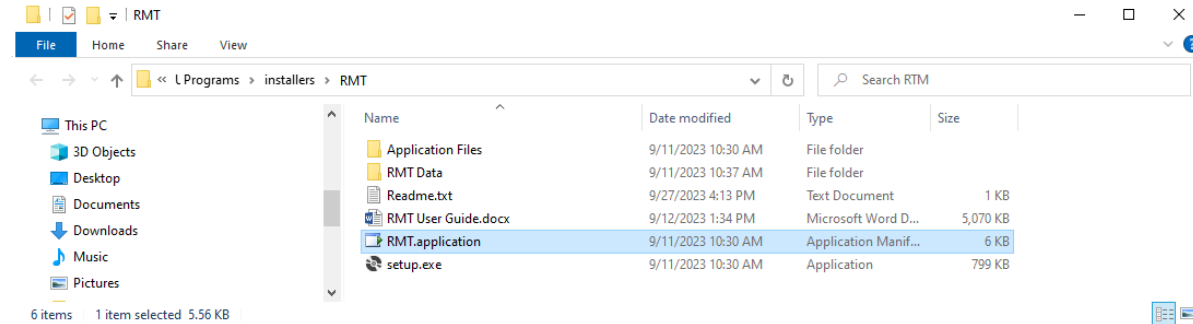

Figure B.1 RMT files.

The unzipped *Readme.txt* file provides installation instructions for the tool. These steps are also discussed below. The user should retain the zipped/compressed file to reinstall the program in the future if necessary.

First, the **RMT Data** folder containing the RMT database must be copied into the PC's **C:\Users\Public** folder (Figure B.2). The program uses two types of data, the first type is *time-independent* data which is related to the infrastructure, and the second one is *time-dependent* data such as weather, speed, and traffic.

B.2.1 Time-Independent Data

A shapefile from ArcGIS was used to store the non-time-dependent data. A shapefile contains both the geometry necessary to locate the segments on the map and the data corresponding to their characteristics. Roadway segments and its characteristics were obtained from INDOT's Road Network Data (RND) and supplemented with Google Earth's historical imagery.

B.2.2 Time-Dependent Data

This data is contained in a csv file (comma delimited values) for each year. Each segment corresponds to 8760 or 8784 records depending on whether or not it is a leap year. Each record represents the values corresponding to 1 hour and contains information on the segment id, date and time, vehicles' speed, weather, and traffic volume. Operating travel speeds at the segment level are assembled from the National Performance Management Research Data Set (NPMRDS). Hourly traffic volumes and vehicle classification are obtained from INDOT's Traffic Count Database System (TCDS). Weather conditions are accessed via the Indiana State Climate Office (INClimate) at Purdue University.
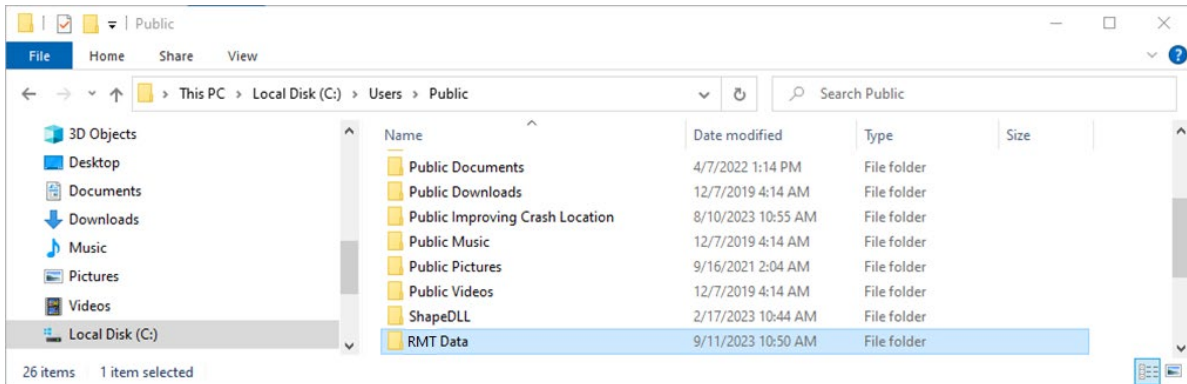
Figure B.2 Proper location of RMT data folder.

The next step is to install the RMT. The user should return to the folder where the content of the zip file was extracted and click on the *setup.exe* file (Figure B.3). The RMT interface should appear (Figure B.4). The splash screen will stay on screen until all data files are read, which might take a few seconds. Once the files are read, the splash screen will disappear.
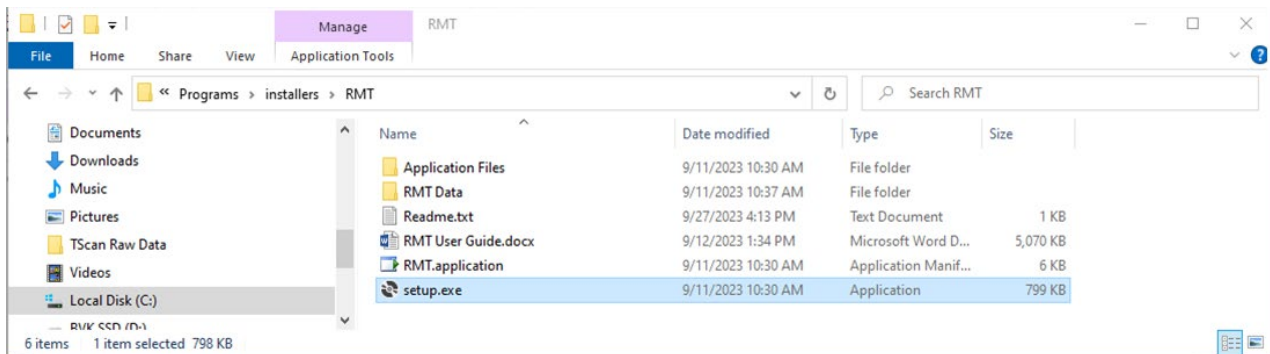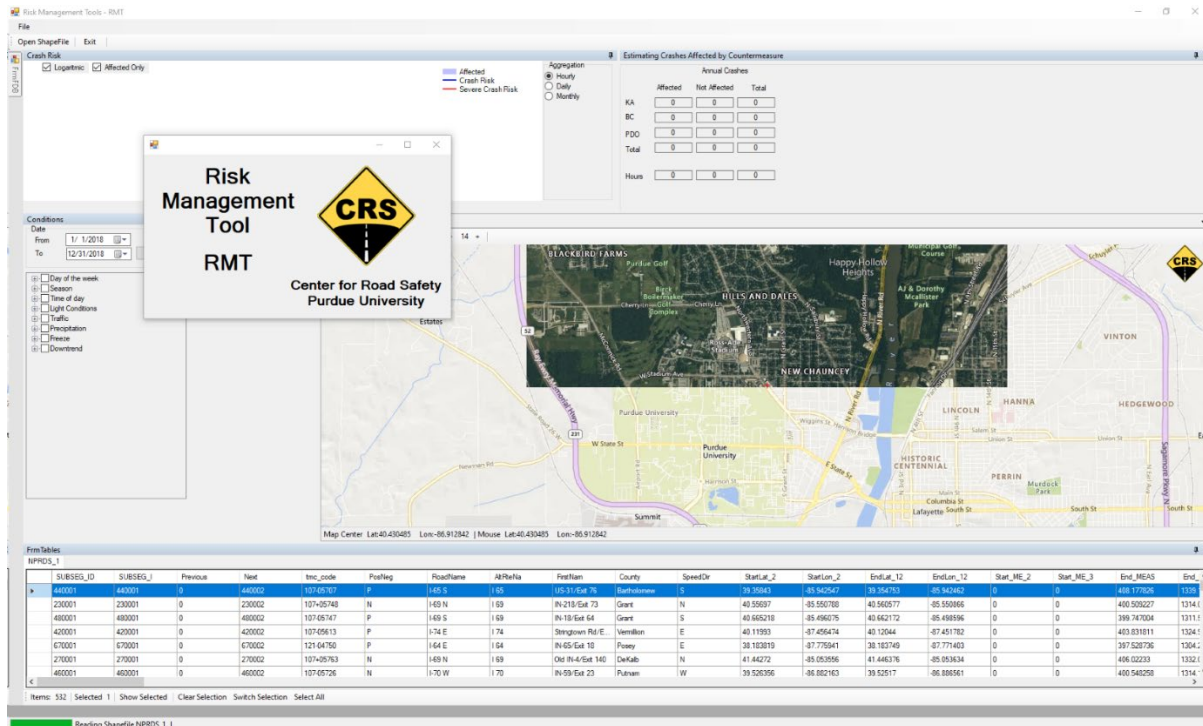

Figure B.3 Running the installation process.

B-3

Figure B.4 RMT initial window.

B.3 Launching RMT

The RMT can be launched using any of the following methods after installation:

Method 1:
Double-clicking the shortcut on the desktop

Method 2:
1. Pressing the "Start" button.
2. Afterwards, RMT should appear in the list of installed programs.
3. Single click on the shortcut.

Method 3:
1. Pressing the "Start" button.
2. Starting to type "RMT", and the program shortcut should appear in the search results.
3. Single clicking on the shortcut.

When executed, the RMT interface window appears within several seconds (Figure B.4). Any temporary tables remaining open from the previous run are closed. The splash screen will stay on screen until all data files are read, which might take a few seconds.

As noted in the Installation section of this manual, the *RMT_Data* folder containing the RMT database must be located in the PC's ***C:\Users\Public\RMT Data*** folder prior to the installation of the software interface. Otherwise, the user will receive an error message at the attempt at running the software.

B.4 Prototype Risk-Based SMS Application

B.4.1 Interface
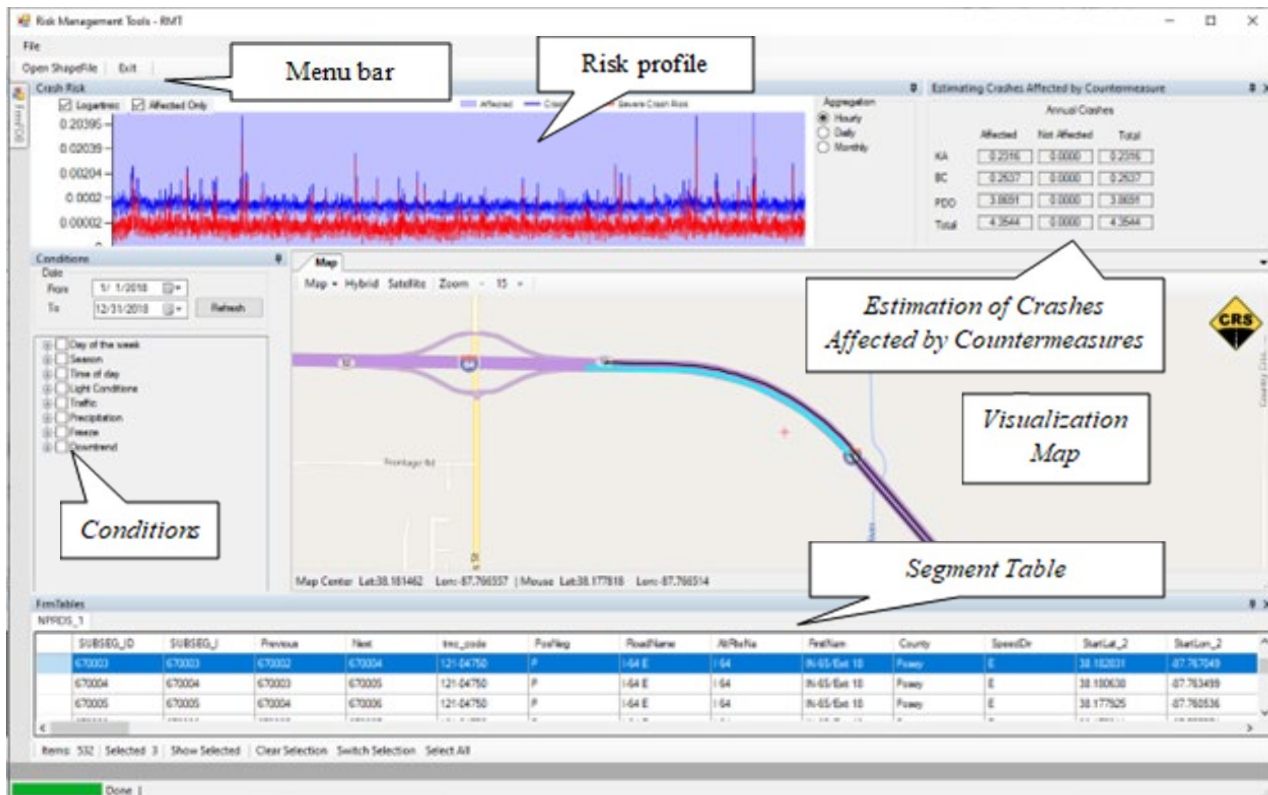
Figure B.5 RMT main interface.

Risk-based SMS (Risk Management Tool, RMT) provides an adaptive interface that allows adjusting its layout and design to fit the user's preferences. It allows the user to move, dock, resize, pin, and unpin windows. It includes a menu bar, the visualization map, the segment table, the risk profile, the conditions window, and the estimation of crashes affected by a countermeasure window (Figure B.5). Each of these components is described below.

To change the layout, users click and hold a window's title bar, drag it, and the interface indicates where it can be docked (Figure B.6).
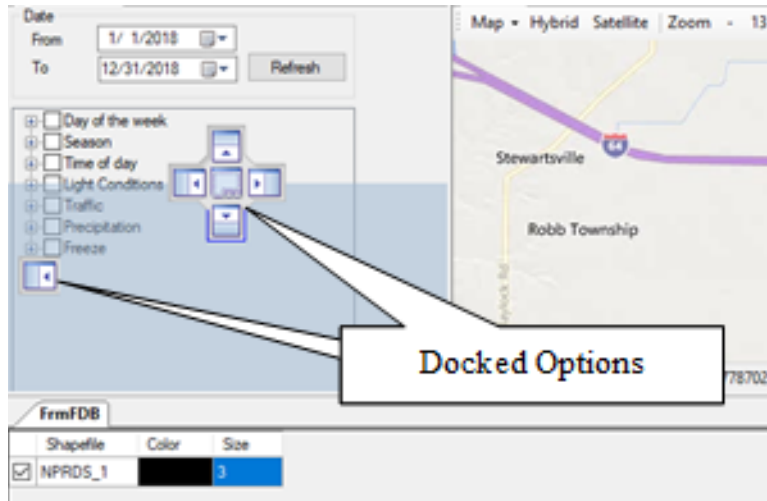
Figure B.6 Adaptive interface.

When hovering the mouse over a tab in a hidden window, the window is displayed to allow its use. The window can be unpinned to make it permanently visible or pinned for hiding it automatically (Figure B.7).
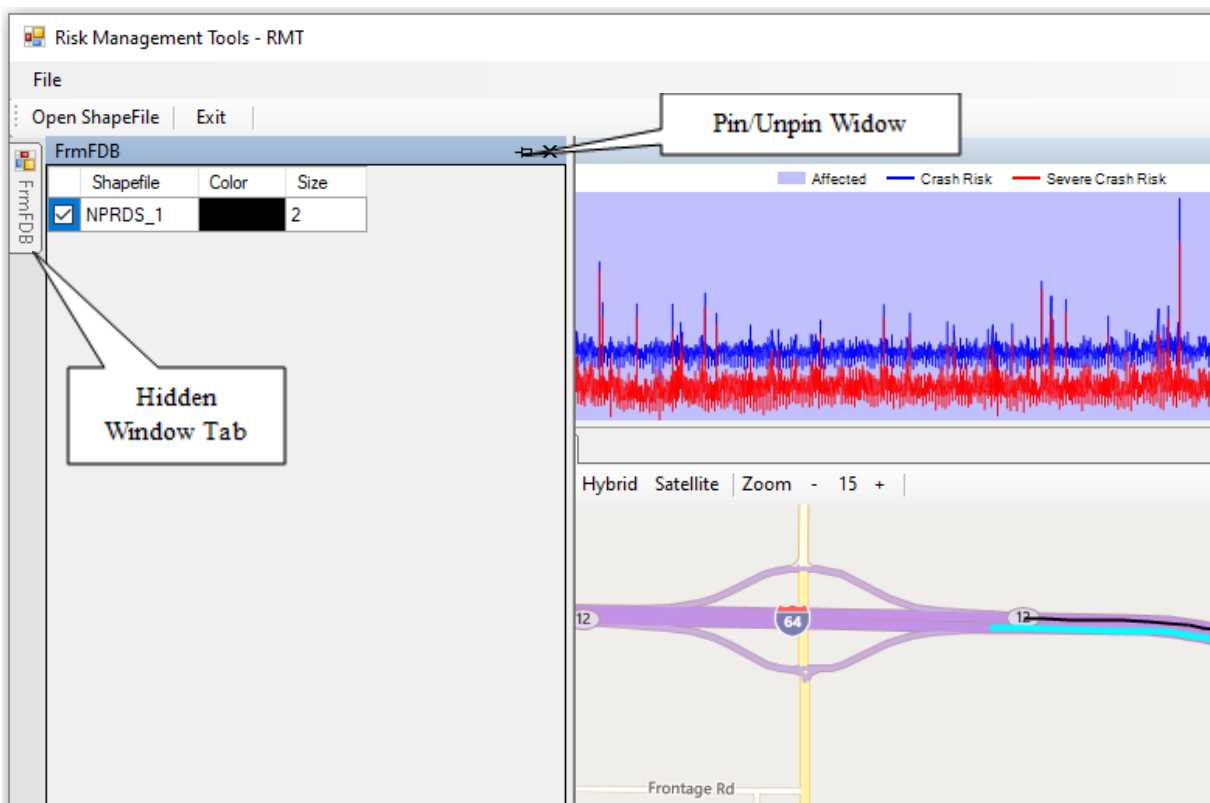

Figure B.7 Pin and unpin windows.

B.4.2 Map Window

Users can select the preferred map mode: Satellite, Hybrid, or Map. On the map option, users can select the preferred online map source between Google maps or Bing Map. In the same menu, users can increase or decrease the zoom level by clicking on the + or – buttons or by moving the mouse wheel (Figure B.8)

Users can move the map by dragging it with the mouse left button. The map window displays the geo coordinates of both the center of the map and of the mouse current position.
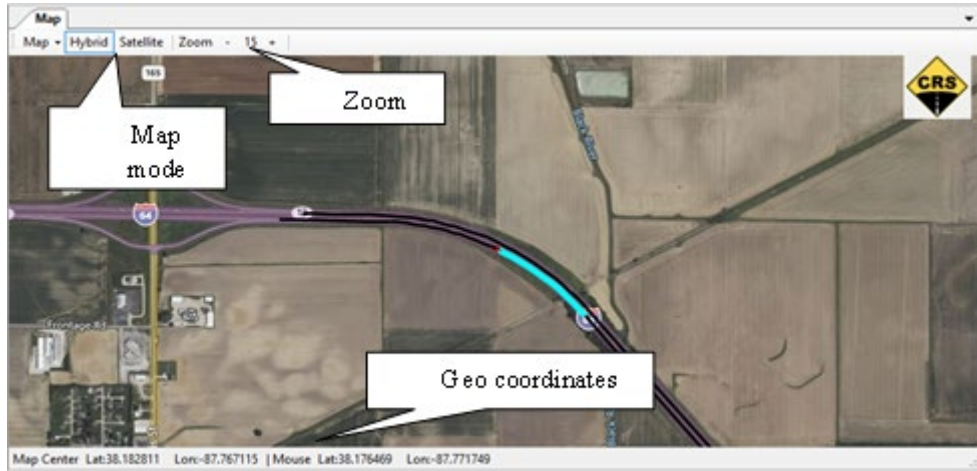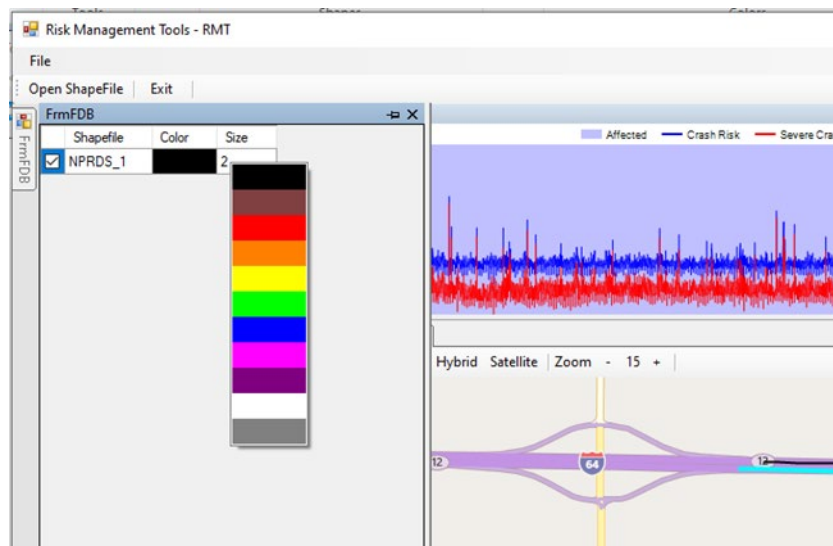

Figure B.8 Map window.

Figure B.9 Shapefile visualization control.

The RMT program automatically loads the segments shape file and displays it on the map. The Segments color and width can be changed using the side panel. Users type a new value on the size column to change the segment's line with on the map. Users right-click on the color to display a color selection option for the segments. (Figure B.9)

Clicking with mouse on segments in the map selects or deselects them. A segment can also be selected in the data table. The selected segments on the map and the corresponding records in the data table are highlighted (Figure B.5). The visualization graphs are updated with every new selection.

B.4.3 Risk Profile Visualization

A change in the segment selection automatically displays the risk profile for the selected segments. The risk profile chart shows three elements: total crash risk profile, severe crash risk profile, and periods selected based on triggers (explained in Section 4.4). User can select the interval of aggregation: hour, day, or month (Figure B.10).
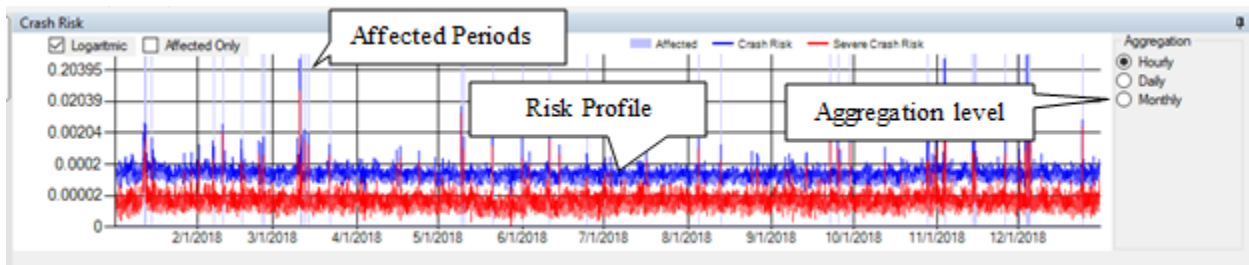

Figure B.10 Risk profile.

Use the logarithmic check box to switch between linear and logarithmic scales. In addition, the axes scale on the risk profile chart can be modified by hovering with mouse over the X or Y axis and use the mouse wheel to adjust the scale.

When the conditions are selected (Section 4.4), use the *Affected Only* check box to display the profile only during the periods with the selected conditions (Figure B.11).
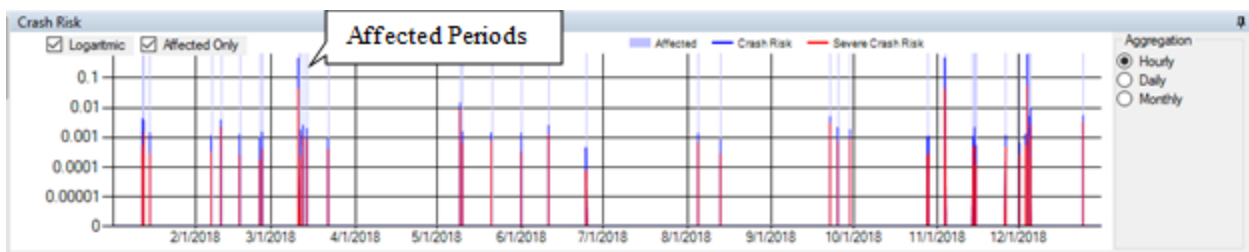

Figure B.11 Risk profile for the affected periods only

B.4.4 Conditions (Triggers)

The risk under certain traffic and environmental conditions, including their combination, can be analyzed in the Conditions Window.

Users select the period of time to analyze by entering or selecting the starting and end dates (Figure B.12). Other triggers available to select include day of week, season, time of day, light conditions, traffic conditions, precipitation, freeze, and traffic speed downtrend. Clicking check boxes select or deselect the conditions. Selected condition is used to filter periods for displaying the risk level. For convenience, all conditions unselected are equivalent to all being selected. After making all desired selections to be used jointly, users must click the *Refresh* button to update the results.
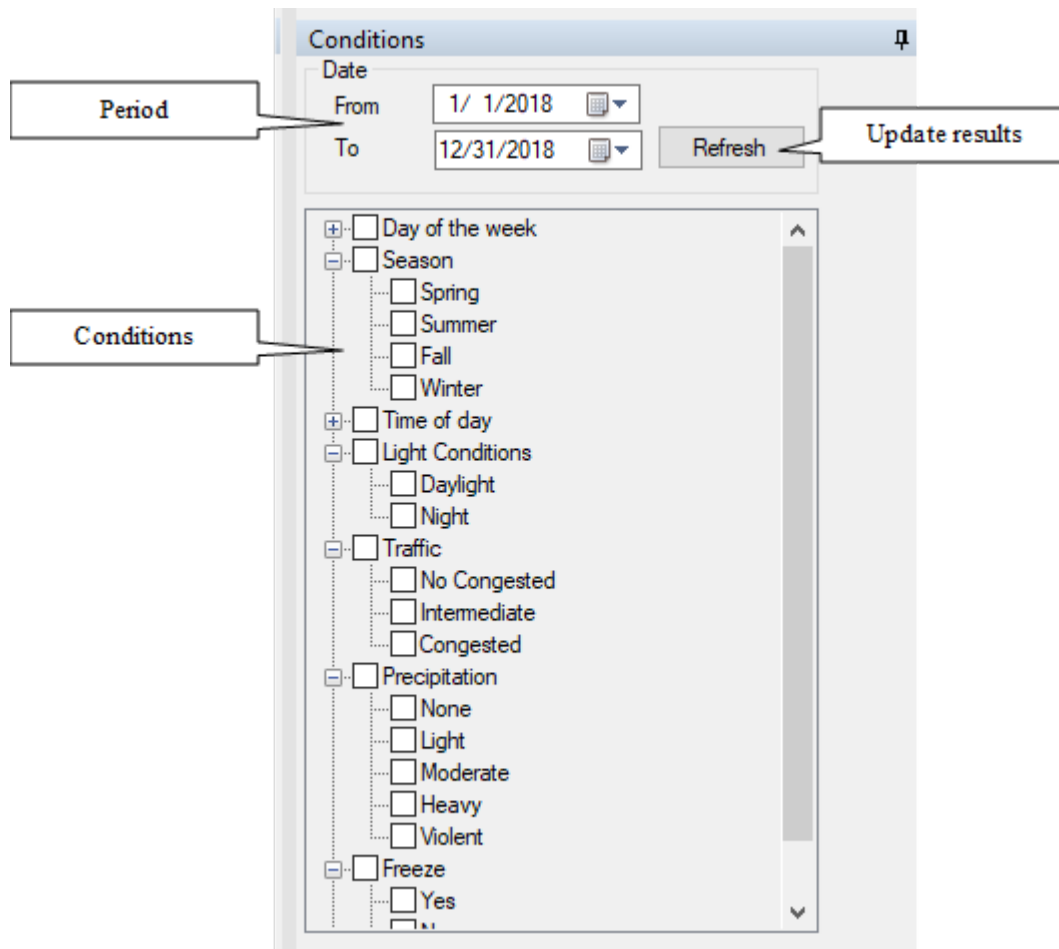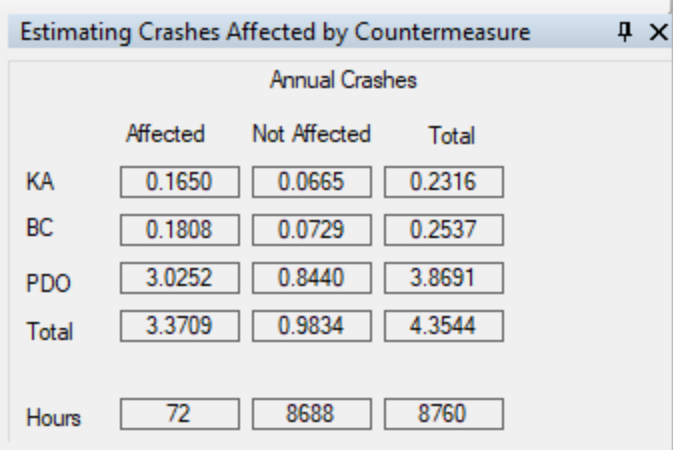

Figure B.12 Filtering window.

B.4.5 Estimating Crashes Affected by Countermeasures

When the segment selection changes or when a user refreshes the results after changing conditions, the numbers of crashes by severity and the total number affected by the countermeasure are estimated and displayed in a table. The table also includes the number of hours when the operational countermeasures are triggered (Figure B.13).

Estimating Crashes Affected by Countermeasure			
Annual Crashes			
	Affected	Not Affected	Total
KA	0.1650	0.0665	0.2316
BC	0.1808	0.0729	0.2537
PDO	3.0252	0.8440	3.8691
Total	3.3709	0.9834	4.3544
Hours	72	8688	8760

Figure B.13 Estimating crashes affected by countermeasures.

 The resulting annual number of crashes can be used in the benefit-cost analysis using the computational elements of the RoadHAT 4D (not implemented yet in RMT).

About the Joint Transportation Research Program (JTRP)

On March 11, 1937, the Indiana Legislature passed an act which authorized the Indiana State Highway Commission to cooperate with and assist Purdue University in developing the best methods of improving and maintaining the highways of the state and the respective counties thereof. That collaborative effort was called the Joint Highway Research Project (JHRP). In 1997 the collaborative venture was renamed as the Joint Transportation Research Program (JTRP) to reflect the state and national efforts to integrate the management and operation of various transportation modes.

The first studies of JHRP were concerned with Test Road No. 1—evaluation of the weathering characteristics of stabilized materials. After World War II, the JHRP program grew substantially and was regularly producing technical reports. Over 1,600 technical reports are now available, published as part of the JHRP and subsequently JTRP collaborative venture between Purdue University and what is now the Indiana Department of Transportation.

Free online access to all reports is provided through a unique collaboration between JTRP and Purdue Libraries. These are available at http://docs.lib.purdue.edu/jtrp.

Further information about JTRP and its current research program is available at http://www.purdue.edu/jtrp.

About This Report

An open access version of this publication is available online. See the URL in the citation below.

Pineda-Mendez, R., Guo, Q., Ahmad, N., Romero, M. A., & Tarko, A. P. (2024). *Incorporating time-dependent data for proactive safety management* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2024/01). West Lafayette, IN: Purdue University. https://doi.org/10.5703/1288284317700