# DREDGE (Disaggregate Realistic Artificial Data Generator)—Design, Development, and Application for Crash Safety Analysis, Volume I

PUBLICATION NO. FHWA-HRT-23-121

JANUARY 2024

U.S. Department of Transportation
**Federal Highway Administration**

Research, Development, and Technology
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

## FOREWORD

Data-driven safety analysis models help State and local agencies quantify safety data, identify high-risk roadway features, and predict the effects of proposed safety measures. However, even a model that performs well overall may not accurately represent interactions between variables for a specific location or crash because underlying relationships in the real world are unknown. One proposed solution is to generate realistic artificial data (RAD) with predetermined safety relationships built into them. Since these datasets are known, the RAD can serve as a test bed, indicating how well a model reflects underlying cause-and-effect relationships.

This report details a study performed by researchers working under the Federal Highway Administration's Exploratory Advanced Research Program. In the study, the researchers created a RAD generation framework on macroscopic and microscopic levels for a diverse selection of roadway facility types and developed a web-based tool. The tool's framework provides users with the ability to generate RAD for multiple years at macroscopic segment and microscopic trip levels. This report should be of interest to academics and researchers developing crash modification factors or functions and statistical models to determine how the models best represent real-world relationships. This volume is the first in a series. Volume II in the series is FHWA-HRT-23-122.

Brian P. Cronin, P.E.
Director, Office of Safety and Operations
Research and Development

# TECHNICAL REPORT DOCUMENTATION PAGE

| 1. Report No.<br>FHWA-HRT-23-121 | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| 4. Title and Subtitle<br>*DREDGE (Disaggregate Realistic Artificial Data Generator)—Design, Development, and Application for Crash Safety Analysis, Volume I* | | 5. Report Date<br>January 2024 |
| | | 6. Performing Organization Code: |
| 7. Author(s)<br>John Ivan (ORCID: 0000-0002-8517-4354), Shanshan Zhao (ORCID: 0000-0001-5476-6894), Kai Wang (ORCID:0000-0003-1452-4000), Oluwaseun Olufowobi (ORCID: 0009-0001-1339-9095), Naveen Eluru (ORCID: 0000-0003-1221-4113), Tanmoy Bhowmik (ORCID:0000-0002-0258-1692), Lauren Hoover (ORCID: 0009-0004-2431-0860), Md I. Jahan (ORCID: 0000-0002-4056-7816), Sudipta Tirtha (ORCID: 0000-0002-6228-0904), and Mohamad Abdel-Aty (ORCID: 0000-0002-4838-1573) | | 8. Performing Organization Report No. |
| 9. Performing Organization Name and Address<br>University of Connecticut<br>Connecticut Transportation Institute<br>270 Middle Turnpike, Unit 5202<br>Storrs, CT 06269-5202 | | 10. Work Unit No. |
| | | 11. Contract or Grant No.<br>693JJ31950017 |
| 12. Sponsoring Agency Name and Address<br>Office of Research and Development<br>Federal Highway Administration<br>6300 Georgetown Pike<br>McLean, VA 22101-2296 | | 13. Type of Report and Period Covered<br>Final report: August 2019–January 2024 |
| | | 14. Sponsoring Agency Code<br>HRSO-2 |
| 15. Supplementary Notes<br>The contracting officer's representative was Yusuf Mohamedshah (HRSO-2; ORCID: 0000-0003-0105-5559). | | |


16. Abstract

Safety analysis primarily focuses on identifying and quantifying the influences of contributing factors to traffic collisions and the consequences of these factors. A practice of relying on observed data only allows relative comparisons between analysis methods and may not lead to determining how well the methods mimic the true underlying crash-generation process, which is often unobserved or known only partially, with varying degrees of certainty. This report describes a study, performed by researchers working under the Federal Highway Administration, to help address this data limitation. Researchers used two approaches to generate realistic artificial data (RAD): In the macroscopic approach, researchers generated RAD at site level by segment or intersection after generating traffic and roadway characteristics by segment or intersection for various facility types. Crashes were then generated using known model structures, based on these characteristics. The microscopic approach built a high-resolution, disaggregate data-generation process that mimics crash occurrences on road facilities at the trip level and accommodates the influence of a full range of crash-contributing factors to generate crashes. This crash generation employs data describing the generated trips and involves identifying the vehicles involved in the crash, crash location, severity of occupant injuries, and crash type. These generated crashes can be aggregated at any spatial or temporal resolution to estimate and evaluate safety models. The researchers thus developed a RAD generator as a standalone, customizable software application tool that can prepare multiple realizations of RAD. The tool was evaluated using two case studies involving segment crashes and intersection crashes. This volume is the first in a series. Volume II in the series is FHWA-HRT-23-122.


| 17. Key Words | 18. Distribution Statement |
|---|---|
| Realistic artificial data, crash analysis, RAD generator | No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161.<br>https://www.ntis.gov |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 135 | N/A |

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

## SI* (MODERN METRIC) CONVERSION FACTORS

### APPROXIMATE CONVERSIONS TO SI UNITS

| Symbol | When You Know | Multiply By | To Find | Symbol |
|---|---|---|---|---|
| | | **LENGTH** | | |
| in | inches | 25.4 | millimeters | mm |
| ft | feet | 0.305 | meters | m |
| yd | yards | 0.914 | meters | m |
| mi | miles | 1.61 | kilometers | km |
| | | **AREA** | | |
| $in^2$ | square inches | 645.2 | square millimeters | $mm^2$ |
| $ft^2$ | square feet | 0.093 | square meters | $m^2$ |
| $yd^2$ | square yard | 0.836 | square meters | $m^2$ |
| ac | acres | 0.405 | hectares | ha |
| $mi^2$ | square miles | 2.59 | square kilometers | $km^2$ |
| | | **VOLUME** | | |
| fl oz | fluid ounces | 29.57 | milliliters | mL |
| gal | gallons | 3.785 | liters | L |
| $ft^3$ | cubic feet | 0.028 | cubic meters | $m^3$ |
| $yd^3$ | cubic yards | 0.765 | cubic meters | $m^3$ |
| | | **NOTE**: volumes greater than 1,000 L shall be shown in $m^3$ | | |
| | | **MASS** | | |
| oz | ounces | 28.35 | grams | g |
| lb | pounds | 0.454 | kilograms | kg |
| T | short tons (2,000 lb) | 0.907 | megagrams (or "metric ton") | Mg (or "t") |
| | | **TEMPERATURE (exact degrees)** | | |
| °F | Fahrenheit | 5 (F-32)/9<br>or (F-32)/1.8 | Celsius | °C |
| | | **ILLUMINATION** | | |
| fc | foot-candles | 10.76 | lux | lx |
| fl | foot-Lamberts | 3.426 | $candela/m^2$ | $cd/m^2$ |
| | | **FORCE and PRESSURE or STRESS** | | |
| lbf | poundforce | 4.45 | newtons | N |
| $lbf/in^2$ | poundforce per square inch | 6.89 | kilopascals | kPa |

### APPROXIMATE CONVERSIONS FROM SI UNITS

| Symbol | When You Know | Multiply By | To Find | Symbol |
|---|---|---|---|---|
| | | **LENGTH** | | |
| mm | millimeters | 0.039 | inches | in |
| m | meters | 3.28 | feet | ft |
| m | meters | 1.09 | yards | yd |
| km | kilometers | 0.621 | miles | mi |
| | | **AREA** | | |
| $mm^2$ | square millimeters | 0.0016 | square inches | $in^2$ |
| $m^2$ | square meters | 10.764 | square feet | $ft^2$ |
| $m^2$ | square meters | 1.195 | square yards | $yd^2$ |
| ha | hectares | 2.47 | acres | ac |
| $km^2$ | square kilometers | 0.386 | square miles | $mi^2$ |
| | | **VOLUME** | | |
| mL | milliliters | 0.034 | fluid ounces | fl oz |
| L | liters | 0.264 | gallons | gal |
| $m^3$ | cubic meters | 35.314 | cubic feet | $ft^3$ |
| $m^3$ | cubic meters | 1.307 | cubic yards | $yd^3$ |
| | | **MASS** | | |
| g | grams | 0.035 | ounces | oz |
| kg | kilograms | 2.202 | pounds | lb |
| Mg (or "t") | megagrams (or "metric ton") | 1.103 | short tons (2,000 lb) | T |
| | | **TEMPERATURE (exact degrees)** | | |
| °C | Celsius | 1.8C+32 | Fahrenheit | °F |
| | | **ILLUMINATION** | | |
| lx | lux | 0.0929 | foot-candles | fc |
| $cd/m^2$ | candela/m2 | 0.2919 | foot-Lamberts | fl |
| | | **FORCE and PRESSURE or STRESS** | | |
| N | newtons | 2.225 | poundforce | lbf |
| kPa | kilopascals | 0.145 | poundforce per square inch | $lbf/in^2$ |

*SI is the symbol for International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AADT | annual average daily traffic |
| AASHTO | American Association of State Highway and Transportation Officials |
| AF | adjustment factor |
| ANN | artificial neural network |
| CMF | crash modification factor |
| CPF | cumulative probability function |
| CRSS | Crash Report Sampling System |
| CS | conflicting speed |
| CV | connected vehicle |
| DOCTOR | Dutch Traffic Conflict Technique |
| DREDGE | disaggregate realistic artificial data generator |
| FCW | forward collision warning |
| FHWA | Federal Highway Administration |
| GIS | geographic information system |
| GOF | goodness of fit |
| GOL | generalized ordered logit |
| HSIS | Highway Safety Information System |
| MAD | mean absolute deviation |
| MDC | multiple discrete-continuous |
| MMNL | mixed multinomial logit |
| MNL | multinomial logit |
| MPR | market penetration rate |
| MSPE | mean-squared prediction error |
| NB | negative binomial |
| NDS | Naturalistic Driving Study |
| NHTSA | National Highway Traffic Safety Administration |
| OL | ordered logit |
| POLARIS | Polychotomous Choice Agent-Based Risk Model for Integrated Travel Demand and Network and Operations Simulation |
| RAD | realistic artificial data |
| SHRP2 | Strategic Highway Research Program 2 |
| SPF | safety performance function |
| STCT | Swedish Traffic Conflict Technique |
| TA | time to accident |
| UConn | University of Connecticut |
| USTCT | U.S. Traffic Conflict Technique |

# CHAPTER 1. INTRODUCTION AND MOTIVATION

Safety analysis primarily focuses on identifying and quantifying the influences of factors that contribute to traffic collisions and the consequences of these factors. A traditional analysis paradigm that relies on observed data only allows relative comparisons between analysis methods and lacks the ability to show how well the methods mimic the true underlying crash-generation process. This process is often unobserved or known only partially, with varying degrees of uncertainty. At the same time, existing data sources and the availability of data for model calibration and validation pose significant challenges to safety performance and crash modification analysis. Most safety performance analysis employs cross-sectional and time series datasets. Researchers make assumptions about the data, but whether these assumptions truly characterize the safety data that are generated in the real world often remains unknown.

One possible solution to address this issue is for researchers to artificially generate realistic artificial data (RAD) by making assumptions about the underlying crash generation process (Bonneson and Ivan 2013). These generated RAD can then provide a context for investigating various questions and verifying assorted assumptions related to safety performance and crash modification analyses. The idea is that these RAD can potentially be aggregated at any spatial or temporal resolution to mimic data from the real world. Furthermore, because researchers generating RAD have complete knowledge of the data-generation process, they can compare the performance of varied safety analysis methods.

In this study, the research team proposed two approaches to generate RAD: a macroscopic and a microscopic approach. In the macroscopic approach, researchers generated RAD at the site level by segment or intersection. Roadway sites that contained traffic and roadway characteristics—such as annual average daily traffic (AADT), lane width, shoulder width, curvature, and speed limit—were generated first, by facility type. The team then generated crashes using known model structures (e.g., Poisson and negative binomial (NB) models), based on generated site-level characteristics. Meanwhile, the microscopic approach built a high-resolution disaggregate data-generation process that mimicked crash occurrences on road facilities at the trip level and accommodated the influence of a full range of crash-contributing factors.

The microscopic approach employed detailed information for trips, including start and end time, start and end location, characteristics (including solo or group), vehicle used, and precise route. The crash generation process involved employing gathered data to identify vehicles involved, location, injury severity (including fatal, incapacitating, non-incapacitating, and uninjured), and crash type (including head-on, rear-end, and vehicle-pedestrian). The generated crashes were further aggregated based on location (intersections versus segments) and time (day, week, month, and year).

Next, the research team conducted two types of case studies to supplement RAD generation and validate the proposed RAD frameworks. The first case study estimated different types of statistical models, using datasets generated by the RAD procedure developed in this study, to illustrate how RAD could be used to make comparisons between analysis methods. The second case study used driving simulation to evaluate the effects of an advanced forward collision warning (FCW) system on crash occurrence.

The three objectives of this study were as follows:

1. Develop frameworks for RAD generation that can be used to evaluate methods used for both safety performance and crash modification analysis: The proposed RAD framework operates at both the site and trip level and incorporates the influence of a full range of crash-contributing factors. RAD frameworks are general enough to generate crash data for all roadway facility types, including segments and intersections. The frameworks can also generate data for different combinations of inputs, including modeling methods, model formulation, input specification, and unobserved heterogeneity. Researchers specifically focused on generating RAD to address known knowledge gaps related to aggregation issues, statistical and structural methodologies for safety analysis, and crash data challenges.

2. Develop DREDGE as a stand-alone software application: The application is customizable and can be executed to prepare multiple realizations of RAD. The application embeds features so RAD can be distributed efficiently; this embedding ensures that the integrity of the RAD process for evaluating methods is maintained. The stand-alone RAD software can also be used to study the uncertainty and stochasticity inherent to RAD generation and to evaluate model robustness.[1]

3. Demonstrate proposed RAD tool feasibility and applicability using two types of case studies: The first case study estimates different types of statistical models using the RAD generated from this study to demonstrate how the RAD can be used to make comparisons between analysis methods. The second case study uses driving simulation to evaluate the effects of an advanced FCW system on crash occurrence.

[1]FHWA. *DREDGE* (standalone RAD software).

# CHAPTER 2. LITERATURE REVIEW

The research team conducted a comprehensive review of previous research efforts on RAD approaches across various domains: statistics, econometrics, computer science, ecology, medicine, and psychology. In all these disciplines, the primary goal was to assess the ability of analysis methods to draw inferences about the underlying assumptions and assertions that generated the data. Researchers followed criteria to select studies for this review based on a simple core principle of RAD generation. The data generated in the research effort had to be based on a framework that was built on research assumptions—as opposed to real observed data-based simulation efforts.

The criteria eliminated two major sets of transportation studies that generate simulated data. First, several travel demand modeling forecast systems, such as activity-based models and synthetic population generators, generate individual-level synthetic data—for example, Eluru, Bhat, and Hensher (2008), Kitamura et al.(2000), and Konduri et al.(2016). However, the generation is entirely based on models estimated using observed data. Second, artificial data are generated in microsimulation frameworks for traffic flow modeling. In these studies, the simulated data are generated based on well-calibrated traffic flow models—for example, Asano, Iryo, and Kuwahara (2010); Mamun et al. (2020); Ranade, Sadek, and Ivan (2007); and Yu and Abdel-Aty (2014). Hence, these studies are also not appropriate for this review.

During the review process, the research team identified 30 research studies (based on the RAD generation criteria) that employed artificial data generation in their analyses. These studies included transportation (including transportation safety and travel behavior), medical science, data science, education, ecological modeling, information analytics, and environmetrics. The research team then prepared a summary of the literature review with the objective of developing a comprehensive RAD framework. Thus, as opposed to providing a study-by-study summary of earlier research, this review provides insight on the important elements of RAD frameworks that can be observed from earlier research efforts.

A concise summary of earlier research efforts on RAD generation is presented in table 1. This table provides information on study objectives, dataset adopted and study region (if known), software and procedures followed for generating RAD, conceptual methods and framework employed, contributing field of the study (for example transportation safety), and exact nature of the dependent variable (categorical or continuous). For the ease of presentation, the studies presented in table 1 are categorized into two groups based on the discipline of the study: first, studies related to transportation and second, studies related to other disciplines, including statistics, economics, ecology, and computer science.

**Table 1. Summary of existing literature on RAD generation.**

| Study | Study Objective(s) | Dataset Adopted (Study Region if Known) | Software/Procedure for RAD Generation | Conceptual Methods | Field | No. of Alt. |
|---|---|---|---|---|---|---|
| Transportation Domain | | | | | | |
| Bhat (2003) | To propose the use of scrambled Halton sequence for simulation estimation. | Simulated dataset | Generated data using unordered discrete choice models using the matrix programming language GAUSS (Aptech Systems 2023). | Mixed probit and multinomial probit models | Travel behavior | 3 |
| Cummings, McKnight, and Weiss (2003) | To review three methods for estimating relative risks in matched-pair crash data. | Simulated and observed dataset | Generated crash data employing Stata® Statistical Software with an assumed probability of fatality as a function of speed and seatbelt use (StataCorp 2023). | Mantel–Haenszel stratified methods, double-pair comparison method, conditional Poisson regression, and Cox proportional hazards regression | Safety | 2 |
| Salim et al. (2007) | To simulate intersection environment and collision and traffic data learning. | Simulated dataset | Vehicles are generated with different speed, position, and trajectory. | Ubiquitous intersection awareness framework | Safety | — |
| Paez and Scott (2007) | To develop a discrete choice model that incorporates elements of social influence and more conventional factors. | Simulated dataset | A Monte Carlo simulation was designed to explore the properties of the econometric model proposed. | Logit probability formulation | Travel behavior | — |
| Bhat et al. (2010) | To propose a CML approach to estimate ordered-response, discreet choice models with flexible, copula-based spatial correlation structures. | Simulated and observed dataset (San Francisco Bay area) | Considered three independent variables. Drew values from univariate normal distribution. Assumed fixed coefficients. Generated error terms using correlation structure. Generated 25 different datasets with 500 observations. | Copula-based, spatial ordered-response model structure | Travel behavior | 4 |

| Study | Study Objective(s) | Dataset Adopted (Study Region if Known) | Software/Procedure for RAD Generation | Conceptual Methods | Field | No. of Alt. |
|---|---|---|---|---|---|---|
| Bhat and Sidharthan (2010) | To investigate the ability of the MACML estimator to recover parameters from finite samples. | Simulated dataset | Considered five independent variables. Drew values from univariate normal distribution. Assumed random coefficients. Generated error terms from univariate normal distribution with 0.5 variance. Generated 20 datasets with 5,000 observations. | Cross-sectional random coefficients model, panel interindividual random coefficients model, panel intraindividual and interindividual random coefficients | Travel behavior | 5 |
| Pinjari and Bhat (2010) | To investigate nonworker out-of-home discretionary activity time-use and activity timing decisions on weekdays. | Simulated and observed dataset (San Francisco Bay area) | Assumed independent variable values were uniformly distributed. Assumed coefficients were nested extreme values. Generated data for 2,500 hypothetical individuals, with an assumption that each individual chose the value to maximize the total random utility. | MDCNEV | Travel behavior | 3 |
| Ferdous et al. (2010) | To model interactions in nonwork activity decisions across household and non-household members at the level of activity generation. | Simulated and observed dataset | Drew values for independent variables from univariate normal distribution. Assumed a fixed coefficient and used it to compute the utility for each individual using a linear combination. Generated error terms using predefined correlation structure. Repeated the process at least 50 times. | Multivariate, ordered-response system framework | Travel behavior | 3, 4, 5 |
| Ye and Lord (2011) | To examine the effects of underreporting crash data. | Simulated and observed dataset (Texas) | Weighted exogenous sample maximum likelihood estimator method was used for appropriately weighting the crash outcomes to address the underreporting issue in crash data. | MNL, OP, and ML models | Safety | 5 |

| Study | Study Objective(s) | Dataset Adopted (Study Region if Known) | Software/Procedure for RAD Generation | Conceptual Methods | Field | No. of Alt. |
|---|---|---|---|---|---|---|
| Geedipally, Lord, and Dhavala (2012) | To apply an NB, generalized linear model with Lindley mixed effects to the analysis of traffic crash data. | Simulated and observed dataset (road segments in Indiana, Michigan) | NB | NB-Lindley generalized linear model | Safety | 0 to α |
| Lord and Kuo (2012) | To examine the effects of site selection criteria. | Simulated dataset | Used R software to generate sites with crash counts with a predefined overall mean for different dispersion parameters (R Foundation 2021). | Compared four types of before and after studies: Naïve method, control group method, EB method based on the method of moment, and EB method based on a control group. | Safety | — |
| Bhat et al. (2013) | To apply the MACML approach for multiple MDCP models. | Simulated dataset (Michigan) | Considers five independent variables. Draws values from univariate normal distribution. Generates error terms from positive covariance matrix. Undertakes data generation process 20 times with different realizations of coefficient and error terms. | MACML | Travel behavior | 5, 10 |
| Eluru (2013) | To investigate the performance of ordered and unordered injury severity response frameworks. | Simulated dataset | Considered three independent variables. Assumed parameters that provide the same aggregate shares. Generated 5 realizations of the data with 5,000 observations each for each proportional value. Generated a total of six aggregate sample shares. | MNL, OL, and GOL | Safety | 4 |
| Paleti and Bhat (2013) | To compare the MSL inference and CML approaches. | Simulated dataset | Drew independent variables from univariate normal distribution while coefficients were assumed and drawn from multivariate normal distribution. Considered both independent and correlated realizations. Generated data at least 50 times. | CML and MSL approach | Travel behavior | 5 |

| Study | Study Objective(s) | Dataset Adopted (Study Region if Known) | Software/Procedure for RAD Generation | Conceptual Methods | Field | No. of Alt. |
|---|---|---|---|---|---|---|
| Wu, Lord and Zou (2015) | To generate CMFs using a regression model to estimate the crash counts and compare with the actual crash distribution | Simulated dataset | Assumed CMF values for lane width, curve density, and pavement friction and used them to generate simulated crash counts. | Evaluated the conditions for adopting NB regression models for before after studies. | Safety | 0 to α |
| Council et al. (2017)[2] | To use ARD to assess the performance of cross-sectional analysis methods. | RAD (rural two-lane highways, Washington) | Implemented data generation by SAS® programs based on an assumed model structure for AADT and roadway geometry factors. | All available methods of crash count and severity analysis | Safety | 0 to α |
| Nontransportation Domain | | | | | | |
| Gamel and Vogel (1997) | To compare parametric and nonparametric survival methods. | Simulated clinical dataset | Estimated three parametric models using data from breast cancer trials. Assumed the relationships generated survival data in HT Basic (Transera 1991). | Parametric (log-normal) and nonparametric test (logrank test, Gray-Tsiatis and Laska-Meisner methods) | Medical science | 0 to α |
| Scott and Wilkins (1999) | To evaluate data mining procedures. | Artificial dataset | Proposed two alternative ways of generating artificial data for testing data mining approaches. | Self-similarity, classification, and lazy trees | Data science | 0 to α |
| Bifulco and Bretschneider (2001) | To compare methods for assessing school performance. | Artificial dataset | Generated datasets from log-linear relationships between three inputs and two outputs and associated parameters. Generated datasets without bias, with measurement error and endogeneity. | Data envelopment analysis and corrected ordinary least squares | Education | — |
| Austin et al. (2006) | To evaluate statistical methods for predicting plant species distributions. | Artificial dataset | Generated data based on two plant community theoretical models using the computer package COMPAS (Minchin 1987). | Generalized linear models and generalized additive models | Ecological modeling | 0 to α |

[2]Council, F., E. Hauer, B. Lan, D. Harwood, and R. Srinivasan. 2017. *Use of "Artificial Realistic Data" (ARD) To Assess the Performance of Cross-Sectional Analysis Methods in Capturing Causal Relationships Between Individual Roadway Attributes and Safety*. Unpublished Report. Washington, DC: Federal Highway Administration.

| Study | Study Objective(s) | Dataset Adopted (Study Region if Known) | Software/Procedure for RAD Generation | Conceptual Methods | Field | No. of Alt. |
|---|---|---|---|---|---|---|
| Bzdusek and Christensen (2006) | To compare a new variant of PMF with other receptor modeling methods. | Artificial dataset | Generated data set using literature source profiles and postulated source contributions. | Eigenvalue-based methods and PMF-based methods | Environmetrics | — |
| Whiting, Hack, and Varley (2008) | To describe a method and toolset for creating realistic, synthetic test data. | Realistic simulated dataset | Generate simulated datasets by embedding ground truths. Core software and utility software were written in JAVA® (Arnold, Goslin, and Holmes 2005). | Threat stream generator | Information analytics | — |
| Potharst, Ben-David, and Van Wezel (2009) | To generate monotone ordinal datasets. | Observed and artificial dataset | Used a machine-learning algorithm to generate the data. | Algorithms for generating structured and unstructured random monotone datasets | Data science | — |
| Zimmermann (2012) | To generate diverse datasets reflecting realistic data characteristics. | Artificial dataset | Implemented data generator in JAVA. | Episode mining | Data science | — |
| Devroye, Felber, and Kohler (2012) | To estimate a density using real and artificial data. | Observed and artificial dataset | Implemented data generator in R software. Generated artificial data from a regression analysis of observed data. | Classical model, finite information model, and full information model | Data science | 0 to α |
| Hazwani et al.(2016) | To develop an automatic artificial data generator for generating artificial datasets based on real data. | Artificial and real dataset | Used random permutation algorithm to generate different sets of artificial data that represent realistic data. | Four-phase framework for data generation | Information and communication technologies | — |
| Dahmen and Cook (2019) | To introduce a synthetic data generation method. | Simulated and real dataset | SynSys, a machine learning-based synthetic data-generation method. | SynSys, similarity measures, and semisupervised learning | Medical science | — |

Alt. = alternatives; CML = composite marginal likelihood; CMF = crash modification factor; no. = number; EB = empirical Bayes; MACML = maximum approximate CML; MDCNEV = multiple discrete-continuous nested extreme value; MDCP = multiple discrete-continuous probit; ML = mixed logit; MSL = maximum-simulated likelihood; OP = ordered probit; PMF = Positive matrix factorization.
*After a concept proposed by Hauer, as presented in Harwood et al. (2003).

Researchers made five important observations based on the information seen in table 1. First, earlier research explored RAD applications for wide-ranging topics, including statistical/econometric model performance and comparison, travel-demand forecasting, route-choice behavior, and data mining. Second, RAD applications have typically been developed using several software packages or platforms, including R, GAUSS, and COMPAS (R Foundation 2021; Aptech Systems 2023; Minchin 1987). Third, employing RAD datasets, researchers considered the performance of several model structures, including ordered logit (OL), multinomial logit (MNL), generalized ordered logit (GOL), mixed multinomial logit (MMNL) models, probit models (and their cross-sectional and panel variants), multiple discrete-continuous (MDC) frameworks with probit and extreme value formulations, and artificial neural networks. Fourth—notably—studies within the transportation domain traditionally adopt RAD approaches for econometric models. However, non-transportation domain research typically is more focused on machine-learning and data-mining approaches. Fifth, the number of alternatives in the RAD variables studied is related to the problem context. The number of alternatives for a RAD variable could range from a small number to a very large number. If the RAD variable is a binary outcome variable, such as crash/no crash, then the number of alternatives will be two. However, if the RAD variable is a continuous value (such as vehicle miles traveled by a household), then the number of possible alternatives is, theoretically, infinite.

The literature review clearly highlights the absence of a single or prevalent software framework for developing RAD across various application domains. Several computer programming languages, matrix programming languages, and statistical software packages have been employed for developing RAD-based frameworks. Thus, RAD framework conceptualization can be platform-agnostic. This fact is particularly beneficial in the current context because no inherent limitations exist in adopting an open-source software platform for RAD implementation. Thus, the framework developed can be widely deployed by Federal Highway Administration (FHWA) without any constraints.

The authors of the report noted that the most common experimental set-up for RAD generation seen in this review was a statistical or econometric model framework assumed to represent the data-generation process of the variable of interest. Within the model framework, independent variables (either based on observed data or random realizations) and their impact (defined by coefficient values) on the dependent variable are assumed. Using the independent variable distribution and coefficient values, a latent propensity (or utility) is computed based on the model system. For example, if the dependent variable is a count variable and the model system is an NB system, a single propensity is generated with an appropriate overdispersion component. However, if the model system represents an unordered discrete outcome model, alternative specific utilities are generated. After the model structure is assumed, a random error term is added to (or multiplied by) the propensity or utility generated. Subsequently, a choice scenario is determined for each record based on the properties of the model. This choice scenario results in the formation of the RAD dataset. The process is repeated several times to generate multiple copies of the datasets. The choice variable of interest, such as crash occurrence, may possibly be a result of several layers of decisions, such as trip, route, and time-of-day decisions. The research team's vision was to use a similar approach to this framework to generate RAD after the conceptual framework for the study was finalized.

Additionally, researchers observed during their literature review that the embedded RAD frameworks were consistently single-level frameworks; in other words, the underlying decision process consisted of only one layer of decisions. Earlier research in modeling crash occurrence related crash occurrence to roadway geometry and traffic volume under prespecified assumptions of what variables would influence crash occurrence (for example, AADT and lane width). This study's research effort was the first to attempt the development of RAD datasets using a multilayered decision process; as such, researchers anticipated the effort would be challenging. Hence, the research team was cautious in determining the number of decisions—and thus the number of layers—that could be modeled to determine the crash dependent variables in this study. As the study commenced in full swing, the research team incorporated multilayered complexity within the decision processes to enhance the current state of RAD generation across multiple transportation domains.

# CHAPTER 3. MACROSCOPIC APPROACH FRAMEWORK

The macroscopic approach has two steps. The first step is to generate roadway data, such as AADT, truck percentage, lane width, and shoulder width. The second step is to generate crash counts by crash type and severity, based on roadway-level data and statistical and econometric methodologies.

## ROADWAY DATA GENERATION

The basic philosophy for generating roadway data follows the principle of the Markov chain (Gagniuc 2017). Given a sequence of random variables $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, … , $\mathbf{X}_n$ (where **X** represents a vector of roadway characteristics), the indices (1, 2, 3, …, n) represent the walking state in time (where the indices represent the contiguous sites for roadway network). The Markov chain assumes that the roadway characteristics in the next state depends only on the previous state. A higher probability may exist that the values for these roadway characteristics on a given site will remain the same as those of the previous site, with lower probabilities that the values will change to other possible combinations.

Overall, a transition probability matrix is required by the Markov chain principle. This matrix presents the probabilities for possible roadway characteristics on one site, given the roadway characteristics on the previous site. Two types of transition probability matrices are included in the Markov chain principle: one matrix is for a discrete Markov chain, and the other matrix is for a continuous Markov chain. The discrete Markov chain transition probability matrix is applied to the categorical variables in this study—such as lane width, shoulder width, and speed limit—and is illustrated in figure 1, which uses numbers that are randomly created to provide a visual example.

$$Q = \begin{array}{c|ccccc} & C_1 & C_2 & C_3 & \dots & C_n \\ P_1 & 0.70 & 0.10 & 0.10 & & 0.10 \\ P_2 & 0.10 & 0.60 & 0.20 & & 0.10 \\ P_3 & 0.00 & 0.10 & 0.80 & & 0.10 \\ \dots & & & & & \\ P_n & 0.00 & 0.00 & 0.10 & & 0.90 \end{array}$$

**Figure 1. Equation. Transition probability matrix for discrete Markov chain.**

Where:

I = 1, 2, 3, ….

n = the *ith* combination of roadway characteristics, such as 1 = 12-ft lane width, 6-ft shoulder width, and 55-mph speed limit while 2 = 11-ft lane width, 4-ft shoulder width, and 45-mph speed limit.

$P_i$ = the *ith* combination of roadway characteristics for the previous site.

$C_i$ = the *ith* combination of roadway characteristics for the subject site.

A probability under row Pi and column Cn in the transition probability matrix = the probability of roadway characteristics to be the nth combination for the subject site, given the *ith* combination of roadway characteristics for the previous site.

Note that the probabilities in each row and column must sum to unity in figure 1.

When a variable is continuous, such as AADT or truck percentage, the transition probability matrix is usually represented by a distribution, which can be written as shown in figure 2.

$$P(X_t = x_t) = N(x_t, \sigma^2)$$

**Figure 2. Equation. Transition probability matrix with continuous variable.**

That is, the probability distribution of a roadway characteristic for the subject site follows a normal distribution, with a mean equal to the roadway characteristic for the previous site and a predefined variance $\sigma^2$.

To illustrate the theoretical process of roadway data generation using the Markov chain, one must first generate an initial probability matrix. This initial probability matrix is used to define the first site in the RAD, since this site cannot be defined from the transition probability matrix in the Markov chain. Table 2 shows a hypothetical probability distribution lookup table, which is used to generate all combinations of roadway characteristics for the first roadway site. For example, suppose four combinations of roadway characteristics are generated in the RAD, such as different combinations of values for lane width, shoulder width, and speed limit. A random number between 0 and 1 can be generated and compared to the cumulative probability in table 2 to define the roadway characteristics for the first site. For example, if the random number generated is equal to 0.8, then the roadway characteristics following under combination C will be set for the first site.

**Table 2. Initial probability table.**

| Roadway Characteristics | Initial Probability | Cumulative Probability |
|---|---|---|
| Combination A | 0.25 | 0.25 |
| Combination B | 0.30 | 0.55 |
| Combination C | 0.35 | 0.90 |
| Combination D | 0.10 | 1.00 |

The next step is to generate all remaining sites based on the first site and the transition probability matrix using the Markov chain principle. First, generate a transition probability matrix from the real data, with the transition probabilities listed in a look-up table (table 3). Following this first step, the first site has roadway characteristics of combination C. A random number between 0 and 1 can now be generated for the second site, and the random number can be compared to the cumulative probability in table 3, where the roadway characteristics for the previous site follow combination C to determine the roadway characteristics for the subject site. For example, if the random number generated is equal to 0.6, the roadway characteristics for the subject site will stay the same as they were. If the random number generated is equal to 0.95, then the roadway characteristics for the subject site will change to those values under combination D. This process will be repeated until the last site is generated.

**Table 3. Transition probability look-up table.**

| Roadway Characteristics for Previous Site | Roadway Characteristics for Subject Site | Transition Probability | Cumulative Probability |
|---|---|---|---|
| Combination A | Combination A | 0.70 | 0.70 |
| | Combination B | 0.10 | 0.80 |
| | Combination C | 0.10 | 0.90 |
| | Combination D | 0.10 | 1.00 |
| Combination B | Combination A | 0.15 | 0.15 |
| | Combination B | 0.75 | 0.90 |
| | Combination C | 0.05 | 0.95 |
| | Combination D | 0.05 | 1.00 |
| Combination C | Combination A | 0.10 | 0.10 |
| | Combination B | 0.15 | 0.25 |
| | Combination C | 0.65 | 0.90 |
| | Combination D | 0.10 | 1.00 |
| Combination D | Combination A | 0.05 | 0.05 |
| | Combination B | 0.05 | 0.10 |
| | Combination C | 0.05 | 0.15 |
| | Combination D | 0.85 | 1.00 |

## CRASH DATA GENERATION

After generating the roadway data, crash counts by crash type and severity are generated using the known model structures (for example, Poisson and NB models) and "realistic" relationships between crash counts and roadway-level characteristics. Details about how to define the "realistic" relationships between crashes and roadway variables are discussed in section "Crash Data Generation Inputs" below.

Suppose a crash prediction model or safety performance function (SPF) exists for the sites generated in the roadway data-generation process, and this model or SPF contains the coefficients for the relationships between crash counts and all roadway characteristics and other model-related parameters. This model or SPF can then be used to estimate the parameters of the assumed underlying distribution for each site and then simulate the crash count as a random variable. For example, using the two commonly used model frameworks, the Poisson and NB models, to illustrate the process. The Poisson distribution for the crash counts can be written as shown in figure 3.

$$Prob[y_i|\mu_i] = \frac{exp(-\mu_i)\mu_i{}^{y_i}}{y_i!}$$

**Figure 3. Equation. Poisson distribution for crash counts.**

Where:

$Prob[y_i|\mu_i]$ = probability of $y$ crashes occurring at site i.
$\mu_i$ = expected number of crashes at site i.

Given the vector of roadway site characteristics $X_i$ and the vector of coefficients **β** in the crash prediction model, the expected crashes $\mu_i$ can be predicted as shown in figure 4.

$$\mu_i = exp(\beta X_i)$$

**Figure 4. Equation. Crash prediction model given vector of roadway.**

If the overdispersion is accommodated by the crash prediction model, then the NB model is used to estimate the vector of coefficients **ββ**, and the expected crashes $\mu_i$ can be predicted as shown in figure 5.

$$\mu_i = exp(\beta X_i + \varepsilon_i)$$

**Figure 5. Equation. Crash prediction model with overdispersion accommodated.**

where:

$exp(\varepsilon_i)$ = error term assumed to follow gamma distribution with mean 1 and variance **Error! Digit expected.**$1/\sigma = k$.

k denotes the overdispersion parameter in the NB model.

Given the expected crashes $\mu_i$, the distribution of crash counts in the NB model can be calculated as shown in figure 6.

$$Prob[y_i|\mu_i] = \frac{\Gamma[(\sigma) + y_i]}{\Gamma(\sigma)y_i!}\left[\frac{\sigma}{(\sigma) + \mu_i}\right]^{\sigma}\left[\frac{\mu_i}{(\sigma) + \mu_i}\right]^{y_i}$$

**Figure 6. Equation. Crash prediction model with expected crash counts.**

Where $\Gamma$ is the common gamma function.

Regardless of which statistical and econometric model is used to estimate the expected crash counts $\mu_i$ for each site, its probability distribution function can be used to calculate both the probabilities and cumulative probabilities of observing 0, 1, 2, and up to $y_i$ crashes, respectively, for each site. For example, figure 3 and figure 6 illustrate the functions for Poisson and NB, respectively. Then, the actual crash counts can be determined for each site using a random variable generation procedure similar to the roadway data generation. For example, a random number between 0 and 1 can be generated and compared to the cumulative probabilities of different crash counts to define the final crash counts.

## DATA COLLECTION

The research team collected data from multiple sources to support generating the transition probability matrix for the Markov chain principle in the macroscopic approach. Specifically, the research team collected the data of all seven States in the current Highway Safety Information System (HSIS), including California, Illinois, Maine, Minnesota, North Carolina, Washington, and Ohio (FHWA n.d.). Additionally, the research team collected extra data from the Ohio,

Connecticut, and Florida departments of transportation (DOTs) to supplement the HSIS data (Ivan et al. 2021).

## ROADWAY FACILITY TYPES

In this study, the research team attempted to cover RAD generation for as many types of roadway facilities as possible, based on the facility categorizations in the current *Highway Safety Manual* (HSM) (American Association of State Highway and Transportation Officials (AASHTO) 2010). However, the team decided (based on experiences in other safety-related studies) that freeway ramp and ramp terminal data would be too challenging to attempt to obtain. Therefore, the team mainly focused on all the HSM roadway facility types except these two ramp types. However, the team also omitted the following facility types from consideration due to too-small sample sizes in their data: urban and suburban three-lane arterials with a center two-way, left-turn lane; urban and suburban five-lane arterials with a center two-way, left-turn lane; and rural eight-lane freeway segments. Specifically, the research team considered the following facility types:

- Rural two-lane, two-way roadways:
  - Two-lane, two-way undivided segments.
  - Three-leg unsignalized intersections with stop control on approaches to minor roads.
  - Four-leg unsignalized intersections with stop control on approaches to minor roads.
  - Four-leg signalized intersections.

- Rural multilane highways:
  - Four-lane undivided segments.
  - Four-lane divided segments.
  - Three-leg unsignalized intersections with stop control on approaches to minor roads.
  - Four-leg unsignalized intersections with stop control on approaches to minor roads.
  - Four-leg signalized intersections.

- Urban and suburban arterials:
  - Two-lane undivided arterials.
  - Four-lane undivided arterials.
  - Four-lane divided arterials (i.e., including a raised or depressed median).
  - Three-leg unsignalized intersections with stop control on approaches to minor roads.
  - Three-leg signalized intersections.
  - Four-leg unsignalized intersections with stop control on approaches to minor roads.
  - Four-leg signalized intersections.

- Freeway segments:
  - Rural four-lane divided freeway segments.
  - Rural six-lane divided freeway segments.
  - Urban four-lane divided freeway segments.
  - Urban six-lane divided freeway segments.
  - Urban eight-lane divided freeway segments.
  - Urban 10-lane divided freeway segments.

## ROADWAY CHARACTERISTICS

Roadway characteristics highly depend on data availability because both the initial probability and transition probability in the Markov chain for each characteristic must be calculated from the existing data. The research team attempted to include as many roadway characteristics as possible in RAD generation, as long as existing data supported the analysis.

Again, the research team defined roadway characteristics for their RAD generation mainly as they were defined in the HSM (AASHTO 2010). The following characteristics were included:

- Segment-related characteristics:
  - AADT.
  - Lane width.
  - Left shoulder width.
  - Right shoulder width.
  - Left shoulder type.
  - Right shoulder type.
  - Median width.
  - Median type.
  - Horizontal curve-related factors.
  - Vertical curve-related factors.
  - Centerline rumble strip.
  - Two-way left-turn lane.
  - Left shoulder rumble strip.
  - Right shoulder rumble strip.
  - Passing lane.
  - Lighting.
  - Roadside characteristics.
  - On-street parking.

- Intersection-related characteristics:
  - Major road AADT.
  - Minor road AADT.
  - Number of approaches with exclusive left-turn lanes.
  - Number of approaches with exclusive right-turn lanes.
  - Intersection lighting.
  - Skew angle.
  - Speed limit.
  - Left-turn signal phasing.
  - Number of approaches with right-turn-on-red prohibition.
  - Number of lanes to be crossed by a pedestrian.
  - Presence of school(s).
  - Number of alcohol stores.
  - Number of bus stops.

## CRASH TYPES AND SEVERITIES

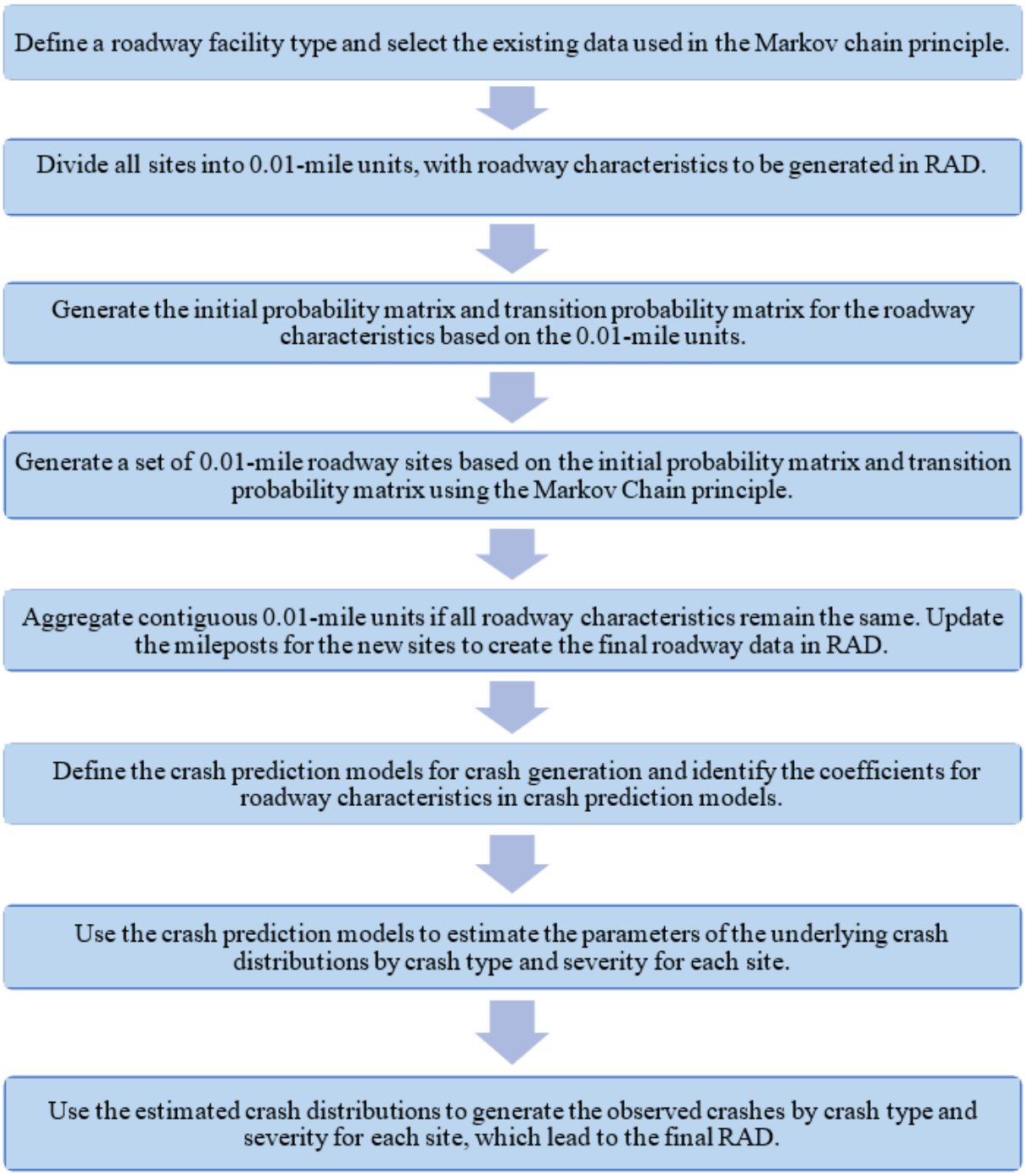
In crash data generation, the research team generated crash counts by different crash types and severities, as defined in *Model Minimum Uniform Crash Criteria* (NHTSA 2017). These crash counts include the following:

- Crash severity:
  - K—Fatal crash.
  - A—Incapacitating crash.
  - B—Non-incapacitating crash.
  - C—Possible crash.
  - PDO—Property damage-only crash.

- Crash type:
  - Single-vehicle crash.
  - Fixed-object crash.
  - Overturn/rollover crash.
  - Multivehicle crash:
    - Angle crash.
    - Head-on crash.
    - Rear-end crash.
    - Sideswipe—same-direction crash.
    - Sideswipe—Opposite-direction crash.

### Macroscopic Approach Consolidation and Enhancement

The overall procedure for generating RAD for the consolidated frameworks for both roadway and crash data generation using the macroscopic approach is presented in figure 7; RAD were generated by facility type.




Source: FHWA.

**Figure 7. Flowchart. RAD generation using a macroscopic approach.**

In terms of a roadway segment-related facility, the research team's first step was to determine the data to be used for generating the initial probability matrix and transition probability matrix in the Markov chain principle. The second step was to divide the original data into 0.01-mi units with the roadway characteristics to be generated in RAD. Researchers performed this second step

because 0.01 mi is usually the smallest resolution in most real data, and dividing all sites into the same length could help prevent biased results. Next, these two matrices were generated based on the 0.01-mi units for all roadway characteristics. After this step, researchers could operate the simulation to generate a set of 0.01-mi roadway sites based on the random number criteria described in the section "Crash Data Generation." These contiguous sites could then be aggregated into longer roadway segments based on the values of roadway characteristics. These steps then culminated with the research team generating the final roadway data in RAD.

At this point, researchers could then generate crash counts by crash type and severity, based on the roadway data. Selecting and defining crash prediction models and coefficients for roadway characteristics were critical steps in predicting the expected crashes for each site. Two options were primarily considered: The first option was using models and parameters in existing research and products, such as the HSM (AASHTO 2010). The second option was using coefficients estimated by the research team. Researchers predicted that estimating coefficients would be challenging. This challenge was due to data limitations, because none of the existing datasets contained all the roadway characteristics to be generated in RAD. Thus, the statistical significance of the parameters estimated in the models for all crash types and severities could not be guaranteed. Therefore, the team selected the first option, which might result in more flexibility and variation in the data.

This option was mainly applicable to the segment-related facilities. The Markov chain was not directly applicable to intersections because intersections are not contiguous. Instead of using the transition probabilities in the Markov chain principle, researchers could simulate the roadway characteristics at intersections, based on the intersection's probability distributions, by intersection type in real data. The crash data-generation process was the same as the generation process for RAD for segments.

## ROADWAY DATA GENERATION INPUTS

During RAD generation, researchers should seek to generate roadway and crash data that are as realistic as possible. In terms of roadway data generation, some roadway characteristics are highly correlated, including lane width and shoulder width, roadway type, and speed limit. Thus, their correlations should be accounted for by the RAD frameworks, so researchers first conducted a correlation test for all available roadway characteristics under each facility type. This testing allowed researchers to identify variables that were highly correlated in real data and so should have been generated together in RAD. To do this testing, Cramer's V statistic was calculated between each pair of roadway characteristics to determine the strength of their correlation (Crewson 2006). This statistic can be expressed as shown in figure 8.

$$V = \sqrt{\frac{\chi^2 \times N}{(C-1, R-1)}}$$

**Figure 8. Equation. Cramer's V statistic.**

where:

$\chi^2$ = Chi-square test for correlation.
N = total number of observations.
C = number of categories in variable 1.
R = number of categories in variable 2.

A Cramer's V statistic equal to or greater than 0.3 indicated that the two variables were moderately or highly correlated and should be generated together. Remaining variables were then generated independently.

Based on the correlation test results, the variables that were generated together in the macroscopic approach by facility type were determined to be as follows:

- Rural two-lane, two-way roadways:
  - Two-lane, two-way undivided segments: AADT, right shoulder width, horizontal curve.
  - Three-leg unsignalized intersections with stop control on approaches to minor roads: Major road AADT, minor road AADT.
  - Four-leg unsignalized intersections with stop control on approaches to minor roads: Major road AADT, minor road AADT.
  - Four-leg signalized intersections:
    - Major road AADT, minor road AADT.
    - Major road AADT, number of approaches with exclusive left-turn lanes.
    - Number of approaches with exclusive left-turn lanes, number of approaches with exclusive right-turn lanes.
    - Intersection lighting, skew angle.

- Rural multilane highways:
  - Four-lane undivided segments: Right shoulder width, right shoulder type, lighting.
  - Four-lane divided segments: Right shoulder width, right shoulder type, speed limit.
  - Three-leg unsignalized intersections with stop control on approaches to minor roads:
    - Major road AADT, minor road AADT.
    - Major road AADT, intersection lighting.
  - Four-leg unsignalized intersections with stop control on approaches to minor roads:
    - Major road AADT, minor road AADT.
    - Major road AADT, intersection lighting.
    - Intersection lighting, skew angle.
  - Four-leg signalized intersections:
    - Major road AADT, minor road AADT.
    - Major road AADT, number of approaches with exclusive left-turn lanes.
    - Major road AADT, intersection lighting.

    - Number of approaches with exclusive left-turn lanes, number of approaches with exclusive right-turn lanes.
    - Intersection lighting, skew angle.

- Urban and suburban arterials:
  - Two-lane undivided arterials:
    - AADT, lighting.
    - Offset to fixed object, on-street parking type, speed limit.
  - Four-lane undivided arterials: Offset to fixed object, on-street parking type, speed limit.
  - Four-lane divided arterials (i.e., including a raised or depressed median):
    - Offset to fixed object, on street-parking type, speed limit.
    - Median width, lighting.
  - Three-leg unsignalized intersections with stop control on approaches to minor roads: Major road AADT, minor road AADT.
  - Three-leg signalized intersections:
    - Major road AADT, minor road AADT.
    - Minor road AADT, speed limit.
    - Number of approaches with exclusive left-turn lanes, number of approaches with exclusive right-turn lanes.
    - Number of approaches with exclusive left-turn lanes, left-turn signal phasing.
    - Number of approaches with exclusive left-turn lanes, number of lanes to be crossed by a pedestrian.
  - Four-leg unsignalized intersections with stop control on approaches to minor roads: Major road AADT, minor road AADT.
  - Four-leg signalized intersections:
    - Major road AADT, minor road AADT.
    - Number of approaches with exclusive left-turn lanes, number of lanes to be crossed by a pedestrian.
    - Number of approaches with exclusive left-turn lanes, left-turn signal phasing.

- Freeway segments:
  - Rural four-lane divided freeway segments:
    - Left shoulder width, right shoulder width.
    - AADT, median width.
  - Rural six-lane divided freeway segments:
    - Left shoulder width, right shoulder width.
    - Lane width, median width.
  - Urban four-lane divided freeway segments:
    - Left shoulder width, right shoulder width.
    - Lane width, speed limit.
    - Left shoulder rumble strip, right shoulder rumble strip, horizontal curve.
  - Urban six-lane divided freeway segments:
    - Left shoulder width, right shoulder width.
    - Lane width, median width.
    - Left shoulder rumble strip, right shoulder rumble strip, horizontal curve.
  - Urban eight-lane divided freeway segments:

    - AADT, median width.
    - Left shoulder rumble strip, right shoulder rumble strip, horizontal curve.
    - Left shoulder width, right shoulder width.
  - Urban 10-lane divided freeway segments:
    - Left shoulder width, right shoulder width.
    - Left shoulder rumble strip, right shoulder rumble strip, horizontal curve.

Researchers then generated both initial probability matrix and transition probability matrix for each group of variables and all remaining single variables by each facility type using the collected data. The initial probability matrix and transition probability matrix were used as the inputs for generating roadway data in the macroscopic approach. Due to the large data size of initial probability matrix and transition probability matrix for all facility types, researchers decided presenting the matrix values in this report was not feasible. Instead, the matrix tables will be included in the source codes accompanying the RAD software developed through this project.

## CRASH DATA GENERATION INPUTS

The key to generating crash data for RAD is to estimate the mean ($\mu_i$) of crashes for each site, based on the characteristics generated from the roadway data-generation process. As noted in the section on Macroscopic Approach Consolidation and Enhancement, this study uses crash prediction models to estimate the mean crashes for each site. The framework for crash prediction models is similar to that in the HSM, where the crash mean is first predicted using a base condition SPF and then adjusted by multiple adjustment factors (AFs), which are expressed as figure 9.

$$u_i = Y * SPF_{base} * AF_{i,1} * AF_{i,2} * \ldots * AF_{i,n}$$

**Figure 9. Equation. SPF adjusted by AFs.**

where:

Y = number of years of crashes to be generated.

$SPF_{BASE}$ = predicted crashes using the base condition SPF, with segment length and AADT used as predictors for segment-related facilities and AADT on both major and minor approaches used as predictors for intersection-related facilities.

AFs = adjustment factors related to all roadway characteristics generated from the roadway data-generation process.

Researchers collected the coefficients of base condition SPF and all AFs for each facility type. Most of the existing literature the research team looked at estimated SPFs by injury severity group instead of by individual severity level, especially for severe injury crashes, such as K and A crashes (Wang et. al 2022; Garber and Rivera 2010). Thus, researchers collected both the SPF coefficients and AFs by three injury severity groups, namely PDO, B + C, and K + A. Further, to account for crash randomness and crash generation variation, researchers generated a uniform distribution for each of the SPF coefficients and AFs, based on all the AF values. The collected coefficients and AFs and generated distributions are not presented in this report to hide the underlying parameters used in generating the data.

For each parameter, if multiple sources provided values, researchers could directly determine the lower and upper limits of the uniform distribution based on all the values. However, if researchers found that only one source included a parameter estimate, they calculated the lower limit of the uniform distribution as 10 percent below that parameter estimate. Similarly, they calculated the upper limit as 10 percent above that parameter estimate. A random value could then be selected for each of the SPF coefficients and AFs from the generated uniform distributions for each site, to be used in predicting the mean of crashes by each injury severity group.

Table 4 shows an example of uniform distributions generated for the SPF coefficients and AFs for the intersection facility, and table 5 shows an example of uniform distributions generated for the SPF coefficients and AFs for the segment facility. The numbers and letters used in these tables are fabricated to demonstrate the overall procedure for generating the uniform distributions.

**Table 4. Crash generation parameters for rural, two-lane, two-way, three-leg unsignalized intersections.**

| Variables | Example Generated Parameters and AFs | | | Uniform Distributions for Parameters and AFs | | |
|---|---|---|---|---|---|---|
| | **PDO** | **B+C** | **K+A** | **PDO** | **B+C** | **K+A** |
| Parameters for Base Condition SPFs | | | | | | |
| Major AADT | A1, A2 | B1, B2, B3 | C1, C2, C3 | U(A1, A2) | U(B1, B3) | U(C1, C3) |
| Minor AADT | D1 | E1, E2 | F1, F2, F3 | U(0.9*D1, 1.1*D1) | U(E1, E2) | U(F1, F3) |
| Parameters for AFs | | | | | | |
| Number of approaches with exclusive left-turn lanes: | | | | | | |
| 1 | G1, G2 | H1, H2 | NA | U(G1, G2) | U(H1, H2) | U(1, 1) |
| 2 | G3, G4 | H3, H4 | NA | U(G3, G4) | U(H3, H4) | U(1, 1) |
| Overdispersion for NB | | | | | | |
| *k* | J1, J2 | K1, K2 | L1, L2 | U(J1, J2) | U(K1, K2) | U(L1, L2) |

**Table 5. Crash generation parameters for rural, two-lane, two-way, undivided highways.**

| Variables | Example Generated Parameters and AFs | | | | Uniform Distributions for Parameters and AFs | | | |
|---|---|---|---|---|---|---|---|---|
| | PDO | | B+C | K+A | PDO | | B+C | K+A |
| Parameters for Base Condition SPFs | | | | | | | | |
| Intercept | A1, A2 | | B1, B2, B3 | C1, C2, C3 | U(A1, A2) | | U(B1, B3) | U(C1, C3) |
| AADT | D1 | | E1, E2 | F1, F2, F3 | U(0.9*D1, 1.1*D1) | | U(E1, E2) | U(F1, F3) |
| Parameters for AFs | | | | | | | | |
| ≤9-ft lane width | G1, G2 | | H1, H2 | J1, J2 | U(G1, G2) | | U(H1, H2) | U(J1, J2) |
| 10-ft lane width | G3, G4 | | H3, H4 | J3, J4 | U(G3, G4) | | U(H3, H4) | U(J3, J4) |
| 11-ft lane width | G5, G6 | | H5, H6 | J5, J6 | U(G5, G6) | | U(H5, H6) | U(J5, J6) |
| ≥12-ft lane width | G7, G8 | | H7, H8 | J7, J8 | U(G7, G8) | | U(H7, H8) | U(J7, J8) |
| Right Shoulder Width | | | | | | | | |
| 2 ft | K1, K2 | | L1, L2 | M1, M2 | U(K1, K2) | | U(L1, L2) | U(M1, M2) |
| 4 ft | K3, K4 | L3, L4 | M3, M4 | U(K3, K4) | U(L3, L4) | U(M3, M4) | — | — |
| 6 ft | K5, K6 | L5, L6 | M5, M6 | U(K5, K6) | U(L5, L6) | U(M5, M6) | — | — |
| ≥8 ft | K7, K8 | L7, L8 | M7, M8 | U(K7, K8) | U(L7, L8) | U(M7, M8) | — | — |
| Overdispersion for NB: $k = \beta/Length$ | | | | | | | | |
| $\beta$ | N1<br>N2 | | O1<br>O2 | P1<br>P2<br>P3 | U(N1, N2) | | U(O1, O2) | U(P1, P3) |

—No data.

*Variables are insignificant at a 90-percent confidence level.

The mean crashes by the three injury severity groups were calculated using the uniform distributions generated in table 4 and table 5, respectively. Then, the crash counts were further simulated using two commonly used crash prediction modeling frameworks: the Poisson and NB models (figure 2 and figure 4). Next, a uniform distribution between 0.6 and 0.8 will be used to generate a weight for NB crash counts for each site, and the final crash counts for each site will be calculated by incorporating the weighted Poisson crash counts. Researchers assigned a greater weight to the NB model because it could account for overdispersion, which is a common issue in crash data. Mixing the Poisson and NB models in the crash generation process may further help account for crash generation variations for the RAD.

After researchers generated crashes for each of the three severity groups, they further disaggregated crash counts for the B+C group into individual B and C crash counts by randomly selecting and multiplying a proportion of B crashes over B and C in total from a pooled sample. They created a pooled sample using the real crash data for the specific facility type collected in this study. They applied the same procedure to K and A crashes and generated crash counts by crash type by each crash severity using a similar approach.

Although crashes could then be generated by both injury severity and crash type for each facility type, the generated crash counts still might not be feasible and realistic, because the research team collected the parameters for both the base condition SPFs and the AFs used in crash generation from multiple sources. The selected parameters are not presented in this report to hide the underlying parameters we used in generating the data to preserve the value of the RAD generated for testing data analysis approaches. Other literature that was not considered might use a different modeling formula, which could suggest a different range of values for the predicted mean of crashes.

For example, if the literature estimates crash prediction models using "Number of Years*365" as the offset, the estimated intercept will be relatively small, and the predicted mean of crashes in the RAD will be extremely low, leading to a zero crash count in the final crash generation. However, if the literature estimates crash prediction models without offset, the estimated intercept will be relatively large, and the predicted mean of crashes in the RAD will be extremely high, leading to unexpectedly high crash counts in the final crash generation. Therefore, the original uniform distributions generated for the intercepts needs to be calibrated to make the generated crash counts realistic. To accomplish this, crash counts are first generated by the three injury severity groups using the initial parameters for each facility type, then the crash counts by the injury severity group are also calculated using the actual crash data collected in this study, based on the same data size as the generated crash counts. For example, if the crash counts are generated from *N* intersections or *M*-mile segments, the crash counts are also calculated from *N* intersections or *M*-mile segments from the collected crash data. Next, a calibration factor is calculated (figure 10) for each of the injury severity groups for each facility type, and the calibration factor is added to the original uniform distribution for the intercept to calibrate the crash generation framework.

$$Ln\left(\frac{Observed\ Crash\ Counts}{Generated\ Crash\ Counts}\right)$$

**Figure 10. Equation. Calibration factor formula for crash count generation.**

# CHAPTER 4. MICROSCOPIC APPROACH FRAMEWORK

This section will outline the research team's development of a high-resolution, disaggregate realistic artificial data generator (DREDGE). The conceptualization of the disaggregate DREDGE framework, the data compiled to build the DREDGE, and the implementation procedures are described. In this research, the authors proposed a general framework of RAD generation embedded with heterogeneous causal structures that generated crash data by considering crash occurrence as a trip-level event impacted by trip-level factors, demographic characteristics, roadway facilities, and vehicle attributes.

The proposed framework was general enough to generate crashes for all roadway facility types, including segments and junctions. Additionally, the framework can generate data for different combinations of inputs, including modeling methods, model formulation, input specification, and unobserved heterogeneity. This chapter presents the motivation for this approach, along with conceptual framework details, a description of the data, and implementation details of the microscopic DREDGE framework.

## CONCEPTUAL FRAMEWORK

Safety analysis primarily focuses on identifying and quantifying the influences of factors that contribute to traffic collisions and the consequences of these factors. The authors of this report proposed and built a high-resolution, disaggregate data-generation process that mimics crash occurrences on transportation facilities at the trip level and accommodates the influence of a full range of crash-contributing factors. The proposed DREDGE recognizes that crashes result from travel decisions made by people. Hence, it is important to examine crash occurrence as a trip-level decision to mimic the true crash-generation process.

Considering such disaggregate treatments will allow the crash process incorporated in the DREDGE to resemble the true process more closely. For example, if data are simulated at the segment level, driver-related factors can be included either in an aggregate form or not at all. In contrast, all contributing factors—including trip-related factors—can be realistically incorporated if the data are simulated at the trip level. Researchers will consider trip-level attributes for DREDGE framework development that include driver and other occupant characteristics, vehicle characteristics, and roadway attributes.

The first step in the proposed framework was to evaluate the crash risk for each trip by employing trip-level travel information. Then, researchers generated detailed crash characteristics for trips identified for crash involvement. The DREDGE framework employs a suite of models to process trips with crashes, including crash type, crash severity, and crash location—as well as driver and vehicle characteristics.

Researchers developed the DREDGE framework by employing multiple datasets, including Strategic Highway Research Program 2 (SHRP2) Naturalistic Driving Study (NDS) data and Crash Report Sampling System (CRSS) data (Virginia Tech Transportation Institute 2020; NHTSA n.d.). The framework is organized around three specific modules: a) disaggregate trip information generation, b) crash data generation, and c) crash data aggregation. This report presents detailed documentation about each module in the next three sections.

### Disaggregate Trip Information Generation Module

The paradigm for regional travel demand modeling has undergone a transformation from an aggregate zonal level statistical framework (such as a four-step or trip-based model) to a disaggregate individual-level framework (tour level and/or activity-based models) (Kamel et al. 2019; Pinjari et al. 2008). The disaggregate frameworks accommodating various influences provide a representation of any individual's travel in continuous time and space. These influences include sociodemographic characteristics (such as income, age, household structure, education, and car ownership), employment characteristics (such as employment industry and location), transportation network characteristics (such as access to travel mode and travel time by mode), and built environmental measures (such as population density, land-use mix, and public transit density).

From these travel patterns, researchers retrieved high-resolution information for trips were retrieved, including trip start and end time, trip start and end location, trip characteristics (such as alone or group trip), vehicle used for trip, and precise route considered. In consultation with FHWA research staff, the research team recognized that developing a standalone, activity-based model system was beyond the scope of this study. Hence, the research team identified Polychotomous Choice Agent-Based Risk Model for Integrated Travel Demand and Network and Operations Simulation (POLARIS), a travel-demand modeling tool developed by Argonne National Laboratory, for use in the study in place of developing a new standalone modeling system. POLARIS is a well-established, activity-based travel demand modeling tool. It is high-performance, open-source, and has an agent-based modeling framework that includes traffic flow simulation, activity-based demand simulation, model building, and geographic information system (GIS) tools (Auld et al. 2016).

To use POLARIS data in the study, the research team reached out to Argonne National Lab personnel, who agreed to provide calibrated outputs from POLARIS. (For more information on POLARIS, please see Auld et al.'s (2016) article titled "POLARIS: Agent-Based Modeling Framework Development and Implementation for Integrated Travel Demand and Network and Operations Simulations.")

### Crash Data Generation Module

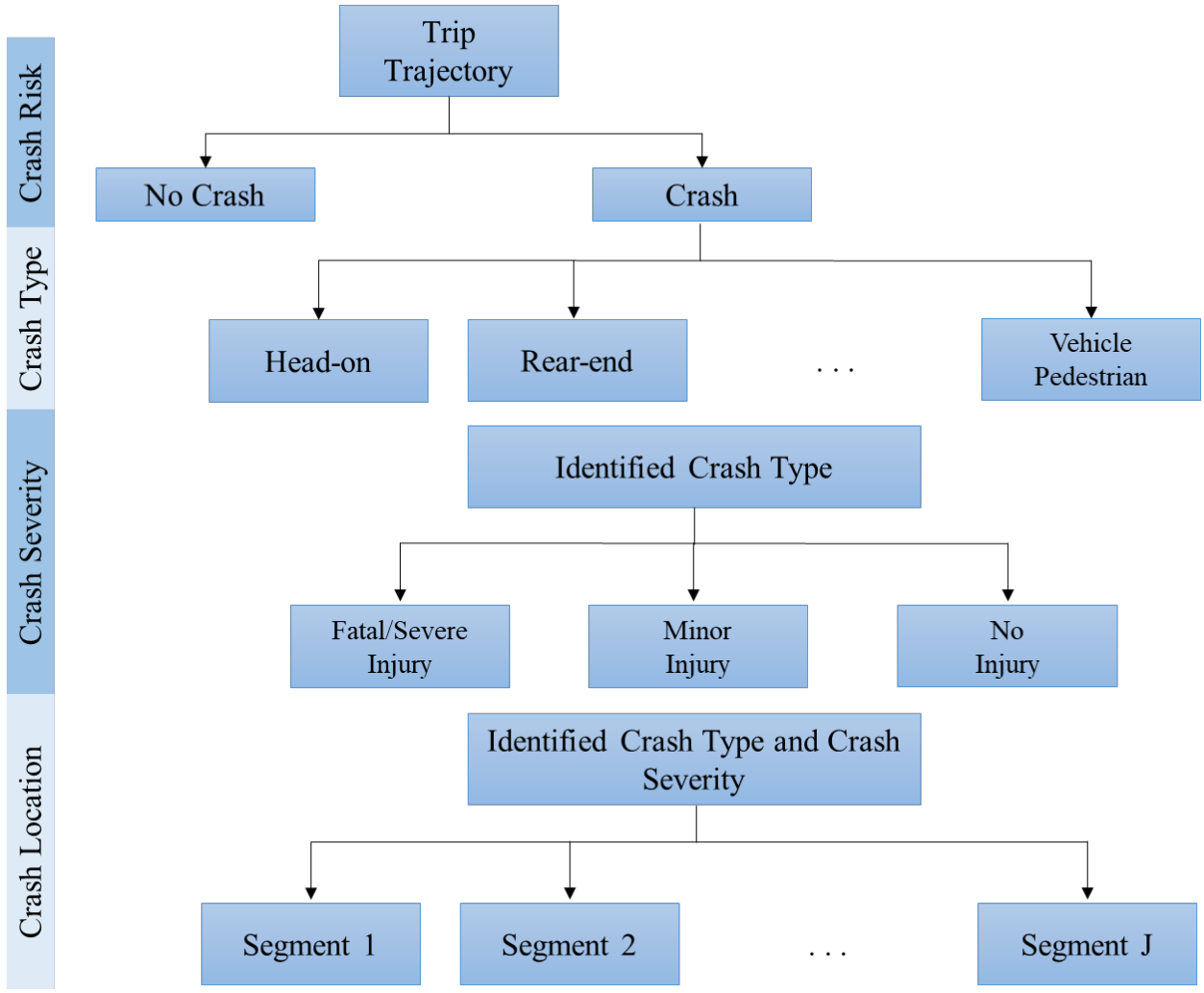
The objective of the crash data generation module is to generate crashes on the transportation system. The framework utilizes the detailed trip information from the trip information module and the disaggregate trip information generation to generate crashes. This process involves identifying the vehicles involved in the crash, crash location, injury severity of the occupants (such as fatal, incapacitating, non-incapacitating, and no injury), and crash type (such as head-on, rear-end, and vehicle-pedestrian). The framework employed for crash generation is described in the following paragraphs.

In the first step of the framework, the research team classified all the trips on the transportation system into two categories: No Crash and Crash. In urban regions, trips in a typical day amount to several million, and their data are likely to take up large amounts of storage space, with high-resolution details on routing characteristics with GIS coordinates. The team's proposed classification process allowed for a reduced number of trips to be used for crash data generation.

Given the relatively small proportion of crash-involved trips, this classification approach provides an elegant solution to computational and data burdens. The classification problem was modeled using a binary logit model. Specifically, the research team estimated the proposed binary model using SHRP2 NDS data (Virginia Tech Transportation Institute 2020). This NDS data provided the study with a large sample of trips and an associated indicator for trips involved in crashes. Further, using the NDS data allowed the research team to incorporate trip-level details, including average speed, travel distance, time of day, and vehicle type, into crash probability (Virginia Tech Transportation Institute 2020).

The second step of the framework took place after the research team identified and tagged the trips with crashes. During the second step, the team processed crash-tagged trips to determine detailed characteristics of the crashes, including crash type, crash location, and injury severity. Notably, crash type and crash severity have fixed and well-defined alternatives, but crash location alternatives are more complicated. Thus, depending on when crash location is examined, alternative structures for crash variable generation become possible. For example, one sequence can be as follows: Researchers estimate a trip-level model for the crash-tagged trips to identify the crash type, such as head-on, rear-end, or vehicle-pedestrian. Researchers then create a subsequent model for crash severity with this crash type and develop a crash location model that is conditional on crash type and severity. Notably, more information is available in the latter models in the sequence (i.e., additional independent variables can be included in the model estimation). For instance, if crash severity follows a crash type model, crash type can be included as an independent variable in the model.

The resulting crash location model can have crash type and crash severity as independent variables. A visual representation of the proposed structure for crash variable generation is presented in figure 11.



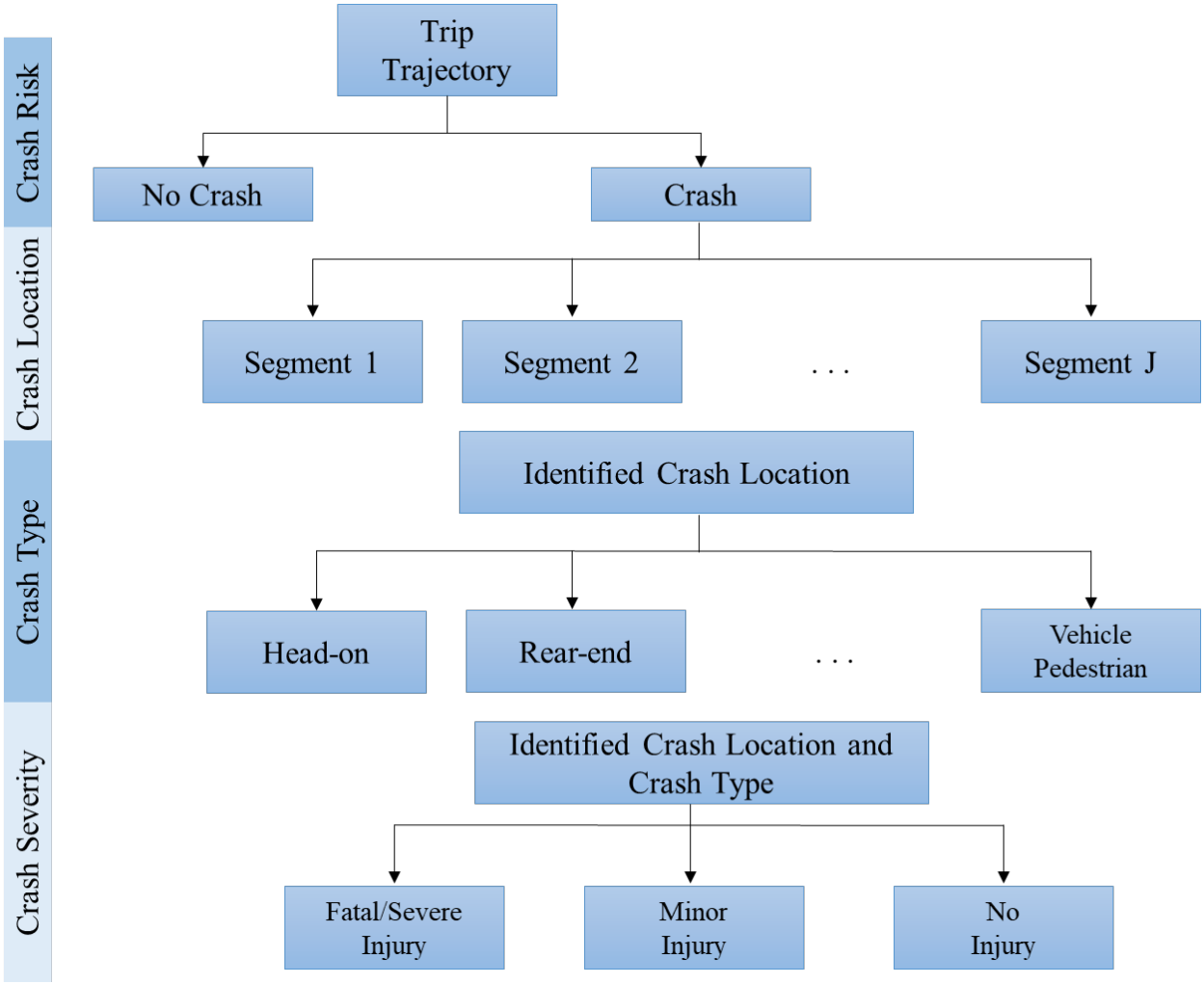

Source: FHWA.
. . . = other options not listed.

**Figure 11. Flowchart. DREDGE sequential approach I: crash risk to crash type to crash severity to crash location.**

Alternatively, the sequence of the variables can be altered to crash location, followed by crash type and crash severity. In this sequence, crash location model estimation will be based on trip-level characteristics, and crash type and crash severity variables will have access to location variables in the model. Figure 12 provides a potential model structure. As is evident, the two sequences described in figure 11 and figure 12 might result in different behavioral frameworks. Safety literature has placed significant emphasis on the role of crash location in modeling severity and crash type.

Hence, the research team chose the sequence listed in figure 12 for this study. The significant data processing and modeling efforts involved did not allow for testing the two sequences. Finally, the research team also generated driver and vehicle information (including the attributes

of other drivers and vehicles for multivehicle crashes) involved in crashes, based on the driver and vehicle characteristics of the crash trips.




Source: FHWA.
. . . = other options not listed.

**Figure 12. Flowchart. DREDGE sequential approach II: crash risk to crash location to crash type to crash severity.**

The final step of the crash data generation framework involved determining the appropriate model framework. Given that crash type and crash location were categorical variables, an MNL model framework was appropriate. For the severity variable—given the inherent ordered nature of the variable—a GOL model structure was employed.

### Crash Aggregation Module

The crash data generation module outputs are crash data—including crash type, crash severity, and crash location—and time and number of vehicle occupants involved in a crash for a typical

day of the year. However, for crash datasets, it might be necessary to aggregate data by facility type both at temporal resolution (such as crashes on a segment or intersection in a 6-mo period or multiple years) and spatial resolution (such as crashes in a zone or county). To generate aggregate data, the research team implemented framework to run a typical day multiple times with different random seeds to ensure duplication of the same crashes did not occur between runs.

## DATA COLLECTION

The microscopic DREDGE generator was developed employing three different datasets: SHRP2 NDS data from Virginia Tech Transportation Institute, CRSS data from NHTSA, and Chicago trip-level data from Argonne National Laboratory (Virginia Tech Transportation Institute 2020; NHTSA n.d.; Auld et al. 2016).

### SHRP2 NDS Data

To develop models for crash risk and crash location, researchers used SHRP2 NDS data, which had been collected during the NDS with cameras and sensors placed in participants' cars that tracked their driving over an extended period (Virginia Tech Transportation Institute 2020). For the current study, researchers requested five different data components from the NDS: time-series data, a trip summary table, an event table, driver demographics, and roadway characteristics data. Time-series data provided trip-related information, (e.g., speed, acceleration, deceleration, location, lighting, and airbag deployment information) for every tenth of a second.

The SHRP2 trip table provided trip details about all 5.4 million trips, including start time, end time, day of week, facility locations, facility speeds (mean/maximum speed), mean/maximum acceleration, mean/maximum deceleration, and headways versus distance traveled. The SHRP2 event table provided information on 1,951 crashes, including event severity, crash severity, traffic condition, and weather condition. Driver demographics included driver age, gender, income, and educational level. Finally, the roadway characteristics data provided information on corresponding road segments, including AADT, number of lanes, and roadway classification (arterial, major/minor collector, freeway) (Virginia Tech Transportation Institute 2020). These variables were included in analyzing crash risk, crash location, crash type, and crash severity appropriately for each analysis variables.

For their model development, the research team selected 1 million trips that did not result in crashes and 1,951 trips that resulted in crashes. The 1 million trips were randomly selected from a full sample of 5,512,900 trips (Hankey, Perez, and McClafferty 2016). For the 1,951 trips during which crashes occurred, 814 crashes were categorized as low-risk tire strikes and removed from the list of trips with crashes. This removal left 1,137 trips with crashes for the final model development.

### CRSS Data

The CRSS was used by the research team to develop models for crash type, drivers and vehicles, and crash severity. The CRSS database contains a nationally representative, weighted and stratified sample of road crashes collected and compiled from about 60 jurisdictions across the 50 States and District of Columbia. The database includes information from reports compiled by

police officers for roadway crashes involving at least one motor vehicle and resulted in property damage, injury, or death (NHTSA n.d.).

The research team obtained these CRSS databases from the U.S. Department of Transportation and NHTSA's National Center for Statistics and Analysis. NHTSA collected the crash data over the last 50 years from 1970 to the present. The data included information on a multitude of factors regarding crash situation and events, including driver characteristics (e.g., driver age, gender), vehicle characteristics (e.g., vehicle type, vehicle age, vehicle model), roadway design and operational attributes (e.g., speed limit, number of lanes), environmental factors (e.g., weather, lighting condition), and crash characteristics (e.g., hour, day, location, crash type, crash severity) (NHTSA n.d.).

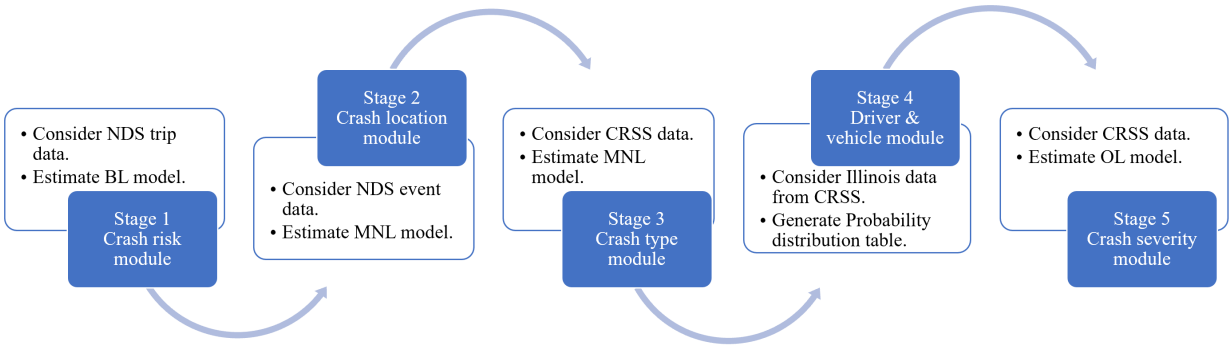
This study focused on the most recent 4 yr of CRSS data to develop crash type, driver, vehicle, and crash severity modules. Between 2016 and 2019, the CRSS crash database recorded a record 204,332 crashes involving 362,596 motor vehicles and 512,312 people (NHTSA n.d.). To prepare the final CRSS dataset for module development, the research team removed records with missing information on essential attributes. Thus, the final CRSS dataset consisted of about 113,983 crashes involving 211,311 motor vehicles and 298,382 people.

### Chicago Trip-Level Data

The research team used the Chicago trip-level data as an input when it implemented DREDGE. The team sourced the trip-level data from an activity-travel realization for an urban region, generated from the POLARIS model for the Chicago region. The data contained 2,256,502 trips, with information on trip data (start time, duration, etc.), driver demographics (age, education, household size, number of vehicles, number of workers, etc.), and roadway segments (segment length, AADT, number of lanes, and roadway type: major/minor arterial, freeway, collector or local, etc.) (Auld et al. 2016).

## MICROSCOPIC RAD MODULE DEVELOPMENT

The research team's proposed microscopic DREDGE framework for implementing the sequential approach presented in figure 12. This framework has five stages/modules of crash data generation, as illustrated in figure 13. Stage 1 (the crash risk module) evaluates a series of trips using a binary logit model to classify each trip as crash or no crash. In stage 2 (the crash location module), the location of each "crash" trip is determined using an MNL model. Stage 3 (the crash type module) is where the type of each crash is determined using an MNL model. In stage 4 (the driver and vehicle module), data on the driver(s) and vehicle(s) associated with each crash are generated using a probability distribution table. In stage 5 (the crash severity module), the severity of the crash is generated for each driver involved using an OL model. Each of these stages/modules is implemented sequentially in DREDGE using Python® programming language (Van and Drake 1995; Python Software Foundation 2023). In the following five sections, this report presents a detailed description of each of the five stages/modules involved in the microscopic approach framework.




Source: FHWA.

**Figure 13. Flowchart. DREDGE generator development for microscopic approach.**

## Crash Risk

The first stage of the DREDGE generator is the crash risk module. The goal of this module is to evaluate each trip and determine stochastically if a crash will occur during the trip. The research team used the SHRP2 NDS dataset to develop the crash risk model. The dataset included 1,137 trips that resulted in crashes and 1 million trips that did not (Virginia Tech Transportation Institute 2020). For this model, the research team removed any trips missing relevant trip or driver information. These removals resulted in 1,004 trips resulting in a crash and 774,873 trips that did not result in a crash. The data were then further filtered because trips with crashes accounted for only 0.13 percent of the total trips, making them difficult to model. Therefore, the trips not resulting in a crash were under sampled, as researchers randomly selected 10 percent to be used for analysis. The final dataset the research team used for model development contained 78,336 trips; 1,004 resulted in crashes, and 77,332 did not.

The research team used a binary logit model for modeling crash risk. The development of the model was based on removing statistically insignificant variables in a systematic manner. Since different datasets were used for modeling and implementation (SHRP2 NDS data were used for modeling and Chicago data were used for implementation), only those variables that were present in both datasets were considered when developing the model. Four variables were common in both datasets, including driver age and income and trip start time and length. Among these variables, driver age was found to have a significant impact on the trip crash risk. Drivers less than 30 years old (with teenage drivers being the most likely) and greater than 74 years old were found to be more likely to be involved in crashes than other drivers.

Finally, as the model was developed with under-sampled, noncrash trips, the constant in the binary logit model will need to be modified for applying the crash risk model for the full population. Researchers calibrated the constant to match the true population crash shares.

## Crash Location

The second stage of the DREDGE framework is the crash location module. The goal of this module is to evaluate the sequence of roadway segments traversed during a trip and determine stochastically where a crash will occur. The SHRP2 NDS dataset provided 1,004 crash records and was used to develop the crash location model (Virginia Tech Transportation Institute 2020).

For this module, any crashes that occurred at an intersection or crashes without a defined location were removed. However, there was missing information for multiple attributes for these 857 crashes. As a result, the research team imputed the missing information in alignment with existing distributions observed in the data.

For modeling the crash segments, the research team considered a sampling of alternative segment approaches to avoid computational complexity for large trips spanning several segments. The sampling process included the crash segment alternative and 29 additional segments randomly sampled from the trip segments. The research team's development of the model was based on removing statistically insignificant variables in a systematic manner. The results indicated that longer segments tended to have a higher risk of crash occurrence. Additionally, for each trip segment with more lanes, roadways with higher AADTs and collector roadways tend to have lower risks of crash occurrence.

### Crash Type

The goal of the third stage of the DREDGE generator, the crash type module, is to generate the type of crash that will occur based on trip and roadway variables. The research team used the CRSS dataset to develop this crash type module (NHTSA n.d.). In the CRSS dataset, researchers randomly selected 25,000 crashes (out of a total of 113,983 crashes) to use for developing the crash type module. The alternatives researchers considered for crash type were rear-end crash, head-on crash, angular crash, sideswipe crash, crash with fixed objects, crash with nonfixed objects, and nonmotorized crash. Since different datasets were used for modeling and implementation, the research team considered only variables that were present in both datasets when developing the module.

Rear-end crashes are used as the base alternative for this module, with angular crashes and crashes with fixed and nonfixed objects having a higher probability of occurrence, and head-on crashes, sideswipe crashes, and nonmotorized crashes having a lower probability of occurrence. Also, as the number of lanes increases, the probability of any crash, except for a rear-end crash, decreases. Crashes on freeways have a higher likelihood of sideswipe crashes and a lower probability of head-on crashes, angular crashes, crashes with nonfixed objects, and nonmotorized crashes. On weekdays, the probability of rear-end crashes increases, and the probability of head-on crashes and crashes with fixed and nonfixed objects decreases. During the morning peak (7 a.m. to 10 a.m.), the probability of crashes with fixed and nonfixed objects and nonmotorized crashes decreases. During the evening peak (4 p.m. to 7 p.m.), the probability of any crash other than a rear-end crash decreases.

### Drivers and Vehicles

The goal of the fourth stage of the DREDGE generator, the drivers and vehicles module, is to generate data for drivers and vehicles involved in crashes. Illinois crashes from the CRSS dataset are used for this module (NHTSA n.d.). From this data, the research team developed a probability distribution of different driver demographics (such as age, gender, and seatbelt use) and vehicle characteristics (such as type and age). The research team then used this distribution to generate driver and vehicle information for the generated crashes.

### Crash Severity

The goal of the fifth stage of the DREDGE framework, the crash severity module, is to generate the severity of the crash for each driver based on trip data, roadway information, driver demographics, vehicle information, and crash type. Of the 25,000 crashes that were used in the crash type module, driver information was available for 24,351 crashes, resulting in 42,039 drivers that researchers used in developing the crash severity module.

Researchers used an OL model to model crash severity. For the crash severity module the alternatives were PDO, minor, major, and severe. Since researchers used different datasets for modeling and implementation (CRSS and Chicago trips, respectively), they only considered those variables that were present in both datasets when developing the model (NHTSA n.d.; Auld et al. 2016). According to this model, drivers under 25 years old are less likely to experience a high-severity crash. Crashes that occur on freeways or involve more than a few lanes are more likely to result in high severity. Crashes that occur on weekdays or during peak hours are likely to be less severe. Using rear-end crashes, crashes with nonfixed objects, and nonmotorized crashes as a base, sideswipe crashes are less likely to result in severe crashes; meanwhile, head-on crashes, angular crashes, and crashes with fixed objects are more likely to result in severe crashes. Using automobiles, motorcycles, and buses as a base, drivers in utility vehicles and trucks are less likely to sustain severe injuries.

## MICROSCOPIC RAD MODULE IMPLEMENTATION

The research team employed Monte Carlo simulation for each DREDGE module's implementation (Spence 1983). Typically, the simulation process involves generating the cumulative probability function (CPF) for all alternatives using the module-specific model. Then, the chosen alternative is identified by generating a uniform random number between 0 and 1 and comparing it with the CPF. Across different modules, different CPF formulae are employed. The rest of the process remains stable across all modules. The implemented routines are validated and checked to ensure the model outcomes follow expected distributions. A detailed documentation on the implementation of microscopic DREDGE modules and validation checks implemented are presented in the following six sections.

### Crash Risk

The research team implemented the binary logit model in the DREDGE framework after developing the crash risk module. For the first step of the DREDGE framework, the Chicago data's 2,256,502 trips were used to simulate a day of traffic. Each trip was evaluated using the binary logit model for crash risk, and the probability of a crash was calculated. A random number between 0 and 1 was then generated. If the random number was less than the probability, then that trip was classified as resulting in a crash and added to the list of crashes for that day. This process repeats 365 times using the same set of trips to simulate a full year of traffic. Since a random number is generated for each trip, different trips will result in a crash each day; however, those with a higher probability of a crash are more likely to be selected. When using DREDGE to generate a full year of data, an average of 563 crashes per day occurred. This average equates to approximately 0.025 percent of the trips resulting in a crash. This average is

comparable to the NDS data that were used for model development, where 0.021 percent of the trips resulted in a crash (Virginia Tech Transportation Institute 2020).

### Crash Location

After developing the crash location model, the research team processed it in the DREDGE tool to determine the likely crash locations. A sequence of roadway segments is provided for each trip that occurs in the Chicago dataset. Each segment in this sequence is considered using the MNL model developed for crash location. This model calculates the probability of a crash for each segment. These probabilities are then combined to create a cumulative probability table. A random number is then generated from a uniform distribution from 0 to 1 and used to select the roadway segment. The random number is compared to each of the segments in the cumulative distribution table, and the segment with a probability greater than the generated number is the one where the crash occurred. The roadway information for this segment is then appended to the trip data to be returned to the user at the end.

### Crash Type

The third step of the DREDGE tool determines the type of crash for each trip where a crash is determined to have occurred. Trip and roadway information are used to calculate the probability of each type of crash using the MNL model. Since the same data are used for each day, a weekday variable is not in the original dataset. Instead, this variable is generated based on the specific day being evaluated. The generator will determine a random day of the week for the first day of generation and then assign each following day accordingly. Similar to the crash location module, this model calculates the probability of each crash type. These probabilities are then combined to create a cumulative probability table. A random number is then generated from a uniform distribution from 0 to 1, which is used to select the crash type. The random number is compared to the cumulative distribution table, and a crash type with a probability greater than the generated number is assigned. The crash type is then appended to the trip data, to be returned to the user at the end.

### Drivers and Vehicles

The first step in generating driver and vehicle information is determining the number of vehicles involved in the crash. This step is partially based on the crash type generated in the crash type module. If the crash type was defined as crash with fixed objects, crash with nonfixed objects, or nonmotorized crash, then it was considered a single vehicle crash. Otherwise, the number of cars was generated as 2 or 3. The probabilities calculated from the Illinois dataset for multivehicle crashes were an 88.1-percent likelihood of two vehicles being involved and an 11.9-percent likelihood of three vehicles being involved. Researchers generated the number using a cumulative probability table. Data was generated for each driver and vehicle involved in a crash once the number of vehicles involved was determined. Researchers assigned the first driver to have the same age as the age of the primary driver in the trip data; subsequently, all other information was generated using probability distributions observed from Illinois CRSS data (NHTSA n.d.).

### Crash Severity

The final step of the DREDGE generator determines the severity of the crash for each trip where a crash occurred. The trip, roadway, and generated crash information are used to calculate the probability of each crash severity using the OL model. Using the OL model, the probability for each crash severity across each driver is calculated. These probabilities are then combined to create a cumulative probability table. A random number is then generated from a uniform distribution from 0 to 1, which is used to select the crash severity. The random number is compared to all the crash severities in the cumulative distribution table, and the severity with a probability greater than the generated number is selected as the assigned crash severity for the driver. The crash severity is then appended to the driver data to be returned to the user at the end. The highest severity for all the drivers involved in a crash is then appended to the crash data to be returned to the user at the end.

### Validation Check

The research team used the DREDGE generator to generate a full year of data to test its accuracy. Data on crash type, driver and vehicle characteristics, and crash severity were generated and compared to the CRSS dataset (see also appendix A, which is in the second volume of this publication). The RAD was similar to the CRSS dataset. The biggest differences were rear-end crashes (a slight underestimation) and crashes with fixed and nonfixed objects (a slight overestimation). Additionally, differences were found in the number of vehicles and driver age distribution. These differences could be partially attributed to inputs from preceding modules. Crash severity results were well-aligned with the input data.

# CHAPTER 5. RAD GENERATION TOOL

The team developed an aggregated RAD software.[3] This software used Python and incorporated both the macroscopic approach and microscopic approach into a single platform (Python Software Foundation 2023). The team compiled all scripts and required inputs into a single executable file, which could be used directly by the end user to operate the RAD tool, with no need to install any software on their local computer.

## MACROSCOPIC RAD DATASETS

The research team implemented the macroscopic approach to provide the user with capabilities to generate RAD for 22 facility types, with 10 intersection facilities and 12 segment facilities. For each facility type, the software produced two data files: the roadway file (containing total crash counts by severity and roadway geometric characteristics) and the crash file (detailed crash severity and crash type information). The two files are cross-linked, and the corresponding ID column can be used to join the two files together when needed.

Researchers can use roadway data generated by the RAD software to compare the crash prediction accuracy among different statistical and econometric methodologies by five injury severity levels: K, A, B, C, and PDO crashes.[4] Additionally, the software can be used to investigate the underlying relationships between crash and/or crash severity and traffic and roadway characteristics (such as AADT, lane width, shoulder width, speed limit, and curvature). The crash data generated by the RAD software can be used to further incorporate crash types into the analysis. The crash data includes a total of 10 crash types: angle, front-to-front, front-to-rear, sideswipe from the opposite direction, sideswipe from the same direction, other multivehicle, fixed object, nonfixed object, overturn/rollover, and other single-vehicle.

## MICROSCOPIC RAD DATASETS

In terms of the microscopic approach, researchers implemented the disaggregate realistic artificial data generator (DREDGE) generator to produce three data files, as follows:

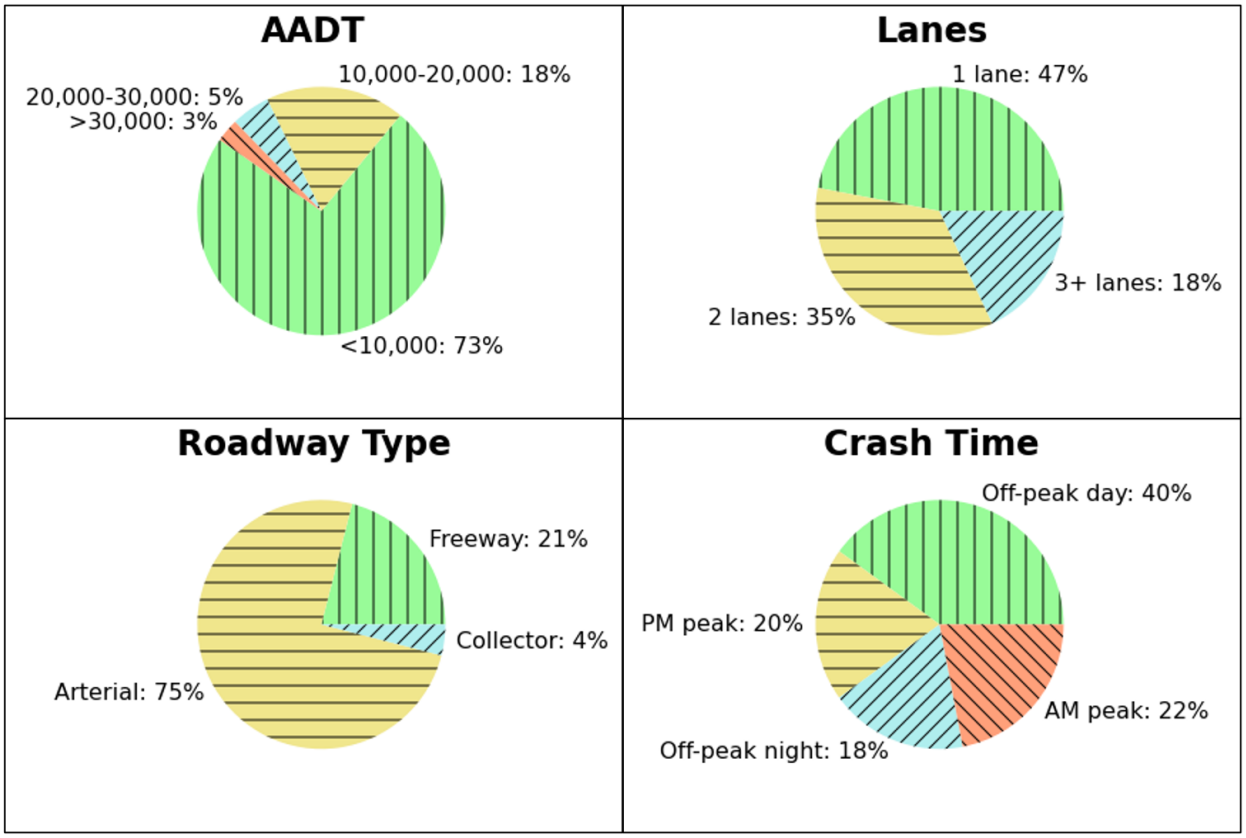
- Crash file: Contains information on crash details, such as location, type, and severity .
- Driver file: Contains information on each driver involved in a crash and their individual injury severity .
- Vehicle file: Contains information on each vehicle involved in a crash.

The three files are cross-linked, and columns from one dataset can be readily merged into the other two files as needed. The user can specify the number of years of crash data to be produced by the DREDGE generator, as well as the number of instances of data for that number of years. For example, a user can specify that they want two sets of 3 yr of crash data. When the

[3]FHWA. *DREDGE* (standalone RAD software).
[4]FHWA. *DREDGE* (standalone RAD software).

DREDGE generator is run, it will then produce two different crash files, two different driver files, and two different vehicle files, each containing three years of data.

The crash dataset produced by the DREDGE generator can be used in a variety of ways. To analyze the crash data produced, it can be aggregated by facility type, such as by crashes on a segment in a 6-mo period or multiple years. It can also be aggregated spatially, such as by crashes in a zone or county. Additionally, multiple variables can be used for analysis. A selection of these variables (and their distributions) are shown in figure 14. A user can analyze the data for roadway characteristics, such as for number of lanes, type of roadway, or AADT. A user can also analyze the data by crash characteristics, such as time of crash, type of crash, or severity of crash.




Source: FHWA.

**Figure 14. Chart. Sample variable distribution.**

The crash databases generated in this study can be employed for future use by transportation agencies to compare frequency models, severity models, crash type, and various other dimensions by facility type. The DREDGE generator the researchers developed can serve as a universal benchmarking system for alternative model frameworks in safety literature.

# CHAPTER 6. CASE STUDY DOCUMENTATION

## RAD GENERATION AND VALIDATION

### Generation and Validation Context

Crash prediction models are essential tools in improving highway safety because they can identify both the expected frequency of crash occurrences and their contributing factors, which can subsequently be addressed by safety countermeasures. Using information from vehicle crashes, these prediction models predict both the frequency of crash occurrence and the degree of crash severity. Although the quantity of crash data has generally expanded over the years, the quality of crash data has not necessarily improved in tandem with methodological advancements in crash analysis. Prior studies tended to concentrate on the modeling component of the overall crash prediction process by developing cutting-edge modeling approaches that offered better fits to the observed data, and the acquired data were implicitly assumed to be suitable representations of reality (Bonneson and Ivan 2013). In developing DREDGE, the research team hoped to mitigate this situation. The purpose of the case studies was to demonstrate the usefulness of the macroscopic DREDGE tool's applications and establish its capability to reliably generate realistic data. Researchers completed the following analyses to support this purpose:

1. Generated a collection of datasets of varying sizes for two different facility types using the RAD tool.

2. Estimated crash prediction models using each dataset and evaluated their predictive performance using mean absolute deviation (MAD) and mean-squared prediction error (MSPE) to select the best model.

3. Examined the estimated model parameters for stability from one dataset to another for each facility type.

The research team intended for these analyses to demonstrate how the RAD tool can serve as a testbed and help determine if any statistical models developed using the RAD can capture underlying relationships between independent variables and resultant crashes. Additionally, the team intended for these analyses to help guide and improve the practical application of statistical methods that influence highway safety policy, eventually leading to more effective safety countermeasures to reduce highway-related injuries and fatalities. The team also sought to demonstrate whether the data generated by the tool reliably represented real-world data and determine the tool's consistency from one random generation session to another.

### Data Generation for Case Studies

The team used datasets produced by the macroscopic RAD tool for two different facility types as an illustration to accomplish the purpose of the case studies. The tool was activated to generate 10 different datasets for each facility, as follows:

- Two sets of segments with 150, 300, 500, 750, and 1,000 mi each, with varying random seed values .

- Two sets of signalized intersections with 1,000, 2,000, 3,000, 4,000, and 5,000 intersections each, with varying random seed values.

**Crash Model Estimation**

A wide variety of methods have been historically applied to address the data and methodological issues associated with crash-frequency data. Articles by Lord and Mannering (2010) and Abdulhafedh (2017) provide detailed reviews of the key issues associated with crash-frequency data and discuss strengths and weaknesses of various methodological approaches that researchers have used to address these issues. The research team considered several different statistical approaches for estimating models with the RAD, as described in the next few sections.

***Poisson Regression Model***

Using standard ordinary least-squares regression (which assumes a continuous dependent variable) is not appropriate for crash prediction because the crash frequency observations are non-negative integers (Lord and Mannering 2010). In a Poisson regression model, the probability of roadway entity (segment, intersection, etc.) $I$ having $y_i$ crashes per some period (where $y_i$ is a non-negative integer) is given as shown in figure 15.

$$P(y_i) = \frac{EXP(-\lambda_i)\lambda_i^{y_i}}{y_i}$$

**Figure 15. Equation. Poisson regression model.**

where:

$P(y_i)$ = probability of roadway entity $I$ having $y_i$ crashes per period.
$\lambda_i$ = Poisson parameter for roadway entity $I$, which is equal to roadway entity $i$'s expected number of crashes per period $E[y_i]$.

Poisson regression models are estimated by specifying the Poisson parameter $\lambda_i$ (the expected number of crashes per period) as a function of explanatory variables, the most common functional form being $\lambda_i = EXP(\boldsymbol{\beta}\mathbf{X}_i)$, where $\mathbf{X}_i$ is a vector of explanatory variables, and $\boldsymbol{\beta}$ is a vector of estimable parameters

However, researchers have often found that crash data exhibit characteristics that make the application of the simple Poisson regression problematic (Washington, Karlaftis, and Mannering 2010). Poisson models cannot handle overdispersion and underdispersion, and they can be adversely affected by low sample means and produce biased results in small samples. These models have been applied to crash-frequency data by Jones, Janssen, and Mannering (1991); Miaou (1994), Greibe (2003), Al-Jabri (2015), Santamarina and Perez (2021), Khattak et al. (2021), and Shahzad et al. (2021).

***NB Regression (Poisson-Gamma) Model***

Some researchers have employed the use of NB as an alternative to Poisson regression (Abbas 2004; Amoros, Martin, and Laumon 2003; Hirst, Mountain, and Maher 2004; Lord and

Bonneson 2007; Miaou and Lord 2003). The NB (or Poisson-gamma) model is an extension of the Poisson model that was developed to overcome possible overdispersion in the data. The NB/Poisson-gamma model assumes that the Poisson parameter follows a gamma probability distribution governed by an additional parameter that is estimated. The NB distribution offers a simple way to accommodate the overdispersion, especially since the optimization function has a closed form, and the mathematics to manipulate the relationship between the mean and the variance structures is relatively simple. The NB model is derived by rewriting the Poisson parameter for each observation *I*, as shown in figure 16.

$$\lambda_i = EXP(\beta \mathbf{X}_i + \varepsilon_i)$$

**Figure 16. Equation. NB model.**

Where $EXP(\varepsilon_i)$ EXP(ε)is a gamma-distributed error term with a mean of 1 and a variance of α.

The addition of this term allows the variance to differ from the mean as $VAR[y_i] = E[y_i][1+ \alpha E[y_i]] = E[y_i]+\alpha E[y_i]$ (Lord and Miranda-Moreno 2008; Miaou and Lord 2003).

The Poisson-gamma/NB model is likely the most frequently used model in crash-frequency modeling. However, the model does have its limitations, most notably its inability to handle underdispersed data and dispersion-parameter-estimation problems when the data are characterized by low sample-mean values and small sample sizes (Washington, Karlaftis, and Mannering 2010).

***Random Parameter Model***

The random parameter model can be viewed as an extension of the random effects model. The latter is designed to address the correlations in crash data arising from spatial considerations (data from the same geographic region may share unobserved effects), temporal considerations (such as in panel data, where data collected from the same observational unit over successive periods may share unobserved effects), or a combination of the two. To account for such correlations, the following models can be considered (Lord and Mannering 2010):

- Random effects models where common, unobserved effects are assumed to be distributed over spatial and/or temporal units according to some distribution, and shared, unobserved effects are assumed to be uncorrelated with explanatory variables.

- Fixed effects models where common, unobserved effects are accounted for by indicator variables, and shared unobserved effects are assumed to be correlated with independent variables.

Random parameter models can be viewed as extensions of random effects models; however, rather than effectively only influencing the intercept of the model, random parameter models allow estimated parameter values to vary across all the observations in the dataset. Random parameters attempt to account for unobserved heterogeneity from one roadway site to another (Milton, Shankar, and Mannering 2008). Some researchers have applied this approach to crash frequency data. For example, Anastasopoulos and Mannering (2009) explored the use of random

parameter count models as a methodological alternative for analyzing crash frequencies. Their findings showed that ignoring the possibility of random parameters when estimating count data models can result in different marginal effects and subsequent inferences relating to the magnitude of the effect of factors affecting accident frequencies. This result can also be observed in studies from Bhat (2001) and Eluru, Bhat, and Hensher (2008).

To allow for such random parameters in count data models, estimable parameters can be written as , $\beta i = \beta + \varphi_i$ where $\varphi_i$ is a randomly distributed term—for example, a normally distributed term with mean zero and variance $\sigma^2$. With this equation, the Poisson parameter becomes $\lambda_i|\boldsymbol{\varphi}_i = EXP(\boldsymbol{\beta}_i\mathbf{X}_i)$ in the Poisson model and $\lambda_i|\boldsymbol{\varphi}_i = EXP(\boldsymbol{\beta}_i\mathbf{X}_i+\varepsilon_i)$ in the NB/Poisson-gamma, with the corresponding probabilities for Poisson or NB now $P(y_i|\boldsymbol{\varphi}_i)$. Each observation has its own parameters; the final model will often provide a statistical fit that is significantly better than a model with traditional fixed parameters. However, random parameter models are complex to estimate, and they are not guaranteed to improve predictive capability (Washington, Karlaftis, and Mannering 2010).

### *Poisson Lognormal Models*

Some researchers have proposed using the Poisson-lognormal model as an alternative to the NB/Poisson-gamma model for modeling crash data (Aguero-Valverde and Jovanis 2008; Lord and Miranda-Moreno 2008). The Poisson-lognormal model is similar to the NB/Poisson-gamma model; however, the $EXP(\varepsilon_i)$ term used to compute the Poisson parameter is lognormal rather than gamma-distributed. The Poisson-lognormal model addresses limitations of the NB model because it is more flexible in handling overdispersion.

### *Multivariate Poisson Lognormal Models*

Researchers have found this method necessary when instead of total crash counts, one wishes to model mutually exclusive categories of crash counts, such as counts by severity or type of collision (Aguero-Valverde and Jovanis 2008; N'Guessan and Langrand 2005; N'Guessan 2010). Modeling the counts of mutually exclusive crash types, as opposed to total crashes, with independent count models can be statistically invalid because the categorical counts are not strictly independent of one another. That is, the counts of crashes resulting in fatalities cannot increase or decrease without affecting the counts of crashes resulting in injuries and no injuries. Bivariate/multivariate models are used to resolve this problem, as they explicitly consider the correlations among the severity levels (Lord and Mannering 2010). The multivariate Poisson Lognormal approach calculates a covariance matrix that captures additional heterogeneity that other models do not.

### *Validating Crash Count Predictions*

An objective assessment of the predictive performance of a particular model can be made only through the evaluation of several goodness of fit (GOF) criteria. GOF measures used to conduct external model validation include MAD and MSPE (Washington, Karlaftis, and Mannering 2010). The model building effort in this case study used the first dataset at each total distance increment for model estimation and the other dataset for cross validation. This process was

reversed, and the data used for cross validation were then used for estimation, and the validation dataset was also used for estimation.

The GOF measures are calculated using the equations shown in figure 17 and figure 18.

$$\text{MAD} = \frac{\sum_{i=1}^{n} |Y_{model} - Y_{observed}|}{n}$$

**Figure 17. Equation. MAD.**

$$\text{MSPE} = \frac{\sum_{i=1}^{n} (Y_{model} - Y_{observed})^2}{n}$$

**Figure 18. Equation. MSPE.**

where:

$Y_{model}$ = predicted crash frequency.
$Y_{observed}$ = observed crash frequency.
$n$ = sample size.

### *Evaluating Parameter Stability*

The parameter estimates from the model estimation are used to check if the stochasticity embedded in the RAD generation process will be consistent for different random seeds used to generate data with the tool. To achieve this goal, researchers examined parameter estimates from the models for each dataset using revised Wald test statistics created by Hoover, Bhowmik, Yasmin, and Eluru (2022), as shown in figure 19.

$$\frac{sample\ parameter - population\ benchmark}{\sqrt{SE\ sample^2} - \sqrt{SE\ population^2}}$$

**Figure 19. Equation. Parameter test statistics.**

Where SE denotes the standard error for the corresponding sample.

If the parameter test statistic computed was higher than the 90 percent t-statistic, the result would indicate significant difference across the parameters. The research team employed this test statistic (figure 19) to compute revised t-statistics for all the parameters across all samples.

## Application To Rural Two-Lane, Two-Way Undivided Segments

### *Generation of Datasets*

Crash prediction models were estimated using each of the alternative approaches, as detailed in table 6.

**Table 6. Summary of developed models.**

| ID | Facility | Datasets | Models | Severity Level | No. of Models |
|---|---|---|---|---|---|
| 1 | Two-lane, two-way, undivided segment | Two sets each of 150, 300, 500, 750, and 1,000 mi | Poisson regression | K, A, B, C, PDO | 5×2 = 10 |
| | | | NB regression | K, A, B, C, PDO | 5×2 = 10 |
| | | | Poisson regression—random parameters | K, A, B, C, PDO | 5×2 = 10 |
| | | | NB—random parameters | K, A, B, C, PDO | 5×2 = 10 |
| | | | Univariate Poisson lognormal | K, A, B, C, PDO | 5×2 = 10 |
| | | | Multivariate Poisson lognormal | K, A, B, C, PDO | 5×2 = 10 |

The descriptive statistics of the datasets generated by the tool are shown in table 7 and table 8. Notably, the same size dataset (i.e., the two sets of 150 mi) resulted in different numbers of observations, as the datasets were randomly generated. The data contained horizontal curve data, roadway data, and crash data. The list of variables included is summarized in table 7 and table 8.

### *Model Estimation*

Researchers estimated crash prediction models using the parameters given in table 6 of this report. A total of 60 models were developed for rural two-lane undivided segments for each severity level: K, A, B, C, and O. For brevity, only the model parameter estimates and model fit statistics for the 1,000-mi dataset models are included in table 9, table 10, table 11, table 12, table 13, and table 14. Parameter estimates and model fit statistics for all models are provided in appendix B, which is in the second volume of this publication.

## Table 7. Descriptive statistics for continuous variables.

| Continuous Variables | Dataset 1 (*n* = 1,351): 150 mi | Dataset 2 (*n* = 2,742): 300 mi | Dataset 3 (*n* = 4,690): 500 mi | Dataset 4 (*n* = 4,140): 750 mi | Dataset 5 (*n* = 8,667): 1,000 mi | Dataset 6 (*n* = 1,361): 150 mi | Dataset 7 (*n* = 2,229): 300 mi | Dataset 8 (*n* = 4,270): 500 mi | Dataset 9 (*n* = 3,749): 750 mi | Dataset 10 (*n* = 7,050): 1,000 mi |
|---|---|---|---|---|---|---|---|---|---|---|
| Crash counts PDO | 0.938[a], 1.752[b] | 0.856, 1.615 | 0.853, 1.684 | 1.231, 2.52 | 0.950, 1.80 | 0.870, 1.751 | 0.988, 1.997 | 0.923, 1.841 | 1.432, 2.751 | 1.087, 2.13 |
| Crash counts K | 0.005, 0.094 | 0.003, 0.060 | 0.005, 0.092 | 0.007, 0.150 | 0.004, 0.088 | 0.004, 0.085 | 0.008 0.232 | 0.0014, 0.037 | 0.008, 0.230 | 0.0079, 0.27 |
| Crash counts A | 0.168, 0.565 | 0.137, 0.477 | 0.143, 0.504 | 0.223, 0.765 | 0.168, 0.583 | 0.153, 0.561 | 0.156 0.536 | 0.165, 0.629 | 0.247, 0.855 | 0.167, 0.63 |
| Crash counts B | 0.041, 0.247 | 0.041, 0.241 | 0.047, 0.278 | 0.060, 0.374 | 0.057, 0.323 | 0.036, 0.234 | 0.052 0.306 | 0.054, 0.336 | 0.073, 0.386 | 0.064, 0.370 |
| Crash counts C | 0.157, 0.454 | 0.165, 0.493 | 0.158, 0.486 | 0.282, 0.908 | 0.193, 0.552 | 0.196, 0.527 | 0.213 0.613 | 0.186, 0.618 | 0.397, 0.810 | 0.219, 0.641 |
| Pavement roughness | 104, 34.04 | 104.5, 34.656 | 105.3, 40.64 | 107.6, 40.30 | 105.9, 39.800 | 104.7, 33.54 | 104.9 34.96 | 105.1, 39.29 | 108.5, 40.83 | 106.7, 40.7 |
| Pavement condition | 40.170, 3.090 | 39.85, 3.077 | 23.0, 3.714 | 39.37, 3.791 | 39.33, 3.730 | 40.28, 3.173 | 39.63 3.18 | 39.59, 3.76 | 39.4, 3.728 | 39.24, 3.70 |
| Average super-elevation | 0.769, 2.603 | 0.769, 2.586 | −8.00, 2.576 | −8.183, 11.94 | −19.61, 9.240 | 0.802, 2.60 | 0.771 2.608 | 0.737, 2.59 | −9.14, 12.14 | −19.07, 9.74 |
| Curvature degree | 4.320, 8.435 | 5.217, 10.354 | 5.21, 9.30 | 0.00, 4.750 | 5.52, 10.91 | 5.155, 10.17 | 4.624 8.563 | 5.062, 8.943 | 3.122, 2.970 | 5.367, 10.27 |
| Arc angle | 36.53, 20.62 | 35.16, 21.904 | 10.0, 22.08 | 43.29, 12.42 | 36, 22.670 | 36.06, 22.72 | 40.75 18.904 | 37.79, 19.92 | 45.33, 5.557 | 37.41, 20.93 |
| Log (radius) | 7.686, 0.904 | 7.189, 0.943 | 7.17, 0.830 | 8.281, 0.842 | 7.153, 0.920 | 7.143, 0.931 | 7.694 1.094 | 7.463, 1.051 | 8.657, 0.304 | 7.397, 1.034 |
| Log segment length | −2.710, 1.066 | −2.72, 1.058 | −2.75, 1.056 | −2.383, 1.22 | −2.63, 1.101 | −2.69, 1.053 | −2.58 1.126 | −2.69, 1.087 | −2.29, 1.249 | −2.576, 1.150 |
| Vertical approach | 0.225, 0.828 | 0.293, 0.910 | 0.355, 0.895 | 0.504, 1.094 | 0.440, 0.960 | 0.232, 0.850 | 0.316 0.950 | 0.361, 0.913 | 0.548, 1.134 | 0.459, 0.990 |
| Vertical leaving | 0.239, 0.873 | 0.213, 0.822 | 0.221, 0.831 | 0.340, 1.017 | 0.26, 0.930 | 0.252, 0.905 | 0.250 0.874 | 0.236, 0.848 | 0.369, 1.055 | 0.277, 0.936 |
| Log (AADT) | 7.735, 0.570 | 7.394, 0.845 | 7.628, 0.832 | 7.381, 0.851 | 7.56, 0.804 | 7.739, 0.575 | 7.411 0.858 | 7.616, 0.835 | 7.357, 0.891 | 7.542, 0.820 |
| Grade | 0.353, 1.075 | 0.402, 1.090 | 0.462, 1.080 | 0.671, 1.30 | 0.56, 1.170 | 0.370, 1.05 | 0.446 1.145 | 0.474, 1.104 | 0.728, 1.35 | 0.594, 1.200 |

[a]mean.
[b]standard deviation.

**Table 8. Descriptive statistics for categorical variables (datasets 1–10).**

| Variable | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Dataset 5 | Dataset 6 | Dataset 7 | Dataset 8 | Dataset 9 | Dataset 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Shoulder Width | | | | | | | | | | |
| 0 ft | 42 | 80 | 280 | 208 | 465 | 16 | 69 | 260 | 194 | 413 |
| 2 ft | 454 | 845 | 939 | 1134 | 2029 | 20 | 697 | 860 | 1044 | 121 |
| 4 ft | 186 | 589 | 1196 | 1083 | 2070 | 444 | 467 | 1081 | 970 | 2876 |
| 6 ft | 387 | 462 | 620 | 593 | 1387 | 171 | 379 | 570 | 532 | 1824 |
| 8 ft | 282 | 755 | 1655 | 1122 | 2116 | 705 | 617 | 1499 | 1005 | 1816 |
| Speed Limit | | | | | | | | | | |
| 25 mph | 95 | 92 | 100 | 78 | 203 | 94 | 109 | 104 | 51 | 157 |
| 30 mph | 26 | 34 | 75 | 73 | 159 | 19 | 41 | 60 | 57 | 129 |
| 35 mph | 26 | 35 | 55 | 81 | 220 | 33 | 35 | 54 | 75 | 129 |
| 40 mph | 54 | 138 | 272 | 229 | 472 | 50 | 101 | 260 | 200 | 393 |
| 45 mph | 85 | 133 | 311 | 258 | 486 | 86 | 122 | 253 | 224 | 418 |
| 50 mph | 31 | 96 | 151 | 258 | 486 | 86 | 122 | 253 | 224 | 418 |
| 55 mph | 1037 | 2205 | 3726 | 3303 | 6339 | 1043 | 1747 | 3399 | 3025 | 5604 |
| Lane Width | | | | | | | | | | |
| 9 ft | 0 | 3 | 37 | 107 | 247 | 0 | 3 | 37 | 103 | 203 |
| 10 ft | 50 | 183 | 354 | 311 | 382 | 50 | 180 | 327 | 267 | 347 |
| 11 ft | 269 | 664 | 11075 | 919 | 1829 | 264 | 558 | 912 | 835 | 1651 |
| 12 ft | 1032 | 1881 | 3224 | 2803 | 5609 | 1042 | 1488 | 2992 | 2540 | 4849 |
| Lighting | | | | | | | | | | |
| Present | 77 | 201 | 353 | 344 | 530 | 72 | 172 | 316 | 328 | 530 |
| Not present | 1274 | 2530 | 4337 | 3796 | 7537 | 1284 | 2057 | 3954 | 3417 | 6520 |

**Table 9. Poisson regression estimates for dataset 10 (1,000 mi).**

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| Intercept | −25.066* | 4.175 | −7.434 | 0.678 | −4.478 | 1.004 | −3.609 | 0.560 | −2.099 | 0.250 |
| Ln (Segment Length) | 1.292 | 0.178 | 0.984 | 0.032 | 1.023 | 0.051 | 1.000 | 0.028 | 1.059 | 0.013 |
| Ln (AADT) | 1.389 | 0.303 | 0.813 | 0.050 | 0.533 | 0.075 | 0.469 | 0.040 | 0.536 | 0.018 |
| Shoulder Width | | | | | | | | | | |
| 8 ft | Base Level | | | | | | | | | |
| <2 ft | −0.580* | 1.059 | 0.317 | 0.140 | 0.656 | 0.214 | 0.324 | 0.129 | 0.231 | 0.062 |
| ≥2 ft < 4 ft | −16.357* | 1672.361 | 0.736 | 0.180 | −0.817 | 0.589 | −0.020 | 0.227 | 0.349 | 0.087 |
| ≥4 ft < 6 ft | −0.080 | 0.397 | 0.260 | 0.076 | 0.369 | 0.126 | 0.359 | 0.066 | 0.340 | 0.030 |
| ≥6 ft < 8 ft | −0.874 | 0.583 | 0.174 | 0.088 | 0.395 | 0.140 | 0.003 | 0.079 | 0.263 | 0.034 |
| Lane Width | | | | | | | | | | |
| >12 ft | Base Level | | | | | | | | | |
| ≤9 ft | 3.028 | 0.567 | 0.279 | 0.160 | 0.011 | 0.249 | 0.246 | 0.142 | 0.145 | 0.064 |
| 9.5 ft−10.5 ft | 0.411 | 0.593 | 0.129 | 0.091 | −0.376 | 0.149 | 0.179 | 0.076 | 0.188 | 0.034 |
| 11 ft−11.5 ft | −12.372* | 1520.181 | 0.478 | 0.229 | −0.341 | 0.393 | 0.343 | 0.194 | 0.223 | 0.089 |
| Speed Limit | | | | | | | | | | |
| 45 mph | Base Level | | | | | | | | | |
| 25 mph | 2.826* | 1.088 | 0.791 | 0.181 | 1.089 | 0.323 | 0.365 | 0.157 | 0.483 | 0.068 |
| 30 mph | −15.228* | 1864.827 | 0.387 | 0.199 | 1.239 | 0.303 | 0.560 | 0.150 | 0.081 | 0.078 |
| 35 mph | 0.773* | 1.250 | 0.642 | 0.184 | 0.696 | 0.329 | 0.038 | 0.167 | −0.359 | 0.088 |
| 40 mph | 1.575* | 1.096 | −0.676 | 0.214 | −0.479 | 0.374 | −0.677 | 0.163 | −1.044 | 0.084 |
| 50 mph | −14.297* | 1595.923 | 0.321 | 0.217 | −0.256 | 0.505 | −0.356 | 0.219 | −0.151 | 0.092 |
| 55 mph | −0.123 | 1.072 | −0.217 | 0.136 | 0.246 | 0.249 | −0.356 | 0.106 | −0.258 | 0.049 |
| Roadside Hazard Rating (Zegeer et al. 1988) | | | | | | | | | | |
| 3 | Base Level | | | | | | | | | |
| 4 | −1.069 | 0.547 | 0.045 | 0.084 | 0.030 | 0.138 | 0.146 | 0.072 | 0.069 | 0.032 |
| 5 | −1.395 | 0.678 | 0.115 | 0.103 | 0.279 | 0.164 | 0.233 | 0.089 | 0.180 | 0.040 |
| 6 | 0.054 | 0.689 | −0.025 | 0.134 | 0.430 | 0.204 | 0.161 | 0.116 | 0.159 | 0.052 |
| 7 | −0.715 | 1.066 | 0.237 | 0.171 | 0.558 | 0.279 | 0.110 | 0.159 | 0.251 | 0.068 |
| Pavement Condition | 0.249 | 0.068 | 0.021 | 0.012 | −0.019 | 0.018 | 0.008 | 0.011 | −0.002 | 0.005 |
| Pavement Roughness | 0.014 | 0.008 | 0.002 | 0.001 | −0.002 | 0.002 | 0.001 | 0.001 | 0.000 | 0.001 |

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** |
| Presence of Lighting | −1.631 | 1.220 | −0.180 | 0.164 | −0.491 | 0.288 | −0.237 | 0.152 | −0.417 | 0.070 |
| Presence of Horizontal Curve | −14.516* | 1042.264 | −1.408 | 0.411 | −14.329* | 262.136 | −1.225 | 0.336 | 0.345 | 0.073 |
| Vertical Curve Approaching Slope | −3.043* | 5.011 | 0.173 | 0.157 | −0.359 | 0.289 | 0.158 | 0.106 | −0.053 | 0.053 |
| Vertical Curve Leaving Slope | −2.276* | 5.020 | −0.104 | 0.114 | 0.163 | 0.324 | 0.332 | 0.107 | 0.077 | 0.051 |
| Grade | 2.775* | 5.118 | −0.058 | 0.175 | −0.061 | 0.374 | −0.336 | 0.140 | −0.005 | 0.065 |
| Curvature Degree | −0.007 | 0.024 | 0.006 | 0.003 | 0.002 | 0.005 | 0.000 | 0.003 | 0.002 | 0.001 |
| Fit Statistics | | | | | | | | | | |
| AIC | 443.88 | N/A | 5693.6 | N/A | 3018.8 | N/A | 6641.1 | N/A | 15812 | N/A |
| BIC | 635.97 | N/A | 5885.67 | N/A | 3210.9 | N/A | 6833.24 | N/A | 16004.4 | N/A |

—No data.
*Variables are insignificant at a 90-percent confidence level.
N/A = not applicable; std. error = standard error.

**Table 10. NB estimates for dataset 10 (1,000 mi).**

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| Intercept | −18.145* | 6.912 | −5.918 | 0.948 | −4.412 | 1.829 | −3.124 | 0.766 | −2.002 | 0.389 |
| Ln (Segment Length) | 0.976 | 0.339 | 1.006 | 0.046 | 1.200 | 0.097 | 1.049 | 0.039 | 1.058 | 0.020 |
| Ln (AADT) | 1.206 | 0.512 | 0.847 | 0.064 | 0.204 | 0.108 | 0.438 | 0.049 | 0.452 | 0.025 |
| Shoulder Width | | | | | | | | | | |
| 8 ft | Base Level | | | | | | | | | |
| <2 ft | 0.005 | 1.406 | 0.322 | 0.210 | −0.515 | 0.631 | 0.837 | 0.164 | 0.240 | 0.094 |
| ≥2 ft < 4 ft | −1.063 | 1.226 | 0.538 | 0.119 | 1.019 | 0.243 | 0.528 | 0.103 | 0.292 | 0.051 |
| ≥4 ft < 6 ft | 1.001 | 0.828 | 0.189 | 0.124 | 0.558 | 0.259 | 0.558 | 0.104 | 0.258 | 0.051 |
| ≥6 ft < 8 ft | 0.338 | 0.910 | 0.115 | 0.144 | 0.482 | 0.296 | 0.367 | 0.122 | 0.246 | 0.059 |
| Lane Width | | | | | | | | | | |
| >12 ft | Base Level | | | | | | | | | |
| ≤9 ft | −49.933* | 38132 | 0.116 | 0.222 | 0.514 | 0.444 | 0.157 | 0.183 | 0.256 | 0.095 |
| 9.5 ft−10.5 ft | 0.920 | 0.818 | 0.027 | 0.123 | 0.051 | 0.247 | 0.126 | 0.100 | 0.091 | 0.052 |
| 11 ft−11.5 ft | −46.906* | 65065 | −0.065* | 0.308 | 0.120 | 0.675 | 0.167 | 0.255 | 0.505 | 0.130 |
| Speed Limit | | | | | | | | | | |
| 45 mph | Base Level | | | | | | | | | |
| 25 mph | 3.088* | 1.793 | 0.303 | 0.313 | −0.222 | 0.660 | 0.549 | 0.252 | 0.041 | 0.131 |
| 30 mph | −0.097* | 3.564 | 0.353 | 0.300 | 1.175 | 0.552 | 0.817 | 0.237 | 0.041 | 0.136 |
| 35 mph | 2.391* | 1.939 | −0.001 | 0.332 | 1.096 | 0.529 | 0.717 | 0.241 | −0.295 | 0.144 |
| 40 mph | −46.975* | 4453 | −0.920 | 0.275 | −0.604 | 0.498 | −0.553 | 0.230 | −1.110 | 0.120 |
| 50 mph | −49.941* | 6182.0 | −0.552 | 0.341 | −0.248 | 0.612 | 0.211 | 0.258 | 0.004 | 0.120 |
| 55 mph | 0.634 | 1.399 | −0.396 | 0.166 | −0.546 | 0.349 | −0.101 | 0.153 | −0.330 | 0.071 |
| Roadside Hazard Rating (Zegeer et al 1988) | | | | | | | | | | |
| 3 | Base Level | | | | | | | | | |
| 4 | 1.392 | 1.159 | 0.122 | 0.120 | −0.146 | 0.238 | −0.092 | 0.100 | 0.015 | 0.049 |
| 5 | 0.485 | 1.391 | 0.083 | 0.152 | 0.073 | 0.301 | 0.093 | 0.125 | 0.051 | 0.064 |
| 6 | 0.867 | 1.585 | 0.019 | 0.191 | 0.313 | 0.371 | 0.205 | 0.156 | 0.251 | 0.080 |
| 7 | 2.325 | 1.695 | −0.078 | 0.261 | −0.083 | 0.536 | 0.348 | 0.206 | 0.180 | 0.110 |
| Pavement Condition | 0.081 | 0.125 | −0.014 | 0.018 | 0.040 | 0.038 | −0.006 | 0.015 | 0.014 | 0.008 |
| Pavement Roughness | −0.002 | 0.012 | 0.003 | 0.002 | 0.002 | 0.004 | 0.001 | 0.002 | 0.000 | 0.001 |

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** |
| Presence of Lighting | 0.688 | 1.347 | −0.404 | 0.238 | −0.240 | 0.449 | −0.722 | 0.211 | −0.531 | 0.102 |
| Presence of Horizontal Curve | −42.583* | 45853 | −2.856 | 1.032 | −1.361 | 1.065 | −1.429 | 0.462 | 0.186 | 0.113 |
| Vertical Curve Approaching Slope | 0.163 | 1.167 | 0.528 | 0.244 | −0.008 | 0.528 | 0.125 | 0.157 | −0.011 | 0.075 |
| Vertical Curve Leaving Slope | 1.851 | 2.132 | 0.107 | 0.151 | −0.243 | 0.409 | 0.079 | 0.135 | −0.050 | 0.064 |
| Grade | −1.984 | 2.366 | −0.546 | 0.274 | 0.033 | 0.587 | −0.140 | 0.189 | 0.060 | 0.089 |
| Curvature Degree | 0.105 | 0.038 | −0.016 | 0.013 | −0.011 | 0.024 | 0.002 | 0.009 | 0.000 | 0.005 |
| Fit Statistics | | | | | | | | | | |
| AIC | 370.24 | N/A | 6386.8 | N/A | 3011.2 | N/A | 7191.9 | N/A | 17318 | N/A |
| BIC | 439.225 | N/A | 3843.97 | N/A | 1686.52 | N/A | 4457.75 | N/A | 9481.62 | N/A |
| Overdispersion | 0.0139 | — | 1.182 | — | 4.329 | — | `0.567 | — | 0.375 | — |

—No data.

*Variables are insignificant at a 90-percent confidence level.

**Table 11. Random parameters—Poisson regression estimates for dataset 10 (1,000 mi).**

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| Intercept | −17.604* | 10.500 | −6.423 | 0.966 | −5.983 | 2.342 | −3.449 | 0.777 | −2.262 | 0.385 |
| Ln (Segment Length) | 0.855 | 0.617 | 0.988 | 0.049 | 1.111 | 0.123 | 1.040 | 0.040 | 1.053 | 0.020 |
| Ln (AADT) | 0.862 | 1.023 | 0.830 | 0.067 | 0.251 | 0.144 | 0.438 | 0.051 | 0.468 | 0.026 |
| Shoulder Width | | | | | | | | | | |
| 8 ft | Base Level | | | | | | | | | |
| <2 ft | 0.265 | 2.649 | 0.204 | 0.224 | −0.460 | 0.810 | 0.805 | 0.168 | 0.241 | 0.092 |
| ≥2 ft < 4 ft | −1.248 | 2.812 | 0.497 | 0.120 | 1.234 | 0.310 | 0.530 | 0.105 | 0.305 | 0.049 |
| ≥4 ft < 6 ft | 0.685 | 1.488 | 0.172 | 0.124 | 0.522 | 0.325 | 0.552 | 0.105 | 0.272 | 0.050 |
| ≥6 ft < 8 ft | 0.059 | 1.608 | 0.068 | 0.145 | 0.066 | 0.399 | 0.355 | 0.124 | 0.253 | 0.057 |
| Lane Width | | | | | | | | | | |
| >12 ft | Base Level | | | | | | | | | |
| ≤9 ft | 0.265 | 2.649 | 0.204 | 0.224 | −0.460 | 0.810 | 0.805 | 0.168 | 0.241 | 0.092 |
| 9.5 ft−10.5 ft | −1.248 | 2.812 | 0.497 | 0.120 | 1.234 | 0.310 | 0.530 | 0.105 | 0.305 | 0.049 |
| 11 ft−11.5 ft | 0.685 | 1.488 | 0.172 | 0.124 | 0.522 | 0.325 | 0.552 | 0.105 | 0.272 | 0.050 |
| Speed Limit | | | | | | | | | | |
| 45 mph | Base Level | | | | | | | | | |
| 25 mph | 1.650* | 4.138 | 0.296 | 0.333 | −0.412 | 0.828 | 0.541 | 0.258 | 0.017 | 0.128 |
| 30 mph | 1.833* | 4.425 | 0.517 | 0.303 | 0.646 | 0.689 | 0.770 | 0.244 | 0.027 | 0.132 |
| 35 mph | 3.153 | 3.556 | −0.071 | 0.351 | −0.180 | 0.823 | 0.596 | 0.251 | −0.377 | 0.142 |
| 40 mph | −3.141* | 20.861 | −0.855 | 0.280 | −1.315 | 0.629 | −0.557 | 0.234 | −1.106 | 0.117 |
| 50 mph | −30.351* | 609.50 | −0.391 | 0.332 | −1.131 | 0.838 | 0.222 | 0.262 | −0.009 | 0.116 |
| 55 mph | −0.242 | 2.618 | −0.342 | 0.168 | −1.005 | 0.398 | −0.123 | 0.154 | −0.339 | 0.068 |
| Roadside Hazard Rating (Zegeer et al 1988) | | | | | | | | | | |
| 3 | Base Level | | | | | | | | | |
| 4 | 2.522* | 1.945 | 0.053 | 0.122 | −0.417 | 0.307 | −0.090 | 0.101 | 0.030 | 0.048 |
| 5 | 1.539* | 2.566 | 0.062 | 0.156 | −0.401 | 0.397 | 0.094 | 0.128 | 0.049 | 0.063 |
| 6 | 1.953* | 2.526 | −0.014 | 0.193 | −0.023 | 0.445 | 0.211 | 0.156 | 0.278 | 0.077 |
| 7 | 2.678* | 2.636 | −0.113 | 0.267 | −0.223 | 0.647 | 0.381 | 0.211 | 0.206 | 0.107 |
| Pavement Condition | 0.037 | 0.165 | −0.009 | 0.018 | 0.014 | 0.047 | −0.003 | 0.015 | 0.016 | 0.008 |
| Pavement Roughness | 0.006 | 0.021 | 0.003 | 0.002 | 0.005 | 0.005 | 0.001 | 0.002 | 0.000 | 0.001 |
| Presence of Lighting | 0.128 | 2.363 | −0.420 | 0.244 | −0.330 | 0.553 | −0.754 | 0.217 | −0.534 | 0.100 |

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** |
| Presence of Horizontal Curve | −27.666* | 2949.6 | −2.918 | 1.050 | −1.583 | 1.350 | −1.479 | 0.470 | 0.175 | 0.111 |
| Vertical Curve Approaching Slope | 0.286 | 2.322 | 0.505 | 0.236 | −0.165 | 0.593 | 0.133 | 0.159 | −0.006 | 0.075 |
| Vertical Curve Leaving Slope | 2.126 | 5.974 | 0.145 | 0.152 | −0.340 | 0.483 | 0.089 | 0.138 | −0.033 | 0.064 |
| Grade | −2.241 | 6.215 | −0.547 | 0.267 | 0.275 | 0.665 | −0.159 | 0.193 | 0.038 | 0.089 |
| Curvature Degree | −0.007 | 0.024 | 0.006 | 0.003 | 0.002 | 0.005 | 0.000 | 0.003 | 0.002 | 0.001 |
| SD (Curvature Degree) | 1.027 | 0.052 | 1.097 | 0.022 | 1.2011 | 0.107 | 0.838 | 0.152 | 0.584 | 0.029 |
| Fit Statistics | | | | | | | | | | |
| AIC | 443.88 | N/A | 5693.6 | N/A | 3018.8 | N/A | 6641.1 | N/A | 15812 | N/A |
| BIC | 635.97 | N/A | 5885.67 | N/A | 3210.9 | N/A | 6833.24 | N/A | 16004.4 | N/A |

—No data.
*Variables are insignificant at a 90-percent confidence level.

**Table 12. Random parameters—NB regression estimates for dataset 10 (1,000 mi).**

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| Intercept | −25.066 | 4.175 | 2.522 | 0.678 | −3.181 | 1.004 | −2.594 | 0.560 | −1.875 | 0.250 |
| Ln (Segment Length) | 1.323 | 0.178 | 1.539 | 0.032 | 1.318 | 0.051 | 1.000 | 0.028 | 1.0946 | 0.013 |
| Ln (AADT) | 0.367 | 0.303 | 1.953 | 0.050 | 0.133 | 0.075 | 0.469 | 0.040 | 0.090 | 0.018 |
| Standard Deviation (Length) | 0.263 | 0.184 | 2.678 | 2.968 | 0.197 | 1.560 | 0.432 | 0.371 | 0.520 | 0.149 |
| Standard Deviation (AADT) | 0.012 | 0.219 | 0.037 | 1.149 | 0.134 | 0.793 | 0.021 | 0.387 | 0.116 | 0.324 |
| Shoulder Width | | | | | | | | | | |
| 8 ft | Base Level | | | | | | | | | |
| <2 ft | −0.580* | 1.059 | 0.317 | 0.140 | 0.656 | 0.214 | 0.324 | 0.129 | 0.231 | 0.062 |
| ≥2 ft < 4 ft | −16.357* | 1672.361 | −0.364 | 0.180 | −0.817 | 0.589 | −0.020 | 0.227 | 0.349 | 0.087 |
| ≥4 ft < 6 ft | −0.080* | 0.397 | 0.260 | 0.076 | 0.369 | 0.126 | 0.359 | 0.066 | 0.340 | 0.030 |
| ≥6 ft < 8 ft | −0.874* | 0.583 | 0.174 | 0.088 | 0.395 | 0.140 | 0.003 | 0.079 | 0.263 | 0.034 |
| Lane Width | | | | | | | | | | |
| >12 ft | Base Level | | | | | | | | | |
| ≤9 ft | 3.028 | 0.567 | 0.279 | 0.160 | 0.011 | 0.249 | 0.246 | 0.142 | 0.145 | 0.064 |
| 9.5 ft−10.5 ft | 0.411 | 0.593 | 0.129 | 0.091 | −0.376 | 0.149 | 0.179 | 0.076 | 0.188 | 0.034 |
| 11 ft−11.5 ft | −12.372* | 1520.181 | 0.478 | 0.229 | −0.341 | 0.393 | 0.343 | 0.194 | 0.223 | 0.089 |
| Speed Limit | | | | | | | | | | |
| 45 mph | Base Level | | | | | | | | | |
| 25 mph | 2.826* | 1.088 | 0.791 | 0.181 | 1.089 | 0.323 | 0.365 | 0.157 | 0.483 | 0.068 |
| 30 mph | −15.228* | 1864.827 | 0.387 | 0.199 | 1.239 | 0.303 | 0.560 | 0.150 | 0.081 | 0.078 |
| 35 mph | 0.773* | 1.250 | 0.642 | 0.184 | 0.696 | 0.329 | 0.038 | 0.167 | −0.359 | 0.088 |
| 40 mph | 1.575* | 1.096 | −0.676 | 0.214 | −0.479 | 0.374 | −0.677 | 0.163 | −1.044 | 0.084 |
| 50 mph | −14.297* | 1595.923 | 0.321 | 0.217 | −0.256 | 0.505 | −0.356 | 0.219 | −0.151 | 0.092 |
| 55 mph | −0.123* | 1.072 | −0.217 | 0.136 | 0.246 | 0.249 | −0.356 | 0.106 | −0.258 | 0.049 |
| Roadside Hazard Rating (Zegeer et al 1988) | | | | | | | | | | |
| 3 | Base Level | | | | | | | | | |
| 4 | −1.690* | 0.547 | 0.045 | 0.084 | 0.030 | 0.138 | 0.146 | 0.072 | 0.069 | 0.032 |
| 5 | −1.395 | 0.678 | 0.115 | 0.103 | 0.279 | 0.164 | 0.233 | 0.089 | 0.180 | 0.040 |
| 6 | 0.054 | 0.689 | −0.025 | 0.134 | 0.430 | 0.204 | 0.161 | 0.116 | 0.159 | 0.052 |
| 7 | −0.715 | 1.066 | 0.237 | 0.171 | 0.558 | 0.279 | 0.110 | 0.159 | 0.251 | 0.068 |

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| Pavement Condition | 0.249 | 0.068 | 0.021 | 0.012 | −0.019 | 0.018 | 0.008 | 0.011 | −0.002 | 0.005 |
| Pavement Roughness | 0.014 | 0.208 | 0.002 | 0.371 | −0.297 | 0.284 | 0.387 | 0.248 | 0.452 | 0.732 |
| Presence of Lighting | −1.631 | 1.220 | −0.180 | 0.164 | −0.491 | 0.288 | −0.237 | 0.152 | −0.417 | 0.070 |
| Presence of Horizontal Curve | −14.516* | 1042.264 | −1.408 | 0.411 | −14.329 | 262.136 | −1.225 | 0.336 | 0.345 | 0.073 |
| Vertical Curve Approaching Slope | −3.043 | 5.011 | 0.173 | 0.157 | −0.359 | 0.289 | 0.158 | 0.106 | −0.053 | 0.053 |
| Vertical Curve Leaving Slope | −2.276 | 5.020 | −0.104 | 0.114 | 0.163 | 0.324 | 0.332 | 0.107 | 0.077 | 0.051 |
| Grade | 2.775 | 5.118 | −0.058 | 0.175 | −0.061 | 0.374 | −0.336 | 0.140 | −0.005 | 0.065 |
| Curvature Degree | −0.007 | 0.024 | 0.006 | 0.003 | 0.002 | 0.005 | 0.000 | 0.003 | 0.002 | 0.001 |
| Fit Statistics | | | | | | | | | | |
| AIC | 443.88 | N/A | 5693.6 | N/A | 3018.8 | N/A | 6641.1 | N/A | 15812 | N/A |
| BIC | 635.97 | N/A | 5885.67 | N/A | 3210.9 | N/A | 6833.24 | N/A | 16004.4 | N/A |

—No data.

*Variables are insignificant at a 90-percent confidence level.

**Table 13. Poisson univariate estimates for dataset 10 (1,000 mi).**

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| Intercept | −25.066* | 4.175 | −7.434 | 0.678 | −4.478 | 1.004 | −3.609 | 0.560 | −2.099 | 0.250 |
| Ln (Segment Length) | 1.292 | 0.178 | 0.984 | 0.032 | 1.023 | 0.051 | 1.000 | 0.028 | 1.059 | 0.013 |
| Ln (AADT) | 1.389 | 0.303 | 0.813 | 0.050 | 0.533 | 0.075 | 0.469 | 0.040 | 0.536 | 0.018 |
| Shoulder Width | | | | | | | | | | |
| 8 ft | Base Level | | | | | | | | | |
| <2 ft | −0.580* | 1.059 | 0.317 | 0.140 | 0.656 | 0.214 | 0.324 | 0.129 | 0.231 | 0.062 |
| ≥2 ft < 4 ft | −16.357* | 1672.361 | 0.736 | 0.180 | −0.817 | 0.589 | −0.020 | 0.227 | 0.349 | 0.087 |
| ≥4 ft < 6 ft | −0.080 | 0.397 | 0.260 | 0.076 | 0.369 | 0.126 | 0.359 | 0.066 | 0.340 | 0.030 |
| ≥6 ft < 8 ft | −0.874 | 0.583 | 0.174 | 0.088 | 0.395 | 0.140 | 0.003 | 0.079 | 0.263 | 0.034 |
| Lane Width | | | | | | | | | | |
| >12 ft | Base Level | | | | | | | | | |
| ≤9 ft | 3.028* | 0.567 | 0.279 | 0.160 | 0.011 | 0.249 | 0.246 | 0.142 | 0.145 | 0.064 |
| 9.5 ft−10.5 ft | 0.411 | 0.593 | 0.129 | 0.091 | −0.376 | 0.149 | 0.179 | 0.076 | 0.188 | 0.034 |
| 11 ft−11.5 ft | −12.372* | 1520.181 | 0.478 | 0.229 | −0.341 | 0.393 | 0.343 | 0.194 | 0.223 | 0.089 |
| Speed Limit | | | | | | | | | | |
| 45 mph | Base Level | | | | | | | | | |
| 25 mph | 2.826* | 1.088 | 0.791 | 0.181 | 1.089 | 0.323 | 0.365 | 0.157 | 0.483 | 0.068 |
| 30 mph | −15.228* | 1864.827 | 0.387 | 0.199 | 1.239 | 0.303 | 0.560 | 0.150 | 0.081 | 0.078 |
| 35 mph | 0.773 | 1.250 | 0.642 | 0.184 | 0.696 | 0.329 | 0.038 | 0.167 | −0.359 | 0.088 |
| 40 mph | 1.575 | 1.096 | −0.676 | 0.214 | −0.479 | 0.374 | −0.677 | 0.163 | −1.044 | 0.084 |
| 50 mph | −14.297* | 1595.923 | 0.321 | 0.217 | −0.256 | 0.505 | −0.356 | 0.219 | −0.151 | 0.092 |
| 55 mph | −0.123 | 1.072 | −0.217 | 0.136 | 0.246 | 0.249 | −0.356 | 0.106 | −0.258 | 0.049 |
| Roadside Hazard Rating (Zegeer et al 1988) | | | | | | | | | | |
| 3 | Base Level | | | | | | | | | |
| 4 | −1.069 | 0.547 | 0.045 | 0.084 | 0.030 | 0.138 | 0.146 | 0.072 | 0.069 | 0.032 |
| 5 | −1.395 | 0.678 | 0.115 | 0.103 | 0.279 | 0.164 | 0.233 | 0.089 | 0.180 | 0.040 |
| 6 | 0.054 | 0.689 | −0.025 | 0.134 | 0.430 | 0.204 | 0.161 | 0.116 | 0.159 | 0.052 |
| 7 | −0.715 | 1.066 | 0.237 | 0.171 | 0.558 | 0.279 | 0.110 | 0.159 | 0.251 | 0.068 |
| Pavement Condition | 0.249 | 0.068 | 0.021 | 0.012 | −0.019 | 0.018 | 0.008 | 0.011 | −0.002 | 0.005 |
| Pavement Roughness | 0.014 | 0.008 | 0.002 | 0.001 | −0.002 | 0.002 | 0.001 | 0.001 | 0.000 | 0.001 |
| Presence of Lighting | −1.631 | 1.220 | −0.180 | 0.164 | −0.491 | 0.288 | −0.237 | 0.152 | −0.417 | 0.070 |
| Presence of Horizontal Curve | −14.516* | 1042.264 | −1.408 | 0.411 | −14.329 | 262.136 | −1.225 | 0.336 | 0.345 | 0.073 |

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** |
| Vertical Curve Approaching Slope | −3.043* | 5.011 | 0.173 | 0.157 | −0.359 | 0.289 | 0.158 | 0.106 | −0.053 | 0.053 |
| Vertical Curve Leaving Slope | −2.276* | 5.020 | −0.104 | 0.114 | 0.163 | 0.324 | 0.332 | 0.107 | 0.077 | 0.051 |
| Grade | 2.775* | 5.118 | −0.058 | 0.175 | −0.061 | 0.374 | −0.336 | 0.140 | −0.005 | 0.065 |
| Curvature Degree | −0.007 | 0.024 | 0.006 | 0.003 | 0.002 | 0.005 | 0.000 | 0.003 | 0.002 | 0.001 |
| Fit Statistics | | | | | | | | | | |
| AIC | 443.88 | N/A | 5693.6 | N/A | 3018.8 | N/A | 6641.1 | N/A | 15812 | N/A |
| BIC | 635.97 | N/A | 5885.67 | N/A | 3210.9 | N/A | 6833.24 | N/A | 16004.4 | N/A |

—No data.

*Variables are insignificant at a 90-percent confidence level.

**Table 14. Multivariate Poisson lognormal estimates for dataset 10 (1,000 mi).**

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| Intercept | −25.066* | 4.175 | −7.434 | 0.678 | −4.478 | 1.004 | −3.609 | 0.560 | −2.099 | 0.250 |
| Ln (Segment Length) | 1.292 | 0.178 | 0.984 | 0.032 | 1.023 | 0.051 | 1.000 | 0.028 | 1.059 | 0.013 |
| Ln (AADT) | 1.389 | 0.303 | 0.813 | 0.050 | 0.533 | 0.075 | 0.469 | 0.040 | 0.536 | 0.018 |
| Shoulder Width | | | | | | | | | | |
| 8 ft | Base Level | | | | | | | | | |
| <2 ft | −0.580 | 1.059 | 0.317 | 0.140 | 0.656 | 0.214 | 0.324 | 0.129 | 0.231 | 0.062 |
| ≥2 ft < 4 ft | −16.357* | 1672.361 | 0.736 | 0.180 | −0.817 | 0.589 | −0.020 | 0.227 | 0.349 | 0.087 |
| ≥4 ft < 6 ft | −0.080 | 0.397 | 0.260 | 0.076 | 0.369 | 0.126 | 0.359 | 0.066 | 0.340 | 0.030 |
| ≥6 ft < 8 ft | −0.874 | 0.583 | 0.174 | 0.088 | 0.395 | 0.140 | 0.003 | 0.079 | 0.263 | 0.034 |
| Lane Width | | | | | | | | | | |
| >12 ft | Base Level | | | | | | | | | |
| ≤9 ft | 3.028* | 0.567 | 0.279 | 0.160 | 0.011 | 0.249 | 0.246 | 0.142 | 0.145 | 0.064 |
| 9.5 ft−10.5 ft | 0.411 | 0.593 | 0.129 | 0.091 | −0.376 | 0.149 | 0.179 | 0.076 | 0.188 | 0.034 |
| 11 ft−11.5 ft | −12.372* | 1520.181 | 0.478 | 0.229 | −0.341 | 0.393 | 0.343 | 0.194 | 0.223 | 0.089 |
| Speed Limit | | | | | | | | | | |
| 45 mph | Base Level | | | | | | | | | |
| 25 mph | 2.826 | 1.088 | 0.791 | 0.181 | 1.089 | 0.323 | 0.365 | 0.157 | 0.483 | 0.068 |
| 30 mph | −15.228* | 1864.827 | 0.387 | 0.199 | 1.239 | 0.303 | 0.560 | 0.150 | 0.081 | 0.078 |
| 35 mph | 0.773 | 1.250 | 0.642 | 0.184 | 0.696 | 0.329 | 0.038 | 0.167 | −0.359 | 0.088 |
| 40 mph | 1.575 | 1.096 | −0.676 | 0.214 | −0.479 | 0.374 | −0.677 | 0.163 | −1.044 | 0.084 |
| 50 mph | −14.297* | 1595.923 | 0.321 | 0.217 | −0.256 | 0.505 | −0.356 | 0.219 | −0.151 | 0.092 |
| 55 mph | −0.123 | 1.072 | −0.217 | 0.136 | 0.246 | 0.249 | −0.356 | 0.106 | −0.258 | 0.049 |
| Roadside Hazard Rating (Zegeer et al 1988) | | | | | | | | | | |
| 3 | Base Level | | | | | | | | | |
| 4 | −1.069 | 0.547 | 0.045 | 0.084 | 0.030 | 0.138 | 0.146 | 0.072 | 0.069 | 0.032 |
| 5 | −1.395 | 0.678 | 0.115 | 0.103 | 0.279 | 0.164 | 0.233 | 0.089 | 0.180 | 0.040 |
| 6 | 0.054 | 0.689 | −0.025 | 0.134 | 0.430 | 0.204 | 0.161 | 0.116 | 0.159 | 0.052 |
| 7 | −0.715 | 1.066 | 0.237 | 0.171 | 0.558 | 0.279 | 0.110 | 0.159 | 0.251 | 0.068 |
| Pavement Condition | 0.249 | 0.068 | 0.021 | 0.012 | −0.019 | 0.018 | 0.008 | 0.011 | −0.002 | 0.005 |
| Pavement Roughness | 0.014 | 0.008 | 0.002 | 0.001 | −0.002 | 0.002 | 0.001 | 0.001 | 0.000 | 0.001 |
| Presence of Lighting | −1.631 | 1.220 | −0.180 | 0.164 | −0.491 | 0.288 | −0.237 | 0.152 | −0.417 | 0.070 |
| Presence of Horizontal Curve | −14.516* | 1042.264 | −1.408 | 0.411 | −14.329 | 262.136 | −1.225 | 0.336 | 0.345 | 0.073 |

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** |
| Vertical Curve Approaching Slope | −3.043* | 5.011 | 0.173 | 0.157 | −0.359 | 0.289 | 0.158 | 0.106 | −0.053 | 0.053 |
| Vertical Curve Leaving Slope | −2.276* | 5.020 | −0.104 | 0.114 | 0.163 | 0.324 | 0.332 | 0.107 | 0.077 | 0.051 |
| Grade | 2.775* | 5.118 | −0.058 | 0.175 | −0.061 | 0.374 | −0.336 | 0.140 | −0.005 | 0.065 |
| Curvature Degree | −0.007 | 0.024 | 0.006 | 0.003 | 0.002 | 0.005 | 0.000 | 0.003 | 0.002 | 0.001 |
| Variance-Covariance Matrix | | | | | | | | | | |
| K | 0.218 | | 0.625 | | 0.248 | | 0.138 | | 0.086 | |
| A | — | | 1.011 | | 0.526 | | 0.491 | | 0.516 | |
| B | — | | — | | 1.600 | | 0.733 | | 0.715 | |
| C | — | | — | | — | | 0.501 | | 0.477 | |
| PDO | — | | — | | — | | — | | 0.531 | |

—No data.
*Variables are insignificant at a 90-percent confidence level.

### *Crash Validation*

The model prediction of dataset 10 used for model estimation is shown in table 15. Dataset 5, which had the same mileage, was used for cross-validation.

**Table 15. Model prediction using dataset 10 (1,000 mi) for estimation.**

| Models | Metric | K | A | B | C | PDO |
|---|---|---|---|---|---|---|
| Poisson regression | MAD | 11.15 | 2.74 | 3.71 | 2.48 | 1.73 |
| | MSPE | 202.4 | 9.25 | 15.5 | 7.49 | 5.17 |
| NB regression | MAD | 14.92 | 2.74 | 3.69 | 2.49 | 1.73 |
| | MSPE | 451.31 | 9.23 | 15.36 | 7.50 | 5.18 |
| Poisson regression—random parameters | MAD | 36.91 | 3.26 | 5.30 | 2.80 | 1.78 |
| | MSPE | 450.7 | 12.29 | 30.74 | 9.00 | 7.57 |
| NB—random parameters | MAD | 24.5 | 3.16 | 6.10 | 2.61 | 1.73 |
| | MSPE | 331.3 | 11.5 | 40.04 | 8.03 | 5.20 |
| Univariate Poisson lognormal | MAD | 29.1 | 2.83 | 4.38 | 2.53 | 1.88 |
| | MSPE | 135.32 | 9.81 | 31.5 | 8.36 | 6.28 |
| Multivariate Poisson lognormal | MAD | 26.92 | 3.72 | 3.98 | 3.13 | 1.34 |
| | MSPE | 472.31 | 10.50 | 15.57 | 7.90 | 5.70 |

From the results, Poisson regression had a lower MSPE and MAD for K crashes. Poisson regression had the lowest MAD and NB had the lowest MSPE for A crashes, which means that it performed better. Univariate Poisson lognormal had the lowest MAD for B crashes; meanwhile, Poisson regression had the lowest MSPE value. Poisson regression performed better for C crashes for MAD while NB did better for MSPE. Finally, Poisson did better for MAD for PDO crashes while univariate Poisson lognormal did better for MSPE.

The cross-validation results are shown in table 16, table 17, and table 18.

**Table 16. Cross validation with dataset 5 (1,000 mi) with parameters estimated using dataset 10.**

| Models | Metric | K | A | B | C | PDO |
|---|---|---|---|---|---|---|
| Poisson regression | MAD | 11.41 | 2.74 | 3.73 | 2.52 | 1.85 |
| | MSPE | 210.24 | 9.34 | 15.82 | 7.73 | 6.21 |
| NB regression | MAD | 15.20 | 2.73 | 3.74 | 2.53 | 1.85 |
| | MSPE | 466.20 | 9.29 | 15.83 | 7.75 | 6.24 |
| Poisson regression—random parameters | MAD | 39.01 | 3.27 | 5.36 | 2.83 | 1.99 |
| | MSPE | 593.8 | 12.70 | 31.36 | 9.48 | 6.86 |
| NB—random parameters | MAD | 25.20 | 3.15 | 6.21 | 2.64 | 1.86 |
| | MSPE | 379.1 | 11.8 | 41.48 | 8.36 | 6.26 |
| Univariate Poisson lognormal | MAD | 14.92 | 2.74 | 3.69 | 2.49 | 1.73 |
| | MSPE | 351.31 | 9.23 | 15.36 | 7.50 | 5.18 |
| Multivariate Poisson lognormal | MAD | 15.92 | 2.92 | 4.68 | 3.51 | 2.73 |
| | MSPE | 462.31 | 10.25 | 15.56 | 7.54 | 5.27 |

**Table 17. Model prediction with dataset 5 (1,000 mi) used for estimation.**

| Models | Metric | K | A | B | C | PDO |
|---|---|---|---|---|---|---|
| Poisson regression | MAD | 9.81 | 2.83 | 4.38 | 2.53 | 1.88 |
| | MSPE | 135.32 | 9.81 | 31.5 | 8.36 | 6.28 |
| NB regression | MAD | 14.03 | 2.86 | 5.05 | 2.51 | 1.88 |
| | MSPE | 533.60 | 10.04 | 66.41 | 7.74 | 6.30 |
| Poisson regression—random parameters | MAD | 24.01 | 3.26 | 5.30 | 2.79 | 1.88 |
| | MSPE | 233.2 | 12.29 | 30.74 | 8.98 | 5.182 |
| NB—random parameters | MAD | 24.59 | 3.14 | 6.11 | 2.60 | 1.739 |
| | MSPE | 301.3 | 11.4 | 40.08 | 8.02 | 5.203 |
| Univariate Poisson lognormal | MAD | 10.1 | 2.75 | 2.52 | 2.01 | 1.12 |
| | MSPE | 254.7 | 8.32 | 12.30 | 6.90 | 10.04 |
| Multivariate Poisson lognormal | MAD | 14.38 | 3.61 | 2.67 | 2.91 | 2.01 |
| | MSPE | 441.5 | 9.19 | 12.99 | 6.98 | 10.24 |

**Table 18. Cross validation with dataset 10 (1,000 mi) with parameters estimated using dataset 5.**

| Models | Metric | K | A | B | C | PDO |
|---|---|---|---|---|---|---|
| Poisson regression | MAD | 13.54 | 2.74 | 4.64 | 2.59 | 1.77 |
| | MSPE | 263.07 | 9.274 | 33.51 | 8.14 | 5.34 |
| NB regression | MAD | 25.19 | 2.80 | 5.25 | 2.58 | 1.78 |
| | MSPE | 335.4 | 9.65 | 66.39 | 8.060 | 5.401 |
| Poisson regression—random parameters | MAD | 27.63 | 3.41 | 5.31 | 2.78 | 2.01 |
| | MSPE | 2671.7 | 13.12 | 31.13 | 8.96 | 6.28 |
| NB—random parameters | MAD | 22.86 | 3.31 | 9.05 | 2.55 | 1.88 |
| | MSPE | 303.1 | 12.59 | 310.34 | 7.84 | 6.31 |
| Univariate Poisson lognormal | MAD | 10.2 | 2.77 | 2.98 | 2.70 | 1.12 |
| | MSPE | 254.7 | 8.24 | 13.67 | 6.79 | 10.14 |
| Multivariate Poisson lognormal | MAD | 15.38 | 2.36 | 2.77 | 2.93 | 2.31 |
| | MSPE | 267.5 | 9.19 | 12.89 | 6.89 | 10.24 |

Cross validation was used to assess the predictive performance of the models and to judge how they perform outside the sample to a new dataset also known as test data. Without cross validation, there is only information on how each model performs relative to the in-sample data. The range of numbers from the cross-validation table shows that the models are robust enough to be relied on which suggests that the data from the tool can be trusted to produce convincing result each time various models are run on them.

### *Empirical Analysis for Evaluating Parameter Stability*

Parameter estimates were examined to determine consistency using the revised Wald statistics. The analysis was conducted for all approaches, but only the results for the NB approach with PDO crashes are shown here, as an example. Dataset 10 (1,000 mi) was used as the population benchmark to evaluate if each parameter in any model is statistically different from the corresponding one in that dataset. If the parameter test statistic is greater than the 90 percent t-statistic, this result indicates a significant difference between the datasets.

In contrast, if the parameter test statistics are below the 90-percent t-statistic, no significant difference exists between the datasets, and the stochasticity in the RAD tool can be assumed to be consistent across different generations with different random seeds. The research team used dataset 10 because it had the largest number of observations and was believed to produce the most convincing parameter estimates. The combined NB estimates of PDO crashes for the 10 datasets are displayed in table 19, along with their standard errors for illustrative purposes. The updated Wald test statistics for the estimated model parameter values are shown in table 20. For example, the test statistics for segment length across the datasets were lower than the 90-percent confidence value of 1.65, which could mean that there is no significant difference between the estimated parameters in the corresponding dataset and dataset 10.

## Application To Urban Four-Leg Signalized Intersections

### *RAD Generation*

The descriptive statistics of the datasets generated by the tool are shown in table 21. The data in table 22 contains traffic, geometric, and intersection feature characteristics. The values for each of the continuous variables fall within a reasonable range, which shows that the dataset generated from the tool, even though randomly generated, will be within expected ranges. The same outcome was noticed for the categorical variables.

**Table 19. NB estimates for PDO crashes.**

| Parameter | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Dataset 5 | Dataset 6 | Dataset 7 | Dataset 8 | Dataset 9 | Dataset 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | −0.62[a] (0.860[b] | −3.870, (0.685) | −1.856, (0.369) | −2.001, (0.389) | −1.965, 0.291 | −1.000*, (1.176) | −3.278, (0.671) | −1.917, (0.448) | −2.438, (0.385) | −2.223, (0.321) |
| Ln (Segment Length) | 0.963, (0.036) | 1.054, (0.034) | 1.023, (0.021) | 1.058, (0.019) | 1.049, (0.017) | 0.965, (0.050) | 1.050, (0.031) | 1.074, (0.025) | 1.048, (0.021) | 1.064, (0.017) |
| Ln (AADT) | 0.392, (0.065) | 0.627, (0.045) | 0.548, (0.027) | 0.452, (0.025) | 0.474, (0.023) | 0.415, (0.089) | 0.533, (0.044) | 0.464, (0.032) | 0.530, (0.043) | 0.546, (0.088) |
| Shoulder Width | | | | | | | | | | |
| 8-ft | Base Level | | | | | | | | | |
| <2 ft | 0.167, (0.329) | 0.610, (0.155) | 0.381, (0.076) | 0.240, (0.093) | 0.349, (0.071) | 0.818, (0.365) | 0.334, (0.159) | 0.266, (0.093) | 0.189, (0.095) | 0.246, (0.076) |
| ≥2 ft < 4 ft | 0.361, (0.092) | 0.497, (0.077) | 0.360, (0.051) | 0.292, (0.050) | 0.383, (0.043) | 0.781, (0.373) | 0.471, (0.076) | 0.249, (0.061) | 0.341, (0.051) | 0.329, (0.118) |
| ≥4 ft < 6 ft | 0.144, (0.083) | 0.446, (0.083) | 0.277, (0.048) | 0.257, (0.051) | 0.243, (0.042) | 0.361, (0.095) | 0.410, (0.083) | 0.218, (0.058) | 0.307, (0.051) | 0.351, (0.094) |
| ≥6 ft < 8 ft | 0.168, (0.096) | 0.479, (0.089) | 0.137, (0.058) | 0.246, (0.059) | 0.217, (0.048) | 0.117*, (0.136) | 0.332, (0.088) | 0.126, (0.070) | 0.177, (0.060) | 0.287, (0.044) |
| Lane Width | | | | | | | | | | |
| 12-ft | Base Level | | | | | | | | | |
| ≤9 ft | — | 0.932, (1.071) | 0.326, (0.196) | 0.5047, (0.129) | 0.415, (0.112) | — | 0.599*, (1.088) | 0.566, (0.232) | 0.403, (0.130) | 0.235, (0.117) |
| 9.5 ft−10.5 ft | 0.210*, (0.208) | 0.252, (0.143) | 0.125, (0.089) | 0.256, (0.094) | 0.289, (0.081) | −0.151, (0.285) | 0.255, (0.142) | 0.184, (0.111) | 0.268, (0.093) | 0.152, (0.084) |
| 11 ft−11.5 ft | 0.008* (0.101) | 0.116, (0.086) | 0.183, (0.051) | 0.091, (0.052) | 0.141, (0.043) | 0.047*, (0.142) | 0.374, (0.083) | 0.102*, (0.064) | 0.152, (0.051) | 0.210, (0.043) |
| Speed Limit | | | | | | | | | | |
| 45-mph | Base Level | | | | | | | | | |
| 25 mph | 0.475, (0.062) | 0.140, (0.188) | 0.439, (0.116) | 0.041, (0.030) | 0.606, (0.102) | 0.291, (0.125) | 0.429, (0.174) | 0.519*, (0.140) | 0.244, (0.139) | 0.483, (0.098) |
| 30 mph | 0.171*, (0.242) | 0.428, (0.244) | 0.210, (0.136) | 0.041, (0.035) | 0.349, (0.110) | 0.313, (0.173) | 0.385*, (0.241) | −0.05*, (0.184) | 0.234, (0.139) | 0.035*, (0.111) |
| 35 mph | −0.84*, (0.271) | −0.701, (0.306) | −0.271, (0.174) | −0.295, (0.143) | 0.069*, (0.109) | 0.268*, (0.343) | −0.045*, (0.253) | −0.567, 0.237 | −0.222, (0.150) | −0.336, (0.112) |
| 40 mph | −0.956, (0.252) | −0.824, (0.197) | −0.822, (0.114) | −1.109, (0.119) | −0.655, (0.099) | −0.879, (0.319) | −0.992, (0.216) | −0.91*, 0.144, | −0.969, (0.123) | −1.021, (0.103) |
| 50 mph | 0.120*, (0.233) | 0.095, (0.172) | −0.163, (0.122) | 0.003*, (0.119) | −0.087, (0.124) | −0.425, (0.367) | −0.036, (0.178) | 0.117*, 0.139 | −0.057, (0.130) | −0.115, (0.116) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 55 mph | −0.28*, (0.132) | −0.332, (0.118) | −0.254, (0.072) | −0.329, (0.071) | −0.011, (0.067) | −0.196, (0.180) | −0.066, (0.115) | −0.226, (0.089) | −0.125, (0.074) | −0.270, (0.064) |
| Presence of Lighting | −0.341, (0.184) | −0.603, (0.151) | −0.523, (0.151) | −0.531, (0.101) | −0.412, (0.085) | −0.300, (0.259) | −0.334, (0.026) | −0.658, (0.122) | −0.306, (0.098) | −0.431, (0.085) |
| Presence of Horizontal Curve | 0.549*, (0.239) | 0.294* (0.205) | 0.078* (0.116) | 0.186 (0.112) | 0.264 (0.088) | 0.269* (0.244) | 0.407 (0.036) | 0.652 (0.111) | 0.168* (0.112) | 0.365 (0.085) |
| Overdispersion | 0.257 | 0.463 | 0.344 | 0.199 | 0.375 | 0.725 | 0.271 | 0.447 | 0.216 | 0.307 |

—No data.
*Variables insignificant at a 90-percent confidence level.
[a]Parameter estimate.
[b]Standard error.

**Table 20. Revised Wald test statistics on NB model parameter estimates (relative to dataset 10).**

| Parameter | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Dataset 5 | Dataset 6 | Dataset 7 | Dataset 8 | Dataset 9 |
|---|---|---|---|---|---|---|---|---|---|
| Ln (Segment Length) | 2.536 | 0.263 | 1.517 | 0.235 | 0.623 | 1.874 | 0.395 | 0.330 | 0.592 |
| Ln (AADT) | 1.407 | 0.819 | 0.022 | 1.027 | 0.792 | 1.046 | 0.132 | 0.875 | 0.163 |
| Shoulder Width | | | | | | | | | |
| 0 ft | 0.233 | 2.108 | 1.256 | 0.049 | 0.990 | 1.534 | 0.499 | 0.166 | 0.468 |
| 2 ft | 0.213 | 1.192 | 0.241 | 0.288 | 0.429 | 1.155 | 1.011 | 0.602 | 0.093 |
| 4 ft | 1.650 | 0.757 | 0.701 | 0.878 | 1.048 | 0.074 | 0.470 | 1.204 | 0.411 |
| 6 ft | 1.126 | 1.933 | 2.060 | 0.557 | 1.075 | 1.189 | 0.457 | 1.947 | 1.478 |
| Lane Width | | | | | | | | | |
| 9 ft | 2.008 | 0.646 | 0.398 | 1.544 | 1.111 | 2.008 | 0.332 | 1.273 | 0.960 |
| 10 ft | 0.258 | 0.602 | 0.220 | 0.825 | 1.174 | 1.019 | 0.624 | 0.229 | 0.925 |
| 11 ft | 0.199 | 0.977 | 0.404 | 1.763 | 1.134 | 1.098 | 1.754 | 1.400 | 0.869 |
| Speed Limit | | | | | | | | | |
| 25 mph | 0.042 | 1.617 | 0.289 | 2.714 | 0.869 | 0.782 | 0.270 | 0.210 | 1.405 |
| 30 mph | 0.510 | 1.466 | 0.996 | 0.034 | 2.009 | 0.123 | 1.319 | 0.437 | 1.118 |
| 35 mph | 1.718 | 1.120 | 0.314 | 0.225 | 2.591 | 1.673 | 1.051 | 0.881 | 0.608 |
| 40 mph | 0.238 | 0.886 | 1.295 | 0.559 | 2.561 | 0.423 | 0.121 | 0.593 | 0.324 |
| 50 mph | 0.902 | 1.012 | 0.285 | 0.710 | 0.164 | 0.805 | 0.371 | 1.281 | 0.332 |
| 55 mph | 0.068 | 0.461 | 0.166 | 0.617 | 3.033 | 0.387 | 1.550 | 0.401 | 1.482 |
| Presence of Lighting | 0.907 | 0.992 | 0.530 | 0.757 | 0.158 | 0.480 | 1.091 | 1.526 | 0.963 |
| Presence of Horizontal Curve | 0.725 | 0.319 | 1.995 | 1.273 | 0.825 | 0.371 | 0.455 | 2.052 | 1.401 |

**Table 21. Descriptive statistics for continuous variables.**

| Continuous Variables | Statistic | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Dataset 5 | Dataset 6 | Dataset 7 | Dataset 8 | Dataset 9 | Dataset 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Crash Counts PDO | Mean | 52.17[a] | 48.0 | 48.43 | 47.94 | 48.03 | 49.03 | 47.51 | 48.94 | 40.070 | 50.76 |
| Crash Counts PDO . | Std. Dev. | 69.974[b] | 62.397 | 59.587 | 62.614 | 65.780 | 63.277 | 63.49 | 69.625 | 64.699 | 68.167 |
| Crash Counts K | Mean | 0.066 | 0.074 | 0.06 | 0.071 | 0.0742 | 0.072 | 0.073 | 0.083 | 0.070 | 0.072 |
| Crash Counts K | Std. Dev. | 0.437 | 0.446 | 0.427 | 0.451 | 0.503 | 0.463 | 0.440 | 0.576 | 0.448 | 0.503 |
| Crash Counts A | Mean | 1.138 | 1.048 | 1.057 | 1.078 | 1.108 | 1.09 | 1.085 | 1.1 | 1.122 | 1.112 |
| Crash Counts A | Std. Dev. | 2.129 | 1.769 | 1.698 | 1.72 | 1.926 | 1.887 | 1.979 | 1.945 | 1.965 | 1.933 |
| Crash Counts B. | Mean | 4.098 | 4.142 | 4.249 | 4.432 | 4.354 | 4.038 | 4.062 | 4.216 | 4.178 | 4.34 |
| Crash Counts B. | Std. Dev. | 8.345 | 8.561 | 9.578 | 11.732 | 10.366 | 7.153 | 8.462 | 8.537 | 8.450 | 9.621 |
| Crash Counts C | Mean | 10.9 | 11.41 | 11.42 | 11.79 | 11.31 | 10.94 | 11.08 | 11.22 | 11.66 | 11.26 |
| Crash Counts C | Std. Dev. | 15.755 | 20.053 | 17.871 | 20.982 | 18.151 | 18.19 | 16.94 | 18.00 | 20.36 | 17.479 |
| Ln (AADT Major) | Mean | 9.565 | 9.547 | 9.55 | 9.557 | 9.557 | 9.563 | 9.54 | 9.541 | 9.570 | 9.566 |
| Ln (AADT Major) | Std. Dev. | 0.565 | 0.551 | 0.554 | 0.5620 | 0.5612 | 0.551 | 0.557 | 0.584 | 0.572 | 0.569 |
| Ln (AADT Minor) | Mean | 8.458 | 8.43 | 8.435 | 8.443 | 8.441 | 8.470 | 8.431 | 8.428 | 8.443 | 8.451 |
| Ln (AADT Minor) | Std. Dev. | 0.812 | 40.798 | 0.812 | 0.815 | 0.818 | 0.815 | 0.79 | 0.817 | 0.821 | 0.819 |

[a]Parameter estimate.
[b]Standard error

**Table 22. Descriptive statistics for categorical variables (datasets 1–10).**

| Variables | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Dataset 5 | Dataset 6 | Dataset 7 | Dataset 8 | Dataset 9 | Dataset 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| LTLs | | | | | | | | | | |
| 1 | 370 | 729 | 1147 | 1479 | 1871 | 411 | 733 | 1145 | 1445 | 1805 |
| 2 | 232 | 505 | 727 | 998 | 1225 | 238 | 488 | 740 | 998 | 1250 |
| 3 | 127 | 256 | 379 | 520 | 622 | 121 | 273 | 407 | 531 | 667 |
| 4 | 271 | 510 | 747 | 1003 | 1282 | 230 | 506 | 708 | 1026 | 1278 |
| RTLs | | | | | | | | | | |
| 1 | 874 | 1730 | 2575 | 3450 | 4313 | 859 | 1721 | 2649 | 3436 | 4334 |
| 2 | 95 | 180 | 303 | 380 | 487 | 106 | 182 | 246 | 396 | 462 |
| 3 | 22 | 64 | 84 | 105 | 118 | 22 | 65 | 64 | 108 | 138 |
| 4 | 9 | 26 | 38 | 65 | 82 | 13 | 32 | 41 | 60 | 66 |
| Speed Limit (mph) | | | | | | | | | | |
| 25 | 175 | 319 | 507 | 660 | 826 | 167 | 322 | 517 | 656 | 882 |
| 30 | 216 | 431 | 627 | 790 | 1016 | 180 | 414 | 549 | 810 | 965 |
| 35 | 281 | 600 | 841 | 1167 | 1399 | 287 | 586 | 867 | 1157 | 1395 |
| 40 | 199 | 406 | 656 | 869 | 1139 | 230 | 433 | 696 | 897 | 1111 |
| 45 | 100 | 205 | 300 | 430 | 495 | 107 | 204 | 309 | 397 | 530 |
| 50 | 20 | 32 | 50 | 72 | 99 | 21 | 33 | 51 | 69 | 96 |
| 55 | 7 | 4 | 14 | 6 | 19 | 6 | 3 | 8 | 10 | 13 |
| 65 | 2 | 3 | 5 | 6 | 7 | 2 | 5 | 3 | 4 | 8 |
| LTLs Permitted | | | | | | | | | | |
| 1 | 715 | 1432 | 2212 | 2936 | 3595 | 760 | 1424 | 2221 | 2883 | 3617 |
| 2 | 180 | 373 | 535 | 715 | 920 | 152 | 388 | 508 | 752 | 928 |
| 3 | 43 | 86 | 116 | 172 | 258 | 51 | 91 | 146 | 182 | 221 |
| 4 | 62 | 109 | 137 | 177 | 227 | 37 | 97 | 126 | 183 | 234 |
| LTLs Protected | | | | | | | | | | |
| 1 | 934 | 1818 | 2744 | 3691 | 4644 | 919 | 1833 | 2765 | 3662 | 4628 |
| 2 | 47 | 134 | 182 | 234 | 253 | 64 | 126 | 179 | 243 | 282 |
| 3 | 5 | 19 | 25 | 34 | 41 | 6 | 17 | 26 | 44 | 40 |
| 4 | 14 | 29 | 49 | 41 | 62 | 11 | 24 | 30 | 51 | 50 |
| LTLs Mix | | | | | | | | | | |
| 1 | 716 | 1475 | 2175 | 2832 | 3604 | 743 | 1465 | 2174 | 2868 | 3510 |
| 2 | 145 | 261 | 445 | 618 | 738 | 133 | 272 | 433 | 604 | 805 |

| Variables | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Dataset 5 | Dataset 6 | Dataset 7 | Dataset 8 | Dataset 9 | Dataset 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 50 | 112 | 143 | 187 | 257 | 49 | 92 | 154 | 181 | 271 |
| 4 | 89 | 152 | 237 | 363 | 401 | 75 | 171 | 239 | 347 | 414 |
| RTOR | | | | | | | | | | |
| 1 | 909 | 1799 | 2711 | 3601 | 4475 | 892 | 1780 | 2680 | 3553 | 4628 |
| 2 | 31 | 79 | 101 | 158 | 183 | 42 | 82 | 94 | 165 | 282 |
| 3 | 35 | 67 | 119 | 148 | 219 | 43 | 93 | 140 | 168 | 40 |
| 4 | 25 | 55 | 69 | 93 | 123 | 23 | 45 | 86 | 114 | 50 |
| Lighting | | | | | | | | | | |
| No | 145 | 245 | 382 | 526 | 617 | 144 | 267 | 408 | 551 | 637 |
| Yes | 855 | 1755 | 2618 | 3474 | 4383 | 856 | 1733 | 2592 | 3449 | 4363 |
| Presence of School Within 1,000 ft of Intersection | | | | | | | | | | |
| No | 874 | 1745 | 2965 | 3503 | 4393 | 893 | 1751 | 2610 | 3538 | 4384 |
| Yes | 126 | 255 | 35 | 487 | 607 | 107 | 249 | 390 | 462 | 616 |
| Number of Bus Stops Within 1,000 ft of Intersection | | | | | | | | | | |
| 0 | 823 | 1696 | 2543 | 3356 | 4246 | 845 | 1697 | 2540 | 3392 | 4226 |
| 1–2 | 45 | 83 | 108 | 147 | 165 | 38 | 71 | 103 | 129 | 185 |
| ≥3 | 132 | 221 | 349 | 497 | 589 | 117 | 232 | 357 | 479 | 539 |
| Number of Alcohol Sale Establishments Within 1,000 ft of Intersection | | | | | | | | | | |
| 0 | 930 | 1846 | 2750 | 3673 | 4625 | 917 | 1857 | 2772 | 3684 | 4634 |
| 1–8 | 70 | 154 | 250 | 327 | 375 | 83 | 143 | 228 | 316 | 366 |
| ≥9 | — | — | — | — | — | 0 | — | — | — | — |
| Maximum Crossing Lanes | | | | | | | | | | |
| 1 | 1 | 1 | 3 | 1 | 2 | 0 | 1 | 3 | 3 | 3 |
| 2 | 189 | 369 | 626 | 798 | 1006 | 230 | 396 | 594 | 793 | 973 |
| 3 | 179 | 314 | 441 | 644 | 864 | 158 | 306 | 463 | 652 | 812 |
| 4 | 234 | 490 | 714 | 958 | 1179 | 246 | 478 | 712 | 955 | 1169 |
| 5 | 252 | 572 | 823 | 1060 | 1297 | 229 | 547 | 806 | 1026 | 1341 |
| 6 | 115 | 198 | 291 | 410 | 489 | 107 | 210 | 330 | 433 | 540 |
| 7 | 30 | 56 | 102 | 129 | 163 | 28 | 62 | 92 | 138 | 162 |

LTL = left-turn lane; RTL = right-turn lane; RTOR = right turn on red.

### *Crash Prediction Model Estimation*

Researchers estimated crash prediction models as described in table 23 section of this report. A total of 60 models were developed and validated for each severity level: K, A, B, C, and O. Table 23 summarizes the crash prediction model created using the dataset from the RAD tool. Estimated model parameters and fit statistics are provided here for only the 1,000-intersection model for brevity. The results for all the models are provided in appendix C through appendix M, which are in the second volume of this publication.

**Table 23. Summary of developed models.**

| ID | Facility | Datasets | Models | Severity Level | No. of Models |
|---|---|---|---|---|---|
| 1 | Urban four-leg signalized intersections | Two sets each of 1,000, 2,000, 3,000, 4,000, and 5,000 intersections | Poisson regression | K, A, B, C, PDO | 5×2 = 10 |
| | | | NB regression | K, A, B, C, PDO | 5×2 = 10 |
| | | | Poisson regression—random parameters | K, A, B, C, PDO | 5×2 = 10 |
| | | | NB—random parameters | K, A, B, C, PDO | 5×2 = 10 |
| | | | Univariate Poisson lognormal | K, A, B, C, PDO | 5×2 = 10 |
| | | | Multivariate Poisson lognormal | K, A, B, C, PDO | 5×2 = 10 |

The statistical software *R: A Language and Environment for Statistical Computing* (version 12.0) was used to estimate the model parameters (R Foundation 2021). Table 24, table 25, table 26, and table 27 summarize the model parameter estimates and their associated statistics under the Poisson, NB, random parameter, and univariate and multivariate Poisson lognormal models, respectively. An examination of the tables indicates that the model parameter estimates are significant at the 95-percent confidence level; however, the estimates that are followed by asterisks in these tables were not significant at a 95-percent confidence level.

## Table 24. Poisson regression model estimation (1,000 intersections).

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| Intercept | −29.068* | 9426.618 | −13.673 | 1.189 | −6.419 | 0.376 | −6.788 | 0.294 | −9.225 | 0.184 |
| Ln (AADT Major) | 0.778 | 0.296 | 0.956 | 0.072 | 0.673 | 0.035 | 0.748 | 0.022 | 0.919 | 0.011 |
| Ln (AADT Minor) | 0.282 | 0.176 | 0.518 | 0.047 | 0.361 | 0.023 | 0.323 | 0.014 | 0.484 | 0.007 |
| Left-Turn Lane | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −1.164 | 0.645 | −0.540 | 0.141 | −0.495 | 0.069 | −0.403 | 0.042 | −0.345 | 0.019 |
| 3 | −2.139 | 1.272 | −0.579 | 0.209 | −0.543 | 0.112 | −0.562 | 0.066 | −0.134 | 0.030 |
| 4 | 0.232 | 1.355 | −1.037 | 0.283 | −0.641 | 0.142 | −0.356 | 0.084 | −0.368 | 0.040 |
| Right-Turn Lane | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −0.708 | 0.529 | −0.167 | 0.108 | −0.328 | 0.061 | −0.279 | 0.037 | −0.100 | 0.016 |
| 3 | 0.009 | 0.744 | 0.050 | 0.174 | −0.721 | 0.136 | −0.524 | 0.078 | 0.203 | 0.025 |
| 4 | 0.878 | 0.757 | −0.490 | 0.340 | −0.642 | 0.189 | 0.572 | 0.074 | −0.661 | 0.057 |
| Speed Limit | | | | | | | | | | |
| 35 mph | Base Level | | | | | | | | | |
| 25 | −0.208 | 0.610 | −0.230 | 0.098 | 0.311 | 0.048 | 0.273 | 0.029 | 0.099 | 0.014 |
| 30 | 0.974* | 0.415 | −0.165 | 0.093 | 0.105 | 0.049 | −0.010 | 0.029 | 0.007 | 0.013 |
| 40 | 1.206* | 0.404 | 0.197 | 0.084 | 0.263 | 0.047 | 0.001 | 0.029 | 0.063 | 0.013 |
| 45 | 1.458* | 0.450 | −0.027 | 0.115 | −0.191 | 0.070 | 0.037 | 0.037 | 0.216 | 0.016 |
| 50 | 0.332* | 1.070 | 0.410 | 0.165 | 0.767 | 0.086 | 0.011 | 0.071 | −0.633 | 0.039 |
| 55 | −13.996* | 3185.047 | 0.063 | 0.463 | 0.735 | 0.175 | −0.059 | 0.148 | −0.321 | 0.076 |
| 65 | −16.579* | 6618.226 | 1.455 | 0.264 | 0.658 | 0.191 | −0.102 | 0.177 | −0.361 | 0.081 |
| Lighting | | | | | | | | | | |
| Not Present | Base Level | | | | | | | | | |
| Present | −0.364 | 0.315 | −0.228 | 0.082 | −0.115 | 0.044 | −0.347 | 0.025 | −0.330 | 0.012 |
| RTOR | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −15.958* | 1347.927 | 0.190 | 0.174 | 0.164 | 0.089 | 0.023 | 0.059 | −0.054 | 0.029 |
| 3 | −0.738 | 1.019 | −0.207 | 0.194 | 0.245 | 0.086 | 0.180 | 0.050 | 0.053 | 0.024 |
| 4 | −16.383* | 1505.027 | −0.227 | 0.231 | −0.262 | 0.114 | −0.213 | 0.071 | −0.222 | 0.033 |
| Maximum No. Crossing Lanes | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 15.660* | 9426.618 | 0.407 | 1.004 | −1.345 | 0.205 | −0.427 | 0.220 | 0.494 | 0.159 |
| 3 | 16.853* | 9426.618 | 0.384 | 1.007 | −1.336 | 0.209 | −0.324 | 0.221 | 0.540 | 0.159 |

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| 4 | 16.052* | 9426.618 | 0.328 | 1.004 | −1.623 | 0.206 | −0.271 | 0.220 | 0.641 | 0.159 |
| 5 | 16.731* | 9426.618 | 0.191 | 1.007 | −1.416 | 0.209 | −0.164 | 0.221 | 0.472 | 0.159 |
| 6 | 16.526* | 9426.618 | 0.330 | 1.010 | −1.419 | 0.213 | −0.298 | 0.223 | 0.453 | 0.160 |
| 7 | 15.758* | 9426.618 | 0.163 | 1.026 | −2.054 | 0.248 | −0.325 | 0.228 | 0.339 | 0.161 |
| LTLs Permitted | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.284 | 0.745 | 0.730 | 0.147 | −0.110 | 0.079 | 0.276 | 0.045 | 0.002 | 0.021 |
| 3 | −15.817* | 1146.280 | 0.337 | 0.256 | 0.349 | 0.129 | 0.041 | 0.082 | −0.380 | 0.038 |
| 4 | −0.408 | 1.377 | 0.623 | 0.312 | 0.159 | 0.155 | −0.096 | 0.091 | −0.012 | 0.043 |
| LTLs Mix | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.000 | 0.745 | 0.060 | 0.162 | −0.077 | 0.081 | −0.214 | 0.049 | −0.152 | 0.022 |
| 3 | −1.393 | 1.623 | 0.224 | 0.267 | −0.153 | 0.139 | −0.174 | 0.084 | −0.173 | 0.037 |
| 4 | −0.355 | 1.351 | 0.812 | 0.287 | 0.061 | 0.147 | −0.507 | 0.090 | −0.194 | 0.041 |
| LTLs Protected | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −1.923 | 1.295 | 0.048 | 0.236 | −0.123 | 0.125 | −0.114 | 0.070 | −0.251 | 0.033 |
| 3 | −16.255* | 3477.698 | 0.720 | 0.621 | −0.364 | 0.395 | −0.409 | 0.225 | −0.575 | 0.110 |
| 4 | −16.865* | 2270.057 | 0.852 | 0.377 | −0.466 | 0.236 | −0.454 | 0.129 | −0.722 | 0.066 |
| Bus | | | | | | | | | | |
| 0 | Base Level | | | | | | | | | |
| 1–2 | −0.285 | 0.547 | 0.052 | 0.137 | 0.032 | 0.073 | 0.142 | 0.041 | −0.092 | 0.020 |
| ≥3 | 0.493 | 0.330 | 0.058 | 0.089 | −0.236 | 0.052 | −0.054 | 0.029 | 0.006 | 0.013 |
| Alcohol | | | | | | | | | | |
| 0 | Base Level | | | | | | | | | |
| 1–8 | −0.210 | 0.609 | 0.246 | 0.116 | −0.477 | 0.078 | −0.170 | 0.042 | −0.017 | 0.019 |
| School | | | | | | | | | | |
| Not Present | Base Level | | | | | | | | | |
| Present | −0.153 | 0.412 | −0.093 | 0.095 | −0.119 | 0.050 | −0.349 | 0.034 | −0.055 | 0.014 |
| Fit Statistics | | | | | | | | | | |
| AIC | 527.48 | N/A | 2879.3 | N/A | 9162.4 | N/A | 13433 | N/A | 29547 | N/A |
| BIC | 718.879 | N/A | 3070.68 | N/A | 9353.81 | N/A | 13624.59 | N/A | 29738.76 | N/A |

—No data.

*Variables insignificant at a 90-percent confidence level.

## Table 25. NB model estimation (1,000 intersections).

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** | **Parameter Estimate** | **Std. Error** |
| Intercept | −52.286* | 6710.0 | −13.408 | 1.523 | −6.687 | 1.722 | −6.839 | 1.125 | −9.720 | 0.759 |
| Ln (AADT Major) | 0.999 | 0.427 | 0.963 | 0.093 | 0.659 | 0.100 | 0.723 | 0.064 | 0.991 | 0.042 |
| Ln (AADT Minor) | 0.178 | 0.267 | 0.473 | 0.061 | 0.443 | 0.069 | 0.362 | 0.044 | 0.460 | 0.029 |
| Left-Turn Lane | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −1.177 | 0.940 | −0.519 | 0.190 | −0.364 | 0.214 | −0.445 | 0.138 | −0.371 | 0.091 |
| 3 | −1.997 | 1.569 | −0.497 | 0.287 | −0.580 | 0.326 | −0.676 | 0.208 | −0.171 | 0.136 |
| 4 | 0.646 | 1.954 | −0.866 | 0.373 | −0.491 | 0.412 | −0.571 | 0.264 | −0.471 | 0.173 |
| Right-Turn Lane | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −0.610 | 0.750 | −0.167 | 0.149 | −0.302 | 0.174 | −0.314 | 0.111 | −0.142 | 0.072 |
| 3 | 0.422 | 1.120 | 0.030 | 0.266 | −0.442 | 0.353 | −0.326 | 0.222 | 0.357 | 0.143 |
| 4 | 1.263 | 1.686 | −0.545 | 0.475 | −0.611 | 0.547 | 0.316 | 0.333 | −0.440 | 0.226 |
| Speed Limit (mph) | | | | | | | | | | |
| 35 | Base Level | | | | | | | | | |
| 25 | −0.422* | 0.743 | −0.162 | 0.131 | 0.321 | 0.152 | 0.154 | 0.097 | 0.084 | 0.064 |
| 30 | 0.800 | 0.584 | −0.117 | 0.124 | 0.077 | 0.144 | −0.068 | 0.092 | 0.037 | 0.060 |
| 40 | 1.041 | 0.559 | 0.136 | 0.119 | 0.340 | 0.145 | 0.007 | 0.094 | 0.055 | 0.062 |
| 45 | 0.832 | 0.752 | 0.025 | 0.157 | −0.302 | 0.191 | −0.085 | 0.120 | 0.136 | 0.078 |
| 50 | 0.587 | 1.373 | 0.448 | 0.271 | 0.345 | 0.359 | −0.109 | 0.235 | −0.188 | 0.154 |
| 55 | −43.061* | 2565.3 | 0.173 | 0.545 | 0.633 | 0.594 | −0.119 | 0.397 | −0.385 | 0.261 |
| 65 | −58.519* | 47453.2 | 1.430 | 0.614 | 0.612 | 1.046 | −0.191 | 0.689 | −0.152 | 0.458 |
| Lighting | | | | | | | | | | |
| Not Present | Base Level | | | | | | | | | |
| Present | 0.145 | 0.537 | −0.181 | 0.117 | −0.401 | 0.140 | −0.376 | 0.090 | −0.346 | 0.060 |
| RTOR | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −54.061* | 12077.2 | 0.151 | 0.242 | 0.121 | 0.288 | 0.087 | 0.185 | −0.074 | 0.122 |
| 3 | −0.588 | 1.331 | −0.118 | 0.244 | 0.224 | 0.272 | 0.153 | 0.173 | 0.041 | 0.115 |
| 4 | −54.394* | 1357.1 | −0.185 | 0.296 | −0.077 | 0.327 | −0.055 | 0.210 | 0.041 | 0.136 |

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| Maximum No. Crossing Lanes | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 37.701* | 671000 | 0.499 | 1.264 | −1.535 | 1.456 | −0.406 | 0.959 | 0.510 | 0.652 |
| 3 | 38.491* | 671000 | 0.388 | 1.268 | −1.547 | 1.461 | −0.315 | 0.962 | 0.449 | 0.654 |
| 4 | 38.019* | 664.000 | 0.269 | 1.264 | −1.681 | 1.456 | −0.234 | 0.959 | 0.600 | 0.652 |
| 5 | 38.271* | 671000 | 0.213 | 1.268 | −1.490 | 1.460 | −0.151 | 0.962 | 0.448 | 0.654 |
| 6 | 38.020* | 67000 | 0.317 | 1.272 | −1.386 | 1.465 | −0.283 | 0.965 | 0.440 | 0.656 |
| 7 | 37.828* | 67000 | 0.016 | 1.295 | −1.578 | 1.485 | −0.380 | 0.978 | 0.588 | 0.663 |
| LTLs Permitted | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.367 | 1.012 | 0.663 | 0.196 | −0.181 | 0.222 | 0.284 | 0.141 | 0.020 | 0.093 |
| 3 | −54.576* | 1024.8 | 0.239 | 0.343 | 0.416 | 0.387 | 0.098 | 0.250 | −0.226 | 0.163 |
| 4 | −0.640 | 2.006 | 0.494 | 0.402 | −0.201 | 0.441 | 0.073 | 0.282 | 0.023 | 0.185 |
| LTLs Mix | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −0.094 | 1.038 | −0.015 | 0.211 | −0.102 | 0.230 | −0.182 | 0.148 | −0.026 | 0.097 |
| 3 | −1.172 | 2.167 | 0.216 | 0.346 | −0.169 | 0.389 | −0.157 | 0.250 | −0.013 | 0.162 |
| 4 | −1.191 | 1.948 | 0.586 | 0.380 | −0.079 | 0.423 | −0.260 | 0.272 | −0.014 | 0.178 |
| LTLs Protected | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −2.032 | 1.684 | −0.017 | 0.304 | −0.195 | 0.332 | 0.053 | 0.211 | −0.133 | 0.139 |
| 3 | −57.193* | 3001.9 | 0.521 | 0.750 | −0.500 | 0.842 | −0.293 | 0.521 | −0.497 | 0.334 |
| 4 | −58.888* | 17935.6 | 0.696 | 0.506 | −0.005 | 0.581 | 0.225 | 0.368 | −0.567 | 0.247 |
| Bus | | | | | | | | | | |
| 0 | Base Level | | | | | | | | | |
| 1–2 | −0.309 | 0.990 | 0.104 | 0.193 | −0.051 | 0.240 | 0.256 | 0.152 | 0.189 | 0.101 |
| ≥3 | 0.260 | 0.611 | 0.092 | 0.122 | −0.108 | 0.150 | 0.031 | 0.095 | −0.031 | 0.063 |
| Alcohol | | | | | | | | | | |
| 0 | Base Level | | | | | | | | | |
| 1–8 | 0.307 | 0.803 | 0.252 | 0.161 | −0.431 | 0.202 | −0.106 | 0.126 | −0.028 | 0.083 |
| School | | | | | | | | | | |
| Not Present | Base Level | | | | | | | | | |
| Present | 0.065 | 0.615 | −0.117 | 0.132 | −0.046 | 0.153 | −0.226 | 0.099 | −0.021 | 0.064 |
| Fit Statistics | | | | | | | | | | |
| AIC | 446.93 | N/A | 2666.8 | N/A | 4579.6 | N/A | 6536.7 | N/A | 9018.6 | N/A |
| BIC | 643.23 | N/A | 2863.10 | N/A | 4775.91 | N/A | 6732.98 | N/A | 9214.88 | N/A |
| Overdispersion | 0.0769 | — | 1.718 | — | 0.486 | — | 1.157 | — | 2.52 | — |

—No data.
*Variables insignificant at a 90-percent confidence level.

## Table 26. Random parameter Poisson regression (1,000 intersections).

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| Intercept | −42.122* | 6.385 | −7.907 | 1.048 | −2.428 | 1.570 | −3.551 | 0.860 | −1.760 | 0.449 |
| Ln (AADT Major) | 1.071 | 0.323 | 1.007 | 0.055 | 1.134 | 0.101 | 1.008 | 0.049 | 1.022 | 0.026 |
| Ln (AADT Minor) | 0.762 | 0.470 | 0.879 | 0.079 | 0.260 | 0.117 | 0.452 | 0.063 | 0.531 | 0.033 |
| Left-Turn Lane | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.585 | 0.550 | 0.449 | 0.110 | 0.456 | 0.189 | 0.466 | 0.107 | 0.274 | 0.046 |
| 3 | 0.253 | 0.547 | 0.103 | 0.111 | 0.438 | 0.185 | 0.328 | 0.104 | 0.292 | 0.043 |
| 4 | −0.741 | 0.781 | −0.004 | 0.130 | −0.015 | 0.232 | 0.187 | 0.122 | 0.151 | 0.051 |
| Right-Turn Lane | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.026 | 0.092 | 0.002 | 0.015 | 0.041 | 0.030 | −0.012 | 0.013 | 0.012 | 0.006 |
| 3 | 0.002 | 0.009 | 0.003 | 0.023 | 0.589 | 0.298 | 0.369 | 0.759 | 0.379 | 0.349 |
| 4 | 0.281 | 1.065 | −0.234 | 0.197 | −0.299 | 0.371 | −0.686 | 0.191 | −0.507 | 0.088 |
| Speed Limit | | | | | | | | | | |
| 35 mph | Base Level | | | | | | | | | |
| 25 | −0.422* | 0.743 | −0.162 | 0.131 | 0.321 | 0.152 | 0.154 | 0.097 | 0.084 | 0.064 |
| 30 | 0.800* | 0.584 | −0.117 | 0.124 | 0.077 | 0.144 | −0.068 | 0.092 | 0.037 | 0.060 |
| 40 | 1.041* | 0.559 | 0.136 | 0.119 | 0.340 | 0.145 | 0.007 | 0.094 | 0.055 | 0.062 |
| 45 | 0.832* | 0.752 | 0.025 | 0.157 | −0.302 | 0.191 | −0.085 | 0.120 | 0.136 | 0.078 |
| 50 | 0.587* | 1.373 | 0.448 | 0.271 | 0.345 | 0.359 | −0.109 | 0.235 | −0.188 | 0.154 |
| 55 | −43.061* | 2565.3 | 0.173 | 0.545 | 0.633 | 0.594 | −0.119 | 0.397 | −0.385 | 0.261 |
| 65 | −58.519* | 47453.2 | 1.430 | 0.614 | 0.612 | 1.046 | −0.191 | 0.689 | −0.152 | 0.458 |
| Lighting | | | | | | | | | | |
| Not Present | Base Level | | | | | | | | | |
| Present | 0.288 | 0.537 | −0.317 | 0.107 | −0.210 | 0.131 | −0.268 | 0.080 | −0.185 | 0.051 |
| RTOR | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.037 | 0.165 | −0.009 | 0.018 | 0.014 | 0.047 | −0.003 | 0.015 | 0.016 | 0.008 |
| 3 | 0.006 | 0.021 | 0.003 | 0.002 | 0.005 | 0.005 | 0.001 | 0.002 | 0.000 | 0.001 |
| 4 | 0.128 | 2.363 | −0.420 | 0.244 | −0.330 | 0.553 | −0.754 | 0.217 | −0.534 | 0.100 |
| Maximum No. Crossing Lanes | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 20.453 | 1.037 | 0.243 | 0.126 | 0.400 | 0.234 | −0.164 | 0.112 | 0.164 | 0.051 |
| 3 | 22.119 | 2.250 | 0.223 | 0.163 | 0.253 | 0.305 | 0.155 | 0.133 | 0.203 | 0.066 |
| 4 | 20.509 | 2.606 | 0.597 | 0.200 | 0.239 | 0.417 | −0.056 | 0.186 | 0.464 | 0.084 |

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error |
| 5 | 20.073 | 0.158 | −0.016 | 0.015 | −0.013 | 0.028 | −0.009 | 0.014 | 0.009 | 0.006 |
| 6 | 20.017 | 0.017 | −0.001 | 0.002 | 0.000 | 0.003 | 0.004 | 0.001 | −0.001 | 0.001 |
| 7 | 22.600 | 1.683 | −0.526 | 0.219 | −0.351 | 0.377 | −0.288 | 0.171 | −0.287 | 0.084 |
| LTLs Permitted | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.892 | 1.323 | 0.093 | 0.178 | 0.341 | 0.299 | −0.204 | 0.162 | 0.242 | 0.071 |
| 3 | 0.684 | 0.582 | 0.182 | 0.095 | 0.237 | 0.175 | 0.094 | 0.086 | 0.141 | 0.040 |
| 4 | 0.259 | 1.662 | 0.279 | 0.220 | 0.186 | 0.440 | −0.128 | 0.213 | 0.343 | 0.095 |
| LTLs Mix | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.376 | 0.682 | 0.267 | 0.094 | 0.122 | 0.166 | 0.232 | 0.089 | 0.207 | 0.040 |
| 3 | 1.849 | 0.720 | 0.275 | 0.122 | 0.609 | 0.197 | 0.151 | 0.117 | 0.309 | 0.050 |
| 4 | 1.050 | 1.230 | 0.245 | 0.169 | 0.405 | 0.282 | 0.638 | 0.142 | 0.377 | 0.067 |
| LTLs Protected | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.130 | 0.202 | 0.015 | 0.017 | −0.012 | 0.028 | −0.019 | 0.014 | 0.005 | 0.007 |
| 3 | 0.005 | 0.016 | 0.002 | 0.002 | −0.006 | 0.003 | 0.002 | 0.002 | 0.002 | 0.001 |
| 4 | 2.760 | 1.848 | 0.071 | 0.204 | −0.022 | 0.352 | −0.139 | 0.191 | −0.642 | 0.098 |
| Bus | | | | | | | | | | |
| 0 | Base Level | | | | | | | | | |
| 1–2 | −0.679 | 1.421 | 0.067 | 0.202 | 0.213 | 0.238 | 0.120 | 0.147 | 0.073 | 0.093 |
| ≥3 | −1.488 | 0.919 | 0.084 | 0.132 | 0.061 | 0.146 | 0.140 | 0.089 | −0.061 | 0.056 |
| Alcohol | | | | | | | | | | |
| 0 | Base Level | | | | | | | | | |
| 1–8 | 2.419 | 0.550 | 0.884 | 0.122 | −0.031 | 0.169 | −0.054 | 0.104 | −0.074 | 0.066 |
| School | | | | | | | | | | |
| Not Present | Base Level | | | | | | | | | |
| Present | 1.093 | 0.574 | 0.432 | 0.118 | −0.073 | 0.151 | 0.083 | 0.092 | −0.097 | 0.058 |
| Fit Statistics | | | | | | | | | | |
| AIC | 446.97 | N/A | 2589.9 | N/A | 4680.3 | N/A | 6346.2 | N/A | 8563.1 | N/A |
| BIC | 638.375 | N/A | 2781.29 | N/A | 4871.75 | N/A | 6537.60 | N/A | 8,754.53 | N/A |

—No data.
*Variables insignificant at a 90-percent confidence level.

## Table 27. Random parameter—NB regression (1,000 intersections).

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate |
| Intercept | 45.292* | 7.476 | −4.946 | 2.538 | −4.689 | 4.723 | −1.963 | 1.874 | −1.003 | 1.176 |
| Ln AADT Major | 1.196 | 0.677 | 0.856 | 0.601 | 1.523 | 0.232 | 1.030 | 0.084 | 0.966 | 0.050 |
| Ln AADT Minor | 0.329 | 1.325 | 0.683 | 0.199 | 0.421 | 0.360 | 0.419 | 0.147 | 0.415 | 0.090 |
| Left-Turn Lane | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.537 | 0.693 | 0.122 | 0.129 | 0.373 | 0.221 | 0.320 | 0.114 | 0.288 | 0.059 |
| 3 | −0.491 | 0.957 | −0.020 | 0.155 | −0.096 | 0.279 | 0.197 | 0.134 | 0.167 | 0.071 |
| 4 | 0.911 | 0.707 | 0.475 | 0.130 | 0.475 | 0.226 | 0.500 | 0.117 | 0.298 | 0.063 |
| Right-Turn Lane | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 1.392 | 1.159 | 0.122 | 0.120 | −0.146 | 0.238 | −0.092 | 0.100 | 0.015 | 0.049 |
| 3 | 0.485 | 1.391 | 0.083 | 0.152 | 0.073 | 0.301 | 0.093 | 0.125 | 0.051 | 0.064 |
| 4 | 0.867 | 1.585 | 0.019 | 0.191 | 0.313 | 0.371 | 0.205 | 0.156 | 0.251 | 0.080 |
| Speed Limit | | | | | | | | | | |
| 35 mph | Base Level | | | | | | | | | |
| 25 | 3.088* | 1.793 | 0.303 | 0.313 | −0.222 | 0.660 | 0.549 | 0.252 | 0.041 | 0.131 |
| 30 | −0.097* | 3.564 | 0.353 | 0.300 | 1.175 | 0.552 | 0.817 | 0.237 | 0.041 | 0.136 |
| 40 | 2.391* | 1.939 | −0.001 | 0.332 | 1.096 | 0.529 | 0.717 | 0.241 | −0.295 | 0.144 |
| 45 | −2.975* | 445322 | −0.920 | 0.275 | −0.604 | 0.498 | −0.553 | 0.230 | −1.110 | 0.120 |
| 50 | −2.941* | 618229 | −0.552 | 0.341 | −0.248 | 0.612 | 0.211 | 0.258 | 0.004 | 0.120 |
| 55 | 0.634* | 1.399 | −0.396 | 0.166 | −0.546 | 0.349 | −0.101 | 0.153 | −0.330 | 0.071 |
| 65 | 0.288* | 0.537 | −0.317 | 0.107 | −0.210 | 0.131 | −0.268 | 0.080 | −0.185 | 0.051 |
| Lighting | | | | | | | | | | |
| Not Present | Base Level | | | | | | | | | |
| Present | −1.923 | 1.295 | 0.048 | 0.236 | −0.123 | 0.125 | −0.114 | 0.070 | −0.251 | 0.033 |
| RTOR | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 7.258* | 17237 | 1.068 | 0.589 | 0.747 | 1.238 | 0.896 | 0.480 | −0.013 | 0.374 |
| 3 | −0.928 | 14010 | −0.386 | 0.869 | 0.807 | 1.332 | −0.061 | 0.656 | 0.268 | 0.343 |
| 4 | 6.742* | 12284 | −1.169 | 0.730 | 0.730 | 1.051 | −1.625 | 0.774 | −0.879 | 0.319 |
| Maximum No. Crossing Lanes | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 22.158* | 1098.5 | 1.219 | 1.723 | −0.001 | 0.840 | 0.130 | 0.568 | 0.487 | 0.435 |
| 3 | 21.745* | 1098.5 | 1.143 | 1.426 | 0.068 | 0.841 | 0.166 | 0.569 | 0.722 | 0.352 |

| Variables | K | | A | | B | | C | | PDO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate | Parameter Estimate | Std. Error | Parameter Estimate | Std. Error | Parameter Estimate |
| 4 | 22.356* | 1088.5 | 1.216 | 1.625 | −0.069 | 0.840 | 0.047 | 0.568 | 1.146 | 0.598 |
| 5 | 22.316* | 1088.5 | 1.181 | 1.129 | −0.086 | 0.841 | 0.124 | 0.569 | 1.403 | 0.360 |
| 6 | 22.162* | 1088.5 | 1.115 | 1.128 | 0.083 | 0.843 | 0.034 | 0.570 | 1.774 | 0.369 |
| 7 | 22.422* | 1088.5 | 1.123 | 1.132 | −0.087 | 0.848 | −0.078 | 0.573 | 1.935 | 0.334 |
| LTLs Permitted | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −1.164 | 0.645 | −0.540 | 0.141 | −0.495 | 0.069 | −0.403 | 0.042 | −0.345 | 0.019 |
| 3 | −2.139 | 1.272 | −0.579 | 0.209 | −0.543 | 0.112 | −0.562 | 0.066 | −0.134 | 0.030 |
| 4 | 0.232 | 1.355 | −1.037 | 0.283 | −0.641 | 0.142 | −0.356 | 0.084 | −0.368 | 0.040 |
| LTLs Mix | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.000 | 0.745 | 0.060 | 0.162 | −0.077 | 0.081 | −0.214 | 0.049 | −0.152 | 0.022 |
| 3 | −1.393 | 1.623 | 0.224 | 0.267 | −0.153 | 0.139 | −0.174 | 0.084 | −0.173 | 0.037 |
| 4 | −0.355 | 1.351 | 0.812 | 0.287 | 0.061 | 0.147 | −0.507 | 0.090 | −0.194 | 0.041 |
| LTLs Protected | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −0.337 | 8413.9 | −0.794 | 0.383 | 0.038 | 0.784 | −1.140 | 0.372 | −1.039 | 0.180 |
| 3 | −1.311 | 84015 | −0.874 | 0.342 | −0.347 | 0.747 | −0.463 | 0.281 | −0.194 | 0.133 |
| 4 | −0.279 | 8926.8 | −0.055 | 0.347 | 0.058 | 0.863 | −0.460 | 0.326 | −0.141 | 0.151 |
| Bus | | | | | | | | | | |
| 0 | Base Level | | | | | | | | | |
| 1–2 | 0.957 | 0.406 | 1.146 | 0.068 | 1.200 | 0.127 | 1.011 | 0.060 | 1.006 | 0.026 |
| ≥3 | 0.490 | 0.546 | 0.636 | 0.085 | 0.414 | 0.150 | 0.519 | 0.076 | 0.612 | 0.034 |
| Alcohol | | | | | | | | | | |
| 0 | Base Level | | | | | | | | | |
| 1−8 | 0.425 | 0.294 | 0.506 | 0.056 | −0.016 | 0.075 | 0.013 | 0.049 | −0.085 | 0.028 |
| School | | | | | | | | | | |
| Not Present | Base Level | | | | | | | | | |
| Present | 0.511 | 0.257 | 0.482 | 0.050 | 0.018 | 0.064 | 0.096 | 0.043 | 0.029 | 0.024 |
| Fit Statistics | | | | | | | | | | |
| AIC | 2018.3 | N/A | 12901 | N/A | 23319 | N/A | 32515 | N/A | 42322 | N/A |
| BIC | 2279.01 | N/A | 13161.73 | N/A | 23579.51 | N/A | 32776.17 | N/A | 42583.07 | N/A |
| Overdispersion | 0.0557 | — | 2.026 | — | 0.5428 | — | 1.1794 | — | 3.907 | — |

—No data.

*Variables insignificant at a 90-percent confidence level.

### Validation of Crash Count Prediction

The model prediction of dataset 10 used for model estimation is shown in table 28; meanwhile, dataset 5, which had the same number of intersections, was used for cross validation.

**Table 28. Model prediction with dataset 10 (5,000 intersections) used for estimation.**

| Models | Metric | K | A | B | C | PDO |
|---|---|---|---|---|---|---|
| Poisson Regression | MAD | 3.24 | 1.38 | 3.16 | 9.13 | 44.56 |
| | MSPE | 11.94 | 4.68 | 113.18 | 402.25 | 6236.01 |
| NB Regression | MAD | 3.25 | 1.37 | 3.14 | 9.23 | 44.58 |
| | MSPE | 13.59 | 6.67 | 113.30 | 402.24 | 6236.42 |
| Poisson regression—random parameters | MAD | 3.35 | 2.37 | 4.14 | 9.34 | 43.88 |
| | MSPE | 14.59 | 4.80 | 131.30 | 522.74 | 5036.42 |
| NB—random parameters | MAD | 3.32 | 1.73 | 3.15 | 9.23 | 44.68 |
| | MSPE | 15.60 | 4.76 | 113.27 | 420.24 | 5236.42 |
| Univariate Poisson lognormal | MAD | 3.20 | 1.38 | 3.12 | 9.21 | 44.56 |
| | MSPE | 12.84 | 4.80 | 112.18 | 402.25 | 5106.01 |
| Multivariate Poisson lognormal | MAD | 4.25 | 1.67 | 3.24 | 9.32 | 45.58 |
| | MSPE | 15.59 | 5.67 | 123.30 | 502.24 | 5236.42 |

In this case, univariate Poisson lognormal performed better in the prediction of MSPE for K crashes while Poisson regression did better for MAD for K crashes. NB had the lowest MAD and MSPE for A crashes, which shows that it performed better. Univariate Poisson lognormal had the lowest MAD for B crashes while Poisson regression had the lowest for MSPE. Poisson regression performed better for C crashes for MAD while NB did better for MSPE. Finally, Poisson did better for MAD for PDO crashes while univariate Poisson lognormal did better for MSPE. The cross-validation results are shown in table 29, table 30, and table 31. Cross validation was used to assess the predictive performance of the models and judge how they performed outside the sample in a new dataset, also known as test data. The cross-validation values for the different models fall within the same range, which could mean the models are robust enough to be relied on.

**Table 29. Cross validation with dataset 5 (5,000 intersections) using dataset 10 model.**

| Models | Metric | K | A | B | C | PDO |
|---|---|---|---|---|---|---|
| Poisson regression | MAD | 3.25 | 1.39 | 3.15 | 9.08 | 47.28 |
| | MSPE | 12.15 | 4.74 | 98.66 | 378.05 | 6800.66 |
| NB regression | MAD | 3.27 | 1.38 | 3.23 | 7.13 | 47.27 |
| | MSPE | 14.21 | 4.72 | 98.68 | 98.69 | 6800.94 |
| Poisson regression—random parameters | MAD | 3.35 | 2.73 | 3.14 | 7.23 | 63.87 |
| | MSPE | 15.90 | 5.60 | 131.30 | 522.74 | 5036.42 |
| NB—random parameters | MAD | 3.19 | 1.40 | 3.20 | 3.32 | 48.27 |
| | MSPE | 13.21 | 3.72 | 97.86 | 99.79 | 6832.49 |
| Univariate Poisson lognormal | MAD | 4.20 | 2.83 | 2.42 | 9.32 | 45.56 |
| | MSPE | 10.84 | 3.79 | 102.82 | 502.25 | 5216.12 |
| Multivariate Poisson lognormal | MAD | 3.30 | 1.38 | 3.13 | 3.13 | 46.27 |
| | MSPE | 15.21 | 5.72 | 99.70 | 377.69 | 5800.94 |

**Table 30. Model prediction with dataset 5 (5,000 intersections) used for estimation.**

| Models | Metric | K | A | B | C | PDO |
|---|---|---|---|---|---|---|
| Poisson regression | MAD | 3.31 | 1.37 | 3.13 | 9.07 | 53.99 |
| | MSPE | 12.58 | 4.66 | 98.43 | 377.75 | 7489.81 |
| NB regression | MAD | 3.33 | 1.37 | 3.12 | 9.07 | 47.24 |
| | MSPE | 14.91 | 4.67 | 98.49 | 377.67 | 6796.74 |
| Poisson regression—random parameters | MAD | 3.25 | 3.01 | 2.01 | 10.98 | 49.6 |
| | MSPE | 15.79 | 5.89 | 96.91 | 383.82 | 7116.87 |
| NB—random parameters | MAD | 4.14 | 1.45 | 2.51 | 10.95 | 48.85 |
| | MSPE | 16.01 | 4.75 | 98.31 | 393.32 | 6920.38 |
| Univariate Poisson lognormal | MAD | 3.1 | 2.14 | 2.91 | 10.23 | 51.99 |
| | MSPE | 13.92 | 4.58 | 95.83 | 370.86 | 6873.26 |
| Multivariate Poisson lognormal | MAD | 5.05 | 2.41 | 2.85 | 9.5 | 54.54 |
| | MSPE | 15.47 | 4.93 | 97.2 | 376.7 | 7092.44 |

**Table 31. Cross validation with dataset 10 (5,000 intersections) using dataset 5 model.**

| Models | Metric | K | A | B | C | PDO |
|---|---|---|---|---|---|---|
| Poisson regression | MAD | 3.3 | 1.38 | 1.38 | 9.07 | 53.98 |
| | MSPE | 12.68 | 36.22 | 54.09 | 103.12 | 7000.66 |
| NB regression | MAD | 3.33 | 1.38 | 3.12 | 9.07 | 47.26 |
| | MSPE | 14.49 | 6.19 | 103 | 388.34 | 6879.1 |
| Poisson regression—random parameters | MAD | 4.78 | 2.79 | 2.59 | 9.95 | 53.4 |
| | MSPE | 12.4 | 7.71 | 99.22 | 159.65 | 6608.34 |
| NB—random parameters | MAD | 3.32 | 1.88 | 1.51 | 8.43 | 48.51 |
| | MSPE | 12.59 | 10 | 99.41 | 120.17 | 6634.81 |
| Univariate Poisson lognormal | MAD | 4.75 | 1.56 | 1.16 | 8.97 | 50.16 |
| | MSPE | 12.8 | 9.93 | 99.29 | 232.12 | 6759.22 |
| Multivariate Poisson lognormal | MAD | 3.36 | 2.68 | 1.87 | 8.78 | 51.07 |
| | MSPE | 13.92 | 8.67 | 65.5 | 169.32 | 6772.57 |

**Empirical Analysis for Evaluating Parameter Stability**

Dataset 10 (5,000 intersections) was used as the population benchmark to evaluate if each parameter in any model is statistically different from the corresponding one in that dataset. The combined NB estimates of PDO crashes for the 10 datasets are displayed in table 32 with their standard errors for illustrative purposes. Table 33 shows the updated Wald test statistics for the estimated model parameter values (Hoover, Bhowmik, Yasmin, and Eluru 2022).

**Table 32. NB estimates for PDO crashes.**

| Variables | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Dataset 5 | Dataset 6 | Dataset 7 | Dataset 8 | Dataset 9 | Dataset 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | −9.719[a], 0.758[b] | −10.514,0.648 | −9.438, 0.364 | −9.829, 0.550 | −9.945, 0.442 | −9.107, 0.339 | −10.705, 0.733 | −9.367, 0.486 | −9.918, 0.364 | −9.858, 0.361 |
| Ln (AADT Major) | 0.991, 0.042 | 0.854, 0.025 | 0.881, 0.020 | 0.867, 0.016 | 0.834, 0.016 | 0.847, 0.036 | 0.888, 0.024 | 0.813, 0.026 | 0.868, 0.017 | 0.855, 0.015 |
| Ln (AADT Minor) | 0.459, 0.028 | 0.576, 0.017 | 0.543, 0.013 | 0.575, 0.011 | 0.561, 0.011 | 0.545, 0.024 | 0.539, 0.017 | 0.602, 0.018 | 0.563, 0.012 | 0.575, 0.011 |
| LTLs | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −0.370, 0.091 | −0.603, 0.053 | −0.564, 0.044 | −0.582, 0.037 | −0.700, 0.035 | −0.652, 0.076 | −0.741, 0.056 | −0.569, 0.062 | −0.679, 0.039 | −0.686, 0.034 |
| 3 | −0.171, 0.136 | −0.669, 0.077 | −0.495, 0.061 | −0.553, 0.053 | −0.652, 0.050 | −0.590, 0.109 | −0.763, 0.076 | −0.645, 0.085 | −0.732, 0.055 | −0.690, 0.049 |
| 4 | −0.471, 0.173 | −0.748, 0.097 | −0.739, 0.081 | −0.632, 0.070 | −0.808, 0.064 | −0.697, 0.140 | −0.952, 0.100 | −0.798, 0.111 | −0.832, 0.072 | −0.844, 0.063 |
| RTLs | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −0.142, 0.072 | −0.083, 0.042 | −0.102, 0.033 | −0.108, 0.029 | −0.039, 0.026 | −0.120, 0.059 | −0.032, 0.042 | −0.012, 0.050 | −0.084, 0.029 | −0.103, 0.027 |
| 3 | 0.356, 0.143 | −0.102, 0.069 | −0.126, 0.060 | −0.186, 0.054 | −0.123, 0.052 | −0.287, 0.128 | −0.037, 0.067 | −0.080, 0.095 | −0.107, 0.054 | −0.186, 0.047 |
| 4 | −0.440, 0.226 | −0.423, 0.107 | −0.146, 0.088 | −0.280, 0.067 | −0.257, 0.063 | −0.286, 0.161 | −0.196, 0.098 | −0.285, 0.118 | −0.242, 0.073 | −0.269, 0.070 |
| Speed Limit (mph) | | | | | | | | | | |
| 35 | Base Level | | | | | | | | | |
| 25 | 0.083, 0.064 | 0.117, 0.037 | 0.038, 0.030 | 0.0107, 0.0260 | 0.011, 0.024 | −0.077, 0.055 | −0.005, 0.037 | −0.018, 0.041 | 0.005, 0.027 | −0.020, 0.023 |
| 30 | 0.036, 0.060 | 0.052, 0.034 | 0.079, 0.028 | −0.027, 0.024 | 0.043, 0.023 | 0.032, 0.054 | 0.010, 0.035 | −0.027, 0.041 | −0.022, 0.025 | −0.013, 0.023 |
| 40 | 0.055, 0.062 | 0.048, 0.035 | 0.045, 0.028 | 0.006, 0.023 | −0.002, 0.022 | 0.010, 0.050 | 0.046, 0.034 | −0.027, 0.038 | −0.029, 0.025 | −0.027, 0.022 |
| 45 | 0.136, 0.078 | 0.004, 0.044 | −0.013, 0.036 | −0.021, 0.030 | −0.017, 0.029 | −0.065, 0.064 | −0.069, 0.044 | 0.058, 0.049 | 0.009, 0.032 | −0.033, 0.028 |
| 50 | −0.187, 0.154 | 0.042, 0.098 | 0.019, 0.078 | 0.128, 0.064 | 0.029, 0.057 | −0.199, 0.131 | −0.080, 0.098 | 0.082, 0.106 | −0.065, 0.068 | 0.064, 0.058 |
| 55 | −0.384, 0.261 | −0.270, 0.280 | −0.004, 0.146 | −0.212, 0.216 | 0.012, 0.124 | 0.125, 0.246 | −0.283, 0.363 | −0.292, 0.272 | −0.314, 0.177 | −0.261, 0.151 |

| Variables | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Dataset 5 | Dataset 6 | Dataset 7 | Dataset 8 | Dataset 9 | Dataset 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 65 | −0.151, 0.458 | −0.368, 0.320 | −0.113, 0.249 | −0.409, 0.222 | 0.215, 0.202 | 0.033, 0.406 | 0.111, 0.241 | −0.510, 0.451 | 0.419, 0.265 | −0.024, 0.192 |
| LTLs Permitted | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.019, 0.093 | 0.144, 0.053 | 0.047, 0.044 | 0.078, 0.038 | 0.161, 0.035 | 0.080, 0.079 | 0.206, 0.055 | 0.065, 0.062 | 0.089, 0.039 | 0.109, 0.034 |
| 3 | −0.225, 0.163 | 0.137, 0.096 | 0.042, 0.078 | 0.032, 0.067 | 0.093, 0.060 | 0.131, 0.134 | 0.309, 0.094 | 0.181, 0.105 | 0.153, 0.068 | 0.146, 0.061 |
| 4 | 0.023, 0.185 | 0.116, 0.106 | −0.004, 0.090 | 0.023, 0.078 | 0.037, 0.071 | −0.009, 0.163 | 0.262, 0.110 | 0.094, 0.125 | 0.077, 0.080 | 0.153, 0.070 |
| LTLs Protected | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −0.133, 0.139 | −0.324, 0.073 | −0.316, 0.059 | −0.302, 0.052 | −0.179, 0.050 | −0.215, 0.107 | −0.221, 0.076 | −0.337, 0.084 | −0.264, 0.054 | −0.256, 0.048 |
| 3 | −0.496, 0.334 | −0.398, 0.154 | −0.352, 0.128 | −0.443, 0.112 | −0.328, 0.104 | −0.302, 0.260 | −0.481, 0.164 | −0.195, 0.175 | −0.350, 0.105 | −0.453, 0.104 |
| 4 | −0.567, 0.247 | −0.535, 0.144 | −0.409, 0.111 | −0.583, 0.109 | −0.502, 0.095 | −0.310, 0.216 | −0.226, 0.149 | −0.963, 0.178 | −0.505, 0.105 | −0.402, 0.101 |
| LTLs Mix | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −0.026, 0.097 | −0.107, 0.058 | −0.131, 0.047 | −0.158, 0.040 | −0.091, 0.037 | −0.103, 0.081 | −0.003, 0.059 | −0.052, 0.064 | −0.099, 0.041 | −0.060, 0.037 |
| 3 | −0.013, 0.162 | −0.115, 0.092 | −0.227, 0.078 | −0.246, 0.067 | −0.119, 0.062 | −0.157, 0.136 | 0.050, 0.096 | −0.225, 0.106 | −0.081, 0.069 | −0.116, 0.060 |
| 4 | −0.014, 0.178 | −0.138, 0.103 | −0.202, 0.085 | −0.298, 0.073 | −0.135, 0.067 | −0.253, 0.149 | 0.017, 0.105 | −0.147, 0.116 | −0.169, 0.075 | −0.170, 0.067 |
| RTOR | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | −0.073, 0.122 | −0.180, 0.063 | −0.137, 0.055 | 0.071, 0.043 | −0.016, 0.041 | −0.109, 0.089 | −0.141, 0.061 | −0.162, 0.078 | −0.057, 0.044 | −0.113, 0.038 |
| 3 | 0.040, 0.115 | −0.117, 0.068 | −0.085, 0.051 | −0.069, 0.045 | −0.051, 0.038 | −0.190, 0.089 | −0.056, 0.057 | −0.079, 0.064 | −0.041, 0.044 | −0.063, 0.041 |
| 4 | 0.041, 0.136 | −0.222, 0.075 | −0.145, 0.066 | 0.013, 0.056 | −0.026, 0.051 | −0.203, 0.119 | −0.035, 0.080 | −0.013, 0.082 | −0.068, 0.052 | −0.150, 0.052 |
| Lighting | | | | | | | | | | |
| No | Base Level | | | | | | | | | |
| Yes | −0.346, 0.060 | −0.213, 0.036 | −0.211, 0.029 | −0.267, 0.024 | −0.297, 0.023 | −0.185, 0.051 | −0.216, 0.035 | −0.247, 0.039 | −0.280, 0.025 | −0.278, 0.023 |

| Variables | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Dataset 5 | Dataset 6 | Dataset 7 | Dataset 8 | Dataset 9 | Dataset 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Presence of School Within 1,000 ft of Intersection | | | | | | | | | | |
| No | Base Level | | | | | | | | | |
| Yes | −0.021, 0.064 | 0.028, 0.036 | 0.057, 0.090 | 0.022, 0.025 | 0.017, 0.024 | −0.097, 0.058 | 0.017, 0.037 | 0.005, 0.040 | −0.085, 0.028 | 0.029, 0.024 |
| Number of Bus Stops Within 1,000 ft of Intersection | | | | | | | | | | |
| 0 | Base Level | | | | | | | | | |
| 1–2 | 0.189, 0.101 | −0.096, 0.061 | −0.010, 0.053 | −0.080, 0.044 | 0.030, 0.043 | 0.073, 0.093 | −0.040, 0.065 | 0.035, 0.074 | −0.028, 0.050 | 0.006, 0.041 |
| ≥3 | −0.031, 0.063 | −0.040, 0.039 | 0.036, 0.030 | −0.027, 0.025 | 0.003, 0.024 | −0.061, 0.056 | 0.015, 0.038 | 0.019, 0.042 | 0.001, 0.027 | −0.015, 0.024 |
| Number of Alcohol Sale Establishments Within 1,000 ft of Intersection | | | | | | | | | | |
| 0 | Base Level | | | | | | | | | |
| 1–8 | −0.027, 0.064 | −0.024, 0.046 | −0.092, 0.036 | −0.005, 0.030 | −0.006, 0.029 | −0.074, 0.066 | −0.025, 0.047 | −0.053, 0.051 | 0.007, 0.032 | −0.010, 0.030 |
| Maximum Crossing Lanes | | | | | | | | | | |
| 1 | Base Level | | | | | | | | | |
| 2 | 0.510, 0.652 | 1.068, 0.609 | −0.042, 0.323 | 0.268, 0.527 | 0.840, 0.419 | 0.291, 0.067 | 1.272, 0.704 | 0.082, 0.418 | 0.540, 0.331 | 0.487, 0.332 |
| 3 | 0.448, 0.654 | 1.209, 0.610 | 0.208, 0.324 | 0.479, 0.527 | 1.062, 0.420 | 0.575, 0.053 | 1.435, 0.704 | 0.368, 0.420 | 0.782, 0.332 | 0.722, 0.332 |
| 4 | 0.600, 0.652 | 1.621, 0.609 | 0.619, 0.323 | 0.889, 0.527 | 1.497, 0.419 | 0.875, 0.065 | 1.862, 0.703 | 0.730, 0.418 | 1.188, 0.331 | 1.146, 0.332 |
| 5 | 0.447, 0.654 | 1.878, 0.610 | 0.839, 0.324 | 1.126, 0.527 | 1.742, 0.420 | 1.241, 0.077 | 2.054, 0.704 | 1.005, 0.419 | 1.473, 0.331 | 1.403, 0.332 |
| 6 | 0.439, 0.656 | 2.240, 0.610 | 1.288, 0.325 | 1.523, 0.528 | 2.101, 0.420 | 1.354, 0.116 | 2.527, 0.705 | 1.375, 0.421 | 1.815, 0.332 | 1.774, 0.332 |
| 7 | 0.587, 0.663 | 2.349, 0.613 | 1.306, 0.327 | 1.624, 0.529 | 2.261, 0.421 | 0.291, 0.067 | 2.596, 0.707 | 1.548, 0.425 | 2.011, 0.334 | 1.935, 0.334 |

*Variables insignificant at a 90-percent confidence level.

[a]Parameter estimate.

[b]Standard error.

**Table 33. Revised Wald test statistics on NB model parameter estimates (relative to dataset 10).**

| Variables | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Dataset 5 | Dataset 6 | Dataset 7 | Dataset 8 | Dataset 9 |
|---|---|---|---|---|---|---|---|---|---|
| Ln (AADT Major) | 0.165 | 0.029 | 1.030 | 0.043 | 0.940 | 0.195 | 1.143 | 1.388 | 0.555 |
| Ln (AADT Minor) | 3.051 | 0.065 | 1.910 | 0.561 | 0.962 | 1.117 | 1.853 | 1.274 | 0.787 |
| LTLs | | | | | | | | | |
| 1 | Base Level | | | | | | | | |
| 2 | 3.251 | 1.303 | 2.170 | 2.011 | 0.290 | 0.400 | 0.840 | 1.644 | 0.134 |
| 3 | 3.602 | 0.231 | 2.500 | 1.910 | 0.546 | 0.841 | 0.811 | 0.466 | 0.567 |
| 4 | 2.018 | 0.829 | 1.018 | 2.232 | 0.395 | 0.955 | 0.915 | 0.358 | 0.126 |
| RTLs | | | | | | | | | |
| 1 | Base Level | | | | | | | | |
| 2 | 0.501 | 0.415 | 0.041 | 0.137 | 1.709 | 0.250 | 1.432 | 1.611 | 0.494 |
| 3 | 3.594 | 0.994 | 0.779 | 0.012 | 0.903 | 0.746 | 1.801 | 1.002 | 1.100 |
| 4 | 0.721 | 1.205 | 1.108 | 0.113 | 0.128 | 0.097 | 0.614 | 0.110 | 0.273 |
| Speed Limit | | | | | | | | | |
| 35 mph | Base Level | | | | | | | | |
| 25 mph | 1.521 | 3.127 | 1.527 | 0.878 | 0.922 | 0.958 | 0.351 | 0.033 | 0.704 |
| 30 mph | 0.772 | 1.590 | 2.525 | 0.421 | 1.743 | 0.767 | 0.561 | 0.301 | 0.258 |
| 40 mph | 1.252 | 1.827 | 2.027 | 1.024 | 0.792 | 0.670 | 1.780 | 0.003 | 0.075 |
| 45 mph | 2.041 | 0.718 | 0.446 | 0.288 | 0.409 | 0.460 | 0.688 | 1.610 | 0.997 |
| 50 mph | 1.525 | 0.189 | 0.462 | 0.753 | 0.430 | 1.830 | 1.272 | 0.150 | 1.438 |
| 55 mph | 0.411 | 0.029 | 1.217 | 0.184 | 1.389 | 1.334 | 0.058 | 0.100 | 0.227 |
| 65 mph | 0.258 | 0.921 | 0.282 | 1.312 | 0.856 | 0.126 | 0.436 | 0.992 | 1.353 |
| LTLs Permitted | | | | | | | | | |
| 1 | Base Level | | | | | | | | |
| 2 | 0.899 | 0.557 | 1.107 | 0.600 | 1.058 | 0.162 | 1.488 | 0.629 | 0.393 |
| 3 | 2.131 | 0.077 | 1.046 | 1.250 | 0.619 | 0.888 | 1.449 | 0.286 | 0.073 |
| 4 | 0.658 | 0.288 | 1.378 | 1.230 | 1.159 | 2.389 | 0.830 | 0.410 | 0.711 |
| LTLs Protected | | | | | | | | | |
| 1 | Base Level | | | | | | | | |
| 2 | 0.835 | 0.787 | 0.787 | 0.659 | 1.112 | 0.175 | 0.383 | 0.844 | 0.115 |
| 3 | 0.124 | 0.300 | 0.613 | 0.063 | 0.853 | 0.597 | 0.141 | 1.268 | 0.702 |
| 4 | 0.617 | 0.755 | 0.047 | 1.215 | 0.722 | 3.466 | 0.980 | 2.734 | 0.700 |
| LTLs Mix | | | | | | | | | |
| 1 | Base Level | | | | | | | | |
| 2 | 0.323 | 0.682 | 1.206 | 1.808 | 0.589 | 0.690 | 0.829 | 0.109 | 0.705 |

| Variables | Dataset 1 | Dataset 2 | Dataset 3 | Dataset 4 | Dataset 5 | Dataset 6 | Dataset 7 | Dataset 8 | Dataset 9 |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 0.595 | 0.002 | 1.135 | 1.446 | 0.038 | 0.853 | 1.466 | 0.902 | 0.381 |
| 4 | 0.819 | 0.264 | 0.298 | 1.300 | 0.377 | 0.358 | 1.506 | 0.170 | 0.014 |
| RTOR | | | | | | | | | |
| 1 | Base Level | | | | | | | | |
| 2 | 0.309 | 0.905 | 0.359 | 3.205 | 1.725 | 0.044 | 0.383 | 0.560 | 0.957 |
| 3 | 0.845 | 0.686 | 0.338 | 0.111 | 0.201 | 1.306 | 0.091 | 0.218 | 0.354 |
| 4 | 1.312 | 0.787 | 0.068 | 2.147 | 1.722 | 0.403 | 1.209 | 1.423 | 1.117 |
| Lighting | | | | | | | | | |
| No | Base Level | | | | | | | | |
| Yes | 1.067 | 1.495 | 1.803 | 0.295 | 0.595 | 1.671 | 1.479 | 0.664 | 0.055 |
| Presence of School Within 1,000 ft of Intersection | | | | | | | | | |
| No | Base Level | | | | | | | | |
| Yes | 0.731 | 0.015 | 0.305 | 0.172 | 0.354 | 1.231 | 0.267 | 0.516 | 3.142 |
| Number of Bus Stops Within 1,000 ft of Intersection | | | | | | | | | |
| 0 | Base Level | | | | | | | | |
| 1–2 | 1.687 | 1.389 | 0.234 | 1.428 | 0.409 | 0.964 | 0.602 | 0.346 | 0.520 |
| ≥3 | 0.242 | 0.549 | 1.328 | 0.343 | 0.528 | 0.848 | 0.666 | 0.699 | 0.447 |
| Number of Alcohol Sale Establishments Within 1,000 ft of Intersection | | | | | | | | | |
| 0 | Base Level | | | | | | | | |
| 1–8 | 0.205 | 0.264 | 1.778 | 0.108 | 0.102 | 1.329 | 0.720 | 0.727 | 0.385 |
| Maximum Crossing Lanes | | | | | | | | | |
| 1 | Base Level | | | | | | | | |
| 2 | 0.032 | 0.837 | 1.142 | 0.351 | 0.660 | 0.580 | 1.009 | 0.758 | 0.114 |
| 3 | 0.372 | 0.702 | 1.108 | 0.389 | 0.635 | 0.438 | 0.916 | 0.661 | 0.128 |
| 4 | 0.746 | 0.684 | 1.139 | 0.413 | 0.657 | 0.802 | 0.920 | 0.781 | 0.088 |
| 5 | 1.303 | 0.684 | 1.218 | 0.443 | 0.633 | 0.477 | 0.836 | 0.745 | 0.149 |
| 6 | 1.815 | 0.671 | 1.047 | 0.401 | 0.611 | 1.192 | 0.966 | 0.743 | 0.088 |
| 7 | 1.814 | 0.593 | 1.344 | 0.495 | 0.606 | 5.402 | 0.845 | 0.714 | 0.161 |
| Overall Percent | 71.4 | 92.8 | 71.42 | 78.57 | 92.85 | 71.4 | 100 | 92.85 | 92.85 |

From table 33, the parameter for major and minor AADT presents a t-statistic lower than the 90-percent confidence value of 1.65 for all instances except for dataset 1 for the minor AADT, perhaps because the parameter was barely significant in the prediction model itself. A lower confidence value of 1.65 for the AADTs indicates no significant differences between the AADTs across the datasets. The parameter for "left-turn lane" presents a range higher than the 90-percent confidence value of 1.65, which is also not surprising, given the variable was only marginally significant. The parameter for "right-turn lane" for all models has a range lower than the confidence value of 1.65. The speed limit parameter, with levels from 25 to 65 mph, had a range lower than the confidence value of 1.65, except for 45 mph in dataset 1. Overall, out of the 14 variables included in the models, 71.4 percent of the variable values in the dataset 1 fell into the range for the tested confidence value. For dataset 2, 92.8 percent of the variables fell within the range of significant value. The exception was "25-mph speed limit," which presented a value higher than the 90-percent confidence value of 1.65. This conjecture applies to the other datasets used for this case study.

Overall, most variables had test statistics lower than the confidence value of 1.65, which indicated there were no significant differences between this dataset and the corresponding dataset (10) that was used for testing the stability of this study's data.

**Discussion of Findings**

***Segments***

The case study generated various crash model predictions, using 10 different datasets for rural, two-lane, undivided segment roads. The significant variables are intuitive relative to other studies. Less variables were significant when the sample size was small. Poisson Regression had a lower MSPE and MAD for K crashes. Poisson regression had the lowest MAD and NB had the lowest MSPE for A crashes, which meant that NB performed better. Univariate Poisson lognormal had the lowest MAD for B crashes while Poisson regression had the lowest for MSPE. Poisson regression performed better for C crashes for MAD while NB did better for MSPE. Finally, Poisson did better for MAD for PDO crashes while univariate Poisson lognormal did better for MSPE.

The research team used cross validation to assess the predictive performance of the models and judge how they performed outside the sample to a new dataset also known as test data. Without cross validation, there is only information on how the model performs relative to in-sample data. Ideally, it is desirable to see the model prediction accuracy on new data. The results from the cross-validation exercise showed that the dataset was robust enough to be relied on. Finally, the t-statistic estimates showed that the differences among the parameter value across the dataset are within a statistically acceptable level. The test statistics across the datasets for the AADT parameter were also lower than the 90-percent confidence value of 1.65 indicating that the variation across the different datasets is within a statistically acceptable level.

***Intersections***

The case study presents the crash prediction models for urban four-leg signalized intersections. Univariate Poisson lognormal performed better in the prediction of MSPE for K crashes while

Poisson regression did better for MAD for K crashes. NB had the lowest MAD and MSPE for A crashes which meant that it performed better. Univariate Poisson lognormal had the lowest MAD for B crashes while Poisson regression had the lowest for MSPE. Poisson regression performed better for C crashes for MAD while NB did better for MSPE. Finally, Poisson did better for MAD for PDO crashes while univariate Poisson lognormal did better for MSPE.

The results from the cross-validation exercise show that the dataset was robust enough to be considered reliable. Finally, the t-statistic estimates show that the differences among the parameter values across the dataset are within a statistically acceptable level.

### *Summary and Conclusions*

The purpose of this case study was to demonstrate the usefulness of the RAD tool, how to apply it, and establish that it could be relied upon to generate realistic data. Hence, the team developed various prediction models for different severity levels. Revised Wald Test Statistics were conducted to check if the variation across the different datasets is within a statistically acceptable level using dataset 10 as the benchmark. The result clearly highlights the stability in various parameter estimates across the datasets. The resulting stability found across the datasets indicates that the parameter estimates using RAD will be consistent, regardless of the miles of segment-related data generated using the tool. With this knowledge, the dataset from the RAD tool can be used for other possible purposes, such as estimating other possible prediction models, comparing the performance of varied safety analysis methods, and helping to determine an adequate sample size to get convincing results, especially for low realizations of fatal crashes. RAD can also be helpful in generating large datasets with consistent conditions in cases where it is not possible to go back many years due to changes in roadway characteristics or drivers. This makes it easier to estimate models for unusual and rare events, such as minor crashes.

The results from this case study may have significant implications on highway safety research in the development of information that can be used to make roads safer and crashes less severe. Since the data-generation process in the RAD tool is completely known, its use will allow objective evaluation and validation of various safety analysis methods used to verify various assumptions related to safety performance. In turn, these capabilities may help transportation agencies produce effective countermeasures to prevent and address crashes.

## DRIVING SIMULATION CASE STUDY

### Introduction

One of the most significant socioeconomic issues facing the world today is the frequency of traffic crashes. According to a recent report by NHTSA, 38,824 people in the United States died in car crashes in 2020 (National Center for Statistics and Analysis 2021). Rear-end collisions accounted for almost one-third of these crashes. Driver inattention was a major causal factor in about 91 percent of rear-end crashes. This inattention may prevent a driver from detecting an object ahead and can be caused by distraction, fatigue, or atmospheric conditions (such as fog or sun glare). Several strategies exist to aid drivers in avoiding collisions (National Transportation Safety Board 2002). These strategies differ in terms of degree of intervention—from alerts that suggest subtle speed adjustments to automatic emergency braking.

With the emergence of these strategies in the future, vehicles with these various safety systems installed are likely to have fewer total crashes. One such strategy is the FCW system. This system is designed to help drivers reduce the severity of collisions or avoid them, especially rear-end crashes, with visual, auditory, or tactile warnings of possibly impending collisions (Kusano and Gabler 2012). Additionally, a 2002 Daimler-Chrysler study found that 60 percent of rear-end collisions could likely be avoided if drivers had 0.5 additional seconds of warning, and 90 percent of rear-end collisions could likely be avoided if drivers had an extra second of warning (NHTSA 2002). This research places FCW high on the list of solutions that can contribute significantly to reducing crash numbers and severity.

The objective of this case study was to quantify the observed number of serious conflicts for drivers in vehicles equipped with an FCW system versus those in vehicles without these systems using driving simulator experiments. For this case study, driving simulation scenarios were generated that exposed participating drivers to road and traffic conditions deliberately designed to test their responses to unexpected forward obstacles with and without warnings from an FCW system. This case study differs from a previous study by Lee et al. (2002), which focused on the value of an FCW system in a scenario where the driver was intentionally distracted as part of the experiment and the hazard was a stopped lead vehicle. This case study explores the value of the FCW system for drivers who may or may not be distracted; the hazard may be a vehicle that stops or crosses the travel path suddenly or a nonmotorized user crossing the street unexpectedly.

## Literature Review

### *Studies on FCW Systems*

FCW systems have been estimated to potentially reduce front to rear-end crash rates by 27 percent and front to rear injury crash rates by 20 percent (Cicchino 2017). Although these systems do not prevent every crash, they have been proven to significantly mitigate the likelihood of a crash. Lee et al. (2002) presented users of a driving simulator with a scenario involving a stopped lead vehicle in which the driver was intentionally distracted and found that the FCW system reduced the number of collisions for that scenario by 80.7 percent. Teoh (2021) estimated the effectiveness of FCW systems using detailed data from exposure measures extracted from video footage and concluded that FCW systems were associated with a statistically significant 22-percent reduction in the rate of police-reportable crashes per vehicle miles traveled.

A common human factor discussion surrounding FCW concerns the criteria that determine the guidance on warning timing. McLaughlin, Hankey, and Dingus (2008) noted that early warning has great potential to prevent crashes. Abe and Richardson (2006) found that early alarm timing may improve driver's trust compared with late alarm timing. In addition, other studies investigated how various driving situations and hazard warning experience can influence driver responses to warning and hazardous situations. Incorrect warnings and failures of the system to produce warnings that drivers find useful and understandable may diminish drivers' reliance on and compliance with warnings (Glassco and Cohen 2001; Lee and Lee 2007; Najim and Smith 2004; Reinmueller and Steinhauser 2019).

A considerable amount of literature describes the possible safety benefits of these advanced technologies. Most of the literature the research team reviewed focused on the market penetration rate (MPR) of these systems. For example, Xiao et al. (2021) carried out a meta-analysis evaluating crash reduction by penetration rate. The results indicated that the number of conflicts is exponentially reduced as MPR goes up; safety is enhanced by 4 percent with an MPR of 10 percent and by 43 percent with 90 percent MPR. Furthermore, Papadoulis, Quddus, and Imprialou (2019) evaluated the safety impact of connected vehicles (CVs) using traffic simulation. The results showed that CVs bring about compelling benefits to road safety, as traffic conflicts significantly reduce even at relatively low market penetration rates. Specifically, estimated traffic conflicts were reduced from 12 to 47 percent, 50 to 80 percent, 82 to 92 percent, and 90 to 94 percent for 25-percent, 50-percent, 75-percent, and 100-percent CV penetration rates, respectively.

Based on 37 precrash scenarios developed by Najm et al. (2010), Jermakian (2011) estimated the maximum potential for U.S. crash reductions for four crash avoidance technologies: side view assistance, FCW, lane departure warning, and adaptive headlights. Jermakian (2011) estimated that FCW holds the greatest potential for preventing crashes of any severity, up to 1.2 million crashes per year in the United States, or 20 percent of the annual 5.8 million police-reported crashes. Kusano and Gabler (2012) examined the safety benefits of FCW systems in rear-end collisions simulating scenarios from real world rear-end crash data extracted from the National Automotive Sampling System/Crashworthiness Data System. Their study indicated a dramatic reduction in serious and fatal injuries when using these safety features.

### *Studies on Crashes and Serious Conflicts*

Surrogate safety assessment is an alternative method of assessing safety that relies on the analysis of safety-critical events known as traffic conflicts. Definitions of traffic conflict vary; the first mention of the term was by Klebelsberg (1964), as cited by Tarko (2018), who defined traffic conflicts as dangerous traffic interactions. Laureshyn and Varhelyi (2020) define a traffic conflict as "an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged." The proximity of road users to each other can be physically measured in temporal and/or spatial dimensions, and thresholds can be used to identify conflicts. Perkins and Harris (1968) used traffic conflicts to define situations necessitating evasive actions, such as braking. According to this definition, conflicts and crashes are of a similar nature but for the presence and success of an evasive action (Zheng, Ismail, and Meng 2014).

Various conflict indicators have been developed to measure the severity of an interaction by quantifying the spatial and temporal proximity of two or more road users. A comprehensive summary of the different indicators is provided in Brown (1994) and Tarko (2009). Typical measures of conflict severity include conflicting speed and severity index (Autey, Sayed, and Zaki 2012; Essa and Sayed 2020). The time-to-collision indicator used in this study has been widely used to measure the severity of traffic conflicts. Time to collision is defined as "…the time that remains until a collision between two vehicles would have occurred if the collision course and speed difference are maintained" (Laureshyn and Varhelyi 2020). The main advantage of conflict indicators is their ability to capture the severity of an interaction in an objective and quantitative way. Various studies have used traffic conflicts together with crashes

for safety analysis and have shown that conflicts can adequately serve as a proxy for crashes (Laureshyn and Varhelyi (2020) ; Charly and Mathew 2019; Peesapati, Hunter, and Rodgers 2013; Xie et al. 2016).

Although previous studies have contributed to examining the effectiveness of FCW systems, none of these studies focused on estimating the relationship between crashes and serious conflict with the aim of providing information for crash modification factors for the FCW systems. In this case study, near crashes, also known as serious conflicts, were used as surrogates for crashes to represent quantitative safety. This substitution was made because researchers expected crashes to be extremely rare, given the number of drivers to be observed in the experiment, and because identifying crashes in the driving simulation environment would be challenging.

## Methodology

This section outlines the general description of methods used for the entire experiment and describes the participants used for the experiment, followed by the description of the Connecticut Transportation Safety Research Center driving simulator used for the main experiments. The driving scenarios used for the experiment are also described in detail. Finally, there is a discussion regarding the traffic conflict technique the research team used to evaluate traffic conflicts.

### *Participants*

A total of 142 participants—64 females, 77 males, and 1 person who declined to identify as either male or female—were recruited for the study. Simulator sickness prevented nine people from completing the experiment. Of the 133 participants who completed the experiment, 58 (41 percent) were between the ages of 18 and 29, 55 (41 percent) were between the ages of 30 and 64, and 20 (15 percent) were 65 or older. Each participant had a valid driver's license and was in good physical health. They were recruited via flyers distributed on college bulletin boards, in accordance with Institutional Review Board regulations. The college bulletin boards they were posted to belonged to University of Connecticut (UConn). Flyers were also distributed at senior and community centers to increase the participation of senior drivers and area residents not associated with UConn.

### *Equipment*

The experiment made use of a driving simulator, made by Technology Company A, which had a full cab and a segmented screen. Additional equipment included a series of cameras, four projectors (rear, front, right-side, and left-side projections), and various screens to provide a high-fidelity virtual environment. All the equipment was owned and maintained by the Connecticut Transportation Safety Research Center located at UConn. Some features of the driving simulator include the Internet Scene Assembler ®, which is used in the modification of the virtual environment, and the SimObserver®, which is integrated with the virtual environment and used for data and video synchronization, video capture, and after-action review (Realtime Technologies 2023). Figure 20 shows the UConn driving simulator.

Source: FHWA.

**Figure 20. Photo. UConn driving simulator vehicle.**

In figure 21, the computer on the left side was used to run the simulator and edit the model. The computer on the right side was used to extract data and video clips of the main driving experiment.

Source: FHWA.

**Figure 21. Photo. UConn driving simulator control center.**

The software programs used in the driving simulator were SimCreator®, SimObserver®, SimCreator DX®, and Data Distillery® (Realtime Technologies 2023). Additionally, audio software and hardware were used to simulate engine sounds, tire sounds, and vehicle noises for the participants.

SimCreator® is used for graphical simulation and as a modeling system. Different components are connected with each other to make a model. Each component can either be a group of different components or a C/C++ code component. Once a model has been developed, it can be simulated, and the result plotted (Realtime Technologies 2023).

SimObserver® is a data collection system designed to capture data for after-action review that is controlled by the SimObserver computer. It procures the following output files (Realtime Technologies 2023):

- Video file (MPEG-2 format): The format of the file is ".mpg file." It contains a video of the entire experiment. The video is used to determine the length of road over which various tasks are performed.

- Log file: The format of this file is ".log file." It includes system messages, errors, and warnings.

- Event file: The format of this file is ".vt file." It contains start and stop times and labels all events logged during video capture.

- DAT file: The format of this file is ".dat file" (in the form of a video on DVD). It is the main data file used to measure driving performance. Data are categorized into different columns. Some of the variables in the DAT file are longitudinal acceleration, lateral acceleration, throttle, and headway distance.

SimCreatorDX® is the graphical user interface used to create, monitor, and control scenarios. It allows the researcher to observe the current state of the SimCreator simulation through graphical displays and detailed data view and is intended to be the primary tool used for development, tuning, and creation of experimental scenarios. After launching SimCreatorDX, the user has the option of either developer or experimental mode. In developer mode, the purpose is to design scenarios to fit the purpose of the study. Experimental mode allows for a quick interface to run multiple participants for any studies (Realtime Technologies 2023).

Data Distillery® is a data review and reduction software package with a main purpose of improving the efficiency of data reduction. This tool is also used for compiling the captured video and data from SimObserver. Data Distillery provides fine details of the collected data to understand the nature of the behavior or system being observed. The log file, video file, and data file are all displayed on the same screen. It also can be utilized to find the position of the vehicle (Realtime Technologies 2023).

### *Procedure and Experimental Design*

Each participant was given an informed consent document to read and sign upon arrival on the day of the study. This consent included permission to record video. Each participant then completed a questionnaire that included demographic questions, such as their sex, age, race/ethnicity, and years of driving experience. Researchers then asked participants to complete an approximately 5-min driving simulator training sequence to acclimatize them to driving in the simulator. If participants requested more time, an additional 5-min training was provided.

After the training sequence, each participant was placed into one of two groups based on the answers to the questionnaire. This allowed researchers to balance the distribution of age, sex, and years of driving experience within each group. The race/ethnicity distribution of each group was also monitored to keep those factors as balanced as possible. Both groups completed a defined driving scenario that included a mix of city and highway driving. The scenario took about 15 min to complete. Participants were asked to drive the course as if they were driving a real car on a real road. They were told that the purpose of the experiment was to observe their reactions to the scenario they drove. At some point during the scenario, a stimulus was introduced that required the driver to execute a sudden, unexpected braking maneuver. For example, a parked vehicle would suddenly pull in front of the vehicle without warning, or another vehicle, animal, or pedestrian would cross the vehicle path.

Half the participants drove the course with a FCW system programmed into the simulation, giving the participant an audible and visual warning on the dashboard about the imminent collision. The other half did not have the FCW system and received no warning. Before the

driving began, participants with the FCW system were informed that the vehicle they were driving provided an audible sound and/or flashing pedestrian on the dashboard to warn of imminent forward collision.

After they completed their turns in the driving simulator, participants were debriefed by a researcher and given a copy of the UCONN debriefing form for the case study, titled "Estimating a Crash Modification Factor for Forward Collision Warning Systems in the Vehicle Fleet." The debriefing form included contact info for the principal investigator and student researcher and listed the study sponsor as FHWA. Additionally, it included the following statement about the study:

> The purpose of this study is to find out how much of a reduction in accidents is possible when vehicles are equipped with an automatic FCW system. Participants were placed in one of two groups: one that included the automatic forward collision avoidance system and one that did not. At one point during the experiment, you were faced with a sudden, unexpected obstacle that required you to stop. We noted the vehicle speed and the distance from the obstacle when you took evasive action and will use them to calculate the severity of the resulting conflict. Once we are finished running experiments, we will compare the conflict severities of the drivers who had the FCW system with those who did not to estimate the safety value of having this system installed.

Additionally, the researcher gave a verbal description of the study and its aims. As the experiment did not involve deception, this debriefing was only informational.

### *Driving Scenarios*

Driving scenarios were created for both urban and rural settings. Researchers programmed pedestrians walking along sidewalks into each scenario to help the scenes be as realistic as possible. Participants were asked to drive as they normally would on their way to work or college but to always stay in the right lane. These instructions allowed for the unexpected event to occur in the right lane. (Although it would be most ideal to run this experiment with no same-way lane restrictions on drivers, programming the driving simulator to accommodate this change was not possible due to constraints in the technology available. Due to its uniqueness, this case study is still extremely valuable for the insights it can provide.)

The urban setting consisted of an undivided roadway section with two lanes of traffic in both directions. The roadway included all the typical features of an urban road, such as speed limit signs and traffic signals, and had paved curbs on both sides of the road. The roadside features included restaurants and gas stations placed along the route with high-rise buildings. Most sections of the road were straight, but there were a few right-angled turns, as shown in figure 22.

Source: FHWA.

**Figure 22. Photo. Section of urban roadway used in experiment.**

The rural scenario consisted of a two-lane undivided roadway section with traffic simulated into the experiment to make it feel like real-world driving. The rural setting was a mix of countryside and residential areas, as shown in figure 23. Most sections of the road were straight, with a few horizontal curves.

Source: FHWA.

**Figure 23. Photo. Section of rural roadway used in experiment.**

During the experiment, participants were asked to follow the designated route as a recorded message that was broadcast through speakers in the car provided directions. Four scenarios were modeled, as follows:

- Scenario 1: Urban setting plus entering car. A parked car suddenly pulls over into the path of the driver.
- Scenario 2: Urban setting plus pedestrian. A person suddenly crosses the road into the path of the driver.
- Scenario 3: Rural setting plus animal. A deer crosses into the path of the driver.
- Scenario 4: Rural setting plus pedestrian. A small child crosses the road unexpectedly.

The demographics of participants in each scenario are demonstrated in table 34. Care was taken to ensure a balance for the age categories by scenario.

**Table 34. Driving simulator participant statistics by scenario.**

| Scenario | Male 18−29 | Male 30–64 | Male 65+ | Female 18−29 | Female 30−64 | Female 65+ | Total |
|---|---|---|---|---|---|---|---|
| Urban plus entering car | 4[a], 4[b] | 4, 5 | 1, 1 | 2, 3 | 3, 1 | 1, 1 | 15,15 |
| Urban plus pedestrian | 4 ,4 | 4, 3 | 0, 1 | 4, 3 | 5, 5 | 2, 3 | 19, 19 |
| Rural plus animal | 2, 2 | 2, 3 | 3, 2 | 4, 5 | 4, 3 | 2, 2 | 17, 17 |
| Rural plus pedestrian | 4, 4 | 3, 3 | 1, 0 | 4, 4 | 2, 3 | 1, 1 | 15, 15 |
| Total | 14, 14 | 13, 14 | 5, 4 | 14, 15 | 14, 12 | 6, 7 | 66, 66 |

[a]With FCW.
[b]Without FCW.

### *Traffic Conflict Methodology*

To consistently use conflict techniques, many countries have formed their own standards and published manuals or handbooks to guide field observations. Some examples include Swedish Traffic Conflict Technique (STCT), U.S. Traffic Conflict Technique (USTCT), Dutch Traffic Conflict Technique (DOCTOR), and the German Traffic Conflict Technique. Both the STCT and USTCT use the time to accident (TA) and conflicting speed (CS) values as a conflict severity indicator; meanwhile, DOCTOR is more concerned with driver error and incorporates it into a safety continuum (Zheng, Ismail, and Meng 2014).

The STCT is used in this effort to evaluate traffic conflicts. Traffic conflicts are typically categorized based on two indicators: the TA value and the CS. The TA is the time between when a particular road user successfully performs an evasive action and when the collision would have occurred. The CS is the speed of the roadway user at the point of taking the evasive action. The TA and CS values are used to define the conflict severity. The CS affects the outcome of the collision (the resulting pedestrian injury depends on the speed), and a higher speed requires longer stopping time and distance. Thus, a higher CS indicates a more severe conflict for the same TA (Laureshyn and Varhelyi 2020). Directly estimating the TA in live traffic can be challenging, so TA can instead be estimated from the vehicle speed and the distance to the collision point using table 35**.** The conflict severity is then determined from the TA and the CS, according to the graph in figure 24.

**Table 35. Chart. TA values estimated from vehicle speed and distance to collision point (Laureshyn and Varhelyi 2020).**

| Speed | | Distance, m | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| km/h | m/s | 0.5 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 |
| 5 | 1.4 | 0.4 | 0.7 | 1.4 | 2.2 | 2.9 | 3.6 | 4.3 | 5.0 | 5.8 | 6.5 | 7.2 | — | — | — | — | — | — | — | — | — |
| 10 | 2,8 | 0.2 | 0.4 | 0.7 | 1.1 | 1.4 | 1.8 | 2.2 | 2.5 | 2.9 | 3.2 | 3.6 | 5.4 | 7.2 | 9.0 | — | — | — | — | — | — |
| 15 | 4.2 | 0.1 | 0.2 | 0.5 | 0.7 | 1.0 | 1.2 | 1.4 | 1.7 | 1.9 | 2.2 | 2.4 | 3.6 | 4.8 | 6.0 | 7.2 | 8.4 | 9.6 | — | — | — |
| 20 | 5.6 | 0.1 | 0.2 | 0.4 | 0.5 | 0.7 | 0.9 | 1.1 | 1.3 | 1.4 | 1.6 | 1.8 | 2.7 | 3.6 | 4.5 | 5.4 | 6.3 | 7.2 | 8.1 | 9.0 | 9.9 |
| 25 | 6.9 | 0.1 | 0.1 | 0.3 | 0.4 | 0.6 | 0.7 | 0.9 | 1.0 | 1.2 | 1.3 | 1.4 | 2.2 | 2.9 | 3.6 | 4.3 | 5.0 | 5.8 | 6.5 | 7.2 | 7.9 |
| 30 | 8.3 | 0.1 | 0.1 | 0.2 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 1.0 | 1.1 | 1.2 | 1.8 | 2.4 | 3.0 | 3.6 | 4.2 | 4.8 | 5.4 | 6.0 | 6.6 |
| 35 | 9.7 | 0.1 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.1 | 2.6 | 3.1 | 3.6 | 4.1 | 4.6 | 5.1 | 5.7 |
| 40 | 11.1 | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.4 | 1.8 | 2.3 | 2.7 | 3.2 | 3.6 | 4.1 | 4.5 | 5.0 |
| 45 | 12.5 | — | 0.1 | 0.2 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.6 | 0.7 | 0.8 | 1.2 | 1.6 | 2.0 | 2.4 | 2.8 | 3.2 | 3.6 | 4.0 | 4.4 |
| 50 | 13.9 | — | 0.1 | 0.1 | 0.2 | 0.3 | 0.4 | 0.4 | 0.5 | 0.6 | 0.6 | 0.7 | 1.1 | 1.4 | 1.8 | 2.2 | 2.5 | 2.9 | 3.2 | 3.6 | 4.0 |
| 55 | 15.3 | — | 0.1 | 0.1 | 0.2 | 0.3 | 0.3 | 0.4 | 0.5 | 0.5 | 0.6 | 0.7 | 1.0 | 1.3 | 1.6 | 2.0 | 2.3 | 2.6 | 2.9 | 3.3 | 3.6 |
| 60 | 16.7 | — | 0.1 | 0.1 | 0.2 | 0.2 | 0.3 | 0.4 | 0.4 | 0.5 | 05 | 0.6 | 0.9 | 1.2 | 1.5 | 1.8 | 2.1 | 2.4 | 2.7 | 3.0 | 3.3 |
| 65 | 18.1 | — | 0.1 | 0.1 | 0,2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 | 0.5 | 0.6 | 0.8 | 1.1 | 1.4 | 1.7 | 1.9 | 2.2 | 2.5 | 2.8 | 3.0 |
| 70 | 19.4 | — | 0.1 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 | 0.5 | 0.5 | 0.8 | 1.0 | 1.3 | 1.5 | 1.8 | 2.1 | 2.3 | 2.6 | 2.8 |
| 75 | 20.8 | — | 0.0 | 0.1 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 | 0.5 | 0.7 | 1.0 | 1.2 | 1.4 | 1.7 | 1.9 | 2.2 | 2.4 | 2.6 |
| 80 | 22.2 | — | 0.0 | 0.1 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 | 0.5 | 0.7 | 0.9 | 1.1 | 1.4 | 1.6 | 1.8 | 2.0 | 2.3 | 2.5 |
| 85 | 23.6 | — | 0.0 | 0.1 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.3 | 0.4 | 0.4 | 0.6 | 0.8 | 1.1 | 1.3 | 1.5 | 1.7 | 1.9 | 2.1 | 2.3 |
| 90 | 25.0 | — | 0.0 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 | 1.4 | 1.6 | 1.8 | 2.0 | 2.2 |
| 95 | 26.4 | — | 0.0 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 | 0.3 | 0.3 | 0.3 | 0.4 | 0.6 | 0.8 | 0.9 | 1.1 | 1.3 | 1.5 | 1.7 | 1.9 | 2.1 |
| 100 | 27.8 | — | 0.0 | 0.1 | 0.1 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.3 | 0.4 | 0.5 | 0.7 | 0.9 | 1.1 | 1.3 | 1.4 | 1.6 | 1.8 | 2.0 |



—No data.

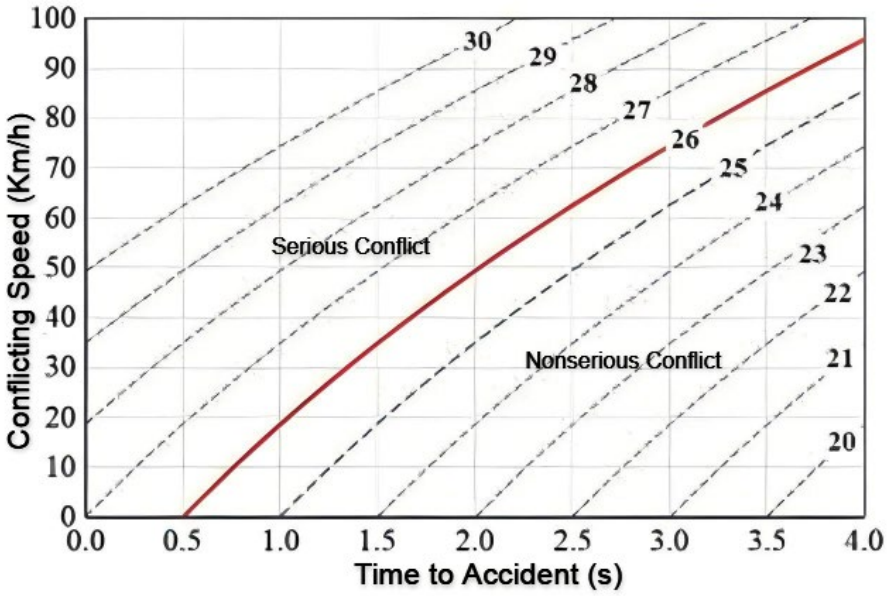





**Figure 24. Graph. Conflict severity diagram (Laureshyn and Varhelyi 2020).**

Conflicts with a severity level above 26 are ranked as serious. For instance, a vehicle approaching at 30 mph (48.3 km/hr) with a TA of 2.8 sec would pass the potential conflict point with too much time remaining for the interaction to be classified as a serious conflict. However, a 45 mph (72.4 km/hr.) approach speed would just exceed that level. These severity levels can be used to understand serious conflicts to estimate the crash modification factor (CMF) for FCW systems. A CMF is the ratio of the number of crashes expected with a countermeasure to the number of crashes expected without it. In other words, when implementing a countermeasure, multiplying the existing expected number of crashes by the CMF gives an estimate of the expected number of crashes after implementation of the countermeasure. CMFs with a value less than one indicate an expected decrease in crashes; meanwhile, a value greater than one indicates an expected increase. The FCW system can be said to be a countermeasure installed in vehicles to help mitigate the occurrence of crashes in vehicles that have it installed. A CMF for this "countermeasure" would have to also account for the penetration rate of the FCW system in the vehicle fleet. Estimating a CMF value for FCW systems would be helpful for adjusting the expected number of crashes in future conditions, as the proportion of vehicles with such systems increases.

The steps to estimate the CMF are as follows:

- Estimate a ratio to account for the safety effect of having a FCW system ($R_{FCW}$):

$R_{FCW} = n_{FCW} / n_o$, where $n_{FCW}$ is the number of serious conflicts in vehicles with the FCW, and $n_o$ is the number of serious conflicts in vehicles without the FCW.

- Define the proportion of the vehicle fleet with the FCW system ($P_{FCW}$): Five different proportions of the vehicle fleet from 10 to 50 percent in increments of 10 percent were investigated.

- Calculate the estimated CMF for the FCW system: $CMF_{FCW} = 1 + P_{FCW}(R_{FCW} - 1)$ **Error! Digit expected.**.

***Data Collection and Analysis***

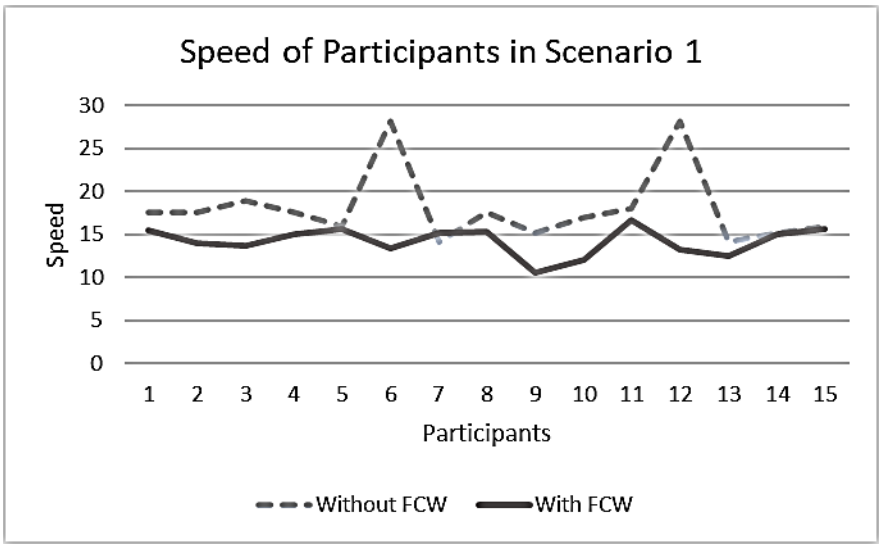
TA and CS are the two variables required to calculate conflict severity. Velocity is the output of the speed in the simulator and reflects the speed with which a participant drove. Lane position can be defined as the position of the vehicle measured from the center of the road, in meters and was used to represent the lateral control of the vehicle. A positive number indicates a vehicle on the right side of the center line. A negative number indicates a vehicle on the left. . Data collected on these two dependent variables were at a frequency of 60 Hz through the SimObserver® proprietary software of the driving simulator (Realtime Technologies 2023)**.**

The lateral position where the event occurred was collected for both the scenarios with and without the FCW system. Table 36 shows an example of the resulting extracted data from the simulator. The X and Y axes indicate the lateral position of each participant at the time the event was introduced into the experiment. The speed is the speed the participant was driving before carrying out the evasive action of either hitting the brakes or swerving.

**Table 36. Example of extracted spreadsheet data.**

| | Participants | X | Y | Speed (m/s) |
|---|---|---|---|---|
| 1 (without FCW) | FHWA01 | 3995.18 | -6895.13 | 17.56 |
| | FHWA03 | 2681.54 | -6895.13 | 17.56 |
| | FHWA 20 | 2694.1 | -6895.31 | 18.89 |
| | FHWA 28 | 2673.54 | -6896.02 | 17.61 |
| | FHWA 30 | 2672.73 | -6895.13 | 15.95 |
| | FHWA 40 | 2711.58 | -6895.38 | 28.23 |
| | FHWA 42 | 2674.42 | -6895.55 | 14.129 |
| | FHWA45 | 2667.56 | -6897.24 | 17.57 |
| | FHWA 51 | 2677.33 | -6895.81 | 15.192 |
| | FHWA 52 | 2686.56 | -6897.07 | 16.96 |
| | FHWA 95 | 2665.24 | -6895.48 | 18.04 |
| | FHWA125 | 2681.99 | -6895.33 | 28.23 |
| | FHWA126 | 2705.15 | -6895.73 | 14.129 |
| | FHWA127 | 2669.16 | -6895.67 | 15.129 |
| | FHWA128 | 2672.73 | -6895.13 | 15.5 |
| 1 (with FCW) | FHWA 09 | 2680.06 | -6895.81 | 13.95 |
| | FHWA 10 | 2678.98 | -6895.1 | 13.75 |
| | FHWA 15 | 2668.53 | -6895.37 | 15.07 |
| | FHWA 29 | 2676.24 | -6895.5 | 15.63 |
| | FHWA 31 | 2671.76 | -6895.96 | 13.44 |
| | FHWA 34 | 2687 | -6895.84 | 15.12 |
| | FHWA 68 | 2681.99 | -6895.33 | 15.38 |
| | FHWA 71 | 2669.16 | -6895.67 | 10.6 |
| | FHWA 74 | 2665.45 | -6895.66 | 12.01 |
| | FHWA 105 | 2667.92 | -6895.83 | 16.61 |
| | FHWA 112 | 2705.15 | -6895.73 | 13.29 |
| | FHWA 122 | 2660.42 | -6897.24 | 12.47 |
| | FHWA 123 | 2667.45 | -6896.09 | 15.07 |
| | FHWA 124 | 3995.18 | -6895.13 | 15.63 |
| | FHWA129 | 2711.58 | -6895.38 | 15.63 |

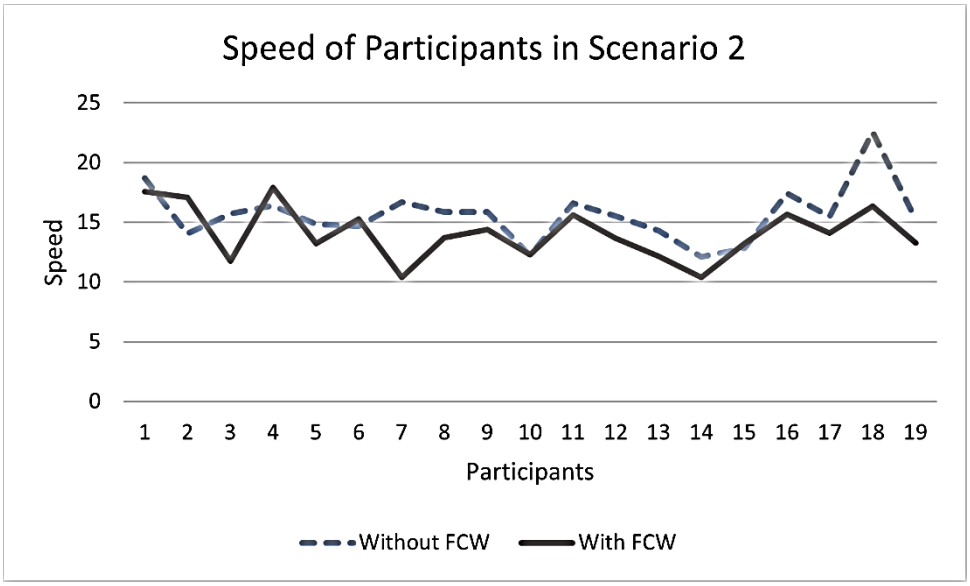
Thirty participants were in scenario 1, 38 in scenario 2, 34 in scenario 3, and 30 in scenario 4. The numbers of participants in each scenario was balanced, as noted in table 34. Figure 25 shows the speed of the participants with and without the FCW. When scenario 1 (a parked car suddenly drives into the path of the participant) occurred, the speed was lower with participants with the FCW—possibly because the warning sound made drivers immediately aware of the upcoming event.




Source: FHWA.

**Figure 25. Graph. Speed of participants in scenario 1.**

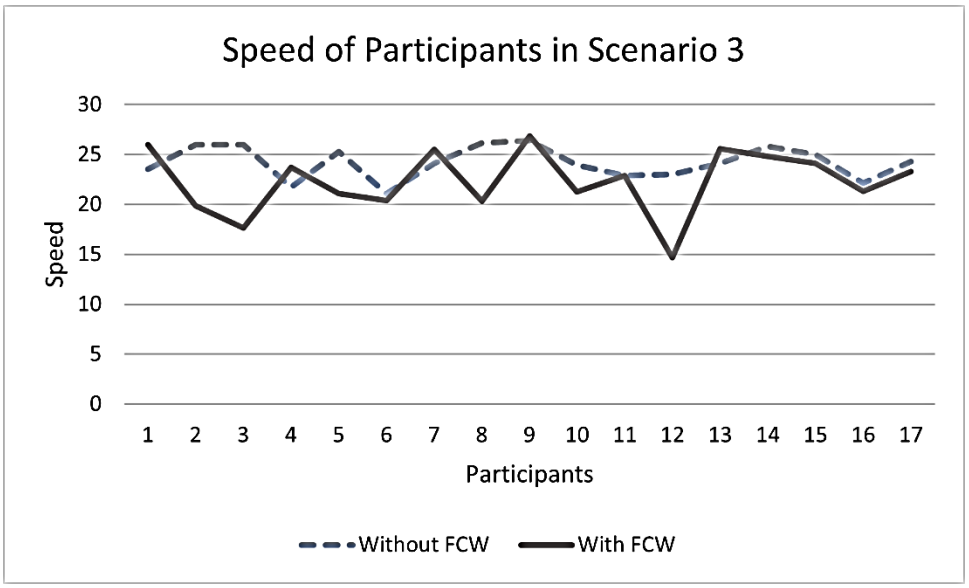
Figure 26 shows the speed of the participants with and without FCW when scenario 2 (a pedestrian crossing into the path of the participant) occurred. It shows that the speed was lower for participants with FCW. This outcome may be explained by the fact that the warning sound made drivers immediately aware of the upcoming event.




Source: FHWA.

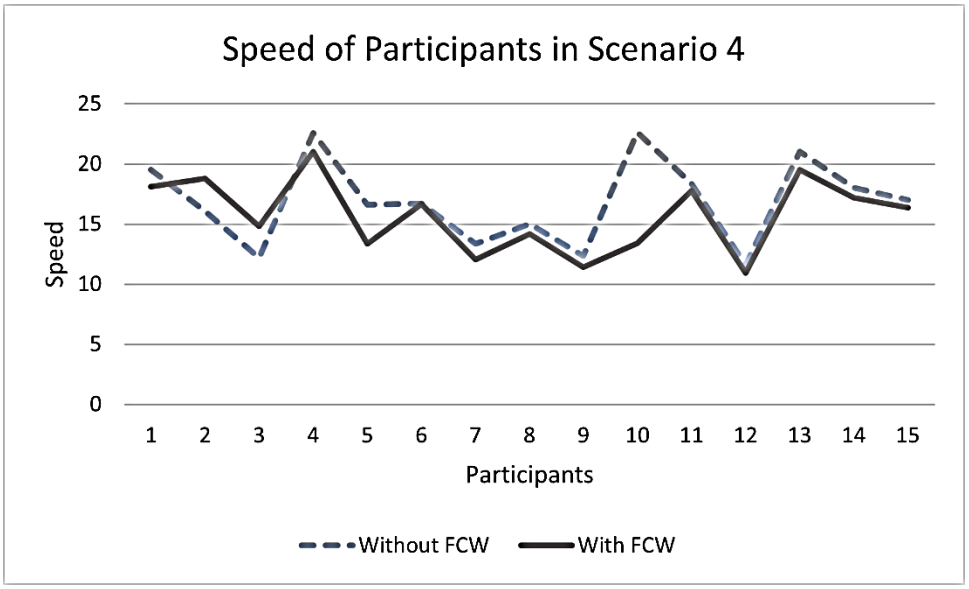
**Figure 26. Graph. Speed of participants in scenario 2.**

Figure 27 and figure 28 show a similar trend, as the speed remains greater for participants without the FCW.




Source: FHWA.

**Figure 27. Graph. Speed of participants in scenario 3.**




Source: FHWA.

**Figure 28. Graph. Speed of participants in scenario 4.**

### *Calculation of the Effectiveness of FCW*

The safety effect of having FCW was calculated by the ratio of serious conflict ($n_{FCW}$) in the experiment with FCW to the ratio of serious conflict without FCW, as discussed in the traffic conflict methodology section. Table 37 shows the raw factor for the different scenarios conducted in the experiment. The raw factor was also calculated for all cases with and without FCW to see if there were substantial differences between individual scenarios and overall cases.

**Table 37. FCW ratios by scenario.**

| Scenario | $n_{FCW}$ | $n_O$ | $R_{FCW}$ |
|---|---|---|---|
| All cases | 24 | 42 | 0.57 |
| Urban plus entering car | 6 | 9 | 0.67 |
| Urban plus pedestrian | 6 | 13 | 0.46 |
| Rural plus animal | 7 | 10 | 0.70 |
| Rural plus pedestrian | 5 | 10 | 0.50 |

The CMF was calculated based on the various penetration rates of the system and is shown in table 38.

**Table 38. CMF values by FCW fleet penetration rate.**

| Scenario | $P_{FCW}$= 0.1 | $P_{FCW}$= 0.2 | $P_{FCW}$= 0.3 | $P_{FCW}$= 0.4 | $P_{FCW}$= 0.5 |
|---|---|---|---|---|---|
| All cases | 0.95 | 0.91 | 0.87 | 0.82 | 0.78 |
| Urban plus entering car | 0.96 | 0.93 | 0.89 | 0.86 | 0.83 |
| Urban plus pedestrian | 0.94 | 0.89 | 0.83 | 0.78 | 0.73 |
| Rural plus animal | 0.97 | 0.94 | 0.91 | 0.88 | 0.85 |
| Rural plus pedestrian | 0.95 | 0.90 | 0.85 | 0.80 | 0.75 |

Considering all simulation scenarios, when the penetration rate is 10 percent, the CMF is 0.95, which translates to a 5-percent reduction in crashes. The CMF increases to 22 percent for a penetration rate of 50 percent. For scenario 1, when the FCW penetration rate is 10 percent, a 4-percent reduction in crashes is expected with a 17-percent reduction at a 50-percent penetration rate. In the case of scenario 2, a CMF of 0.94 for a 10-percent penetration rate means a 6-percent reduction in crashes, with a 27-percent reduction expected when the penetration rate of FCW is at 50-percent. Scenario 3 had a CMF of 0.97, which translates to a 3-percent reduction in crashes when the penetration rate is at 10 percent and 0.85, a 15-percent reduction when the penetration rate is at 50 percent. Scenario 4 had a CMF of 0.95, a 5-percent reduction in crashes when the penetration rate is at 10 percent and a CMF of 0.75 and a 25-percent reduction in crashes when the penetration is at 50 percent. Notably, as the overall penetration rate increases, the number of crashes decreases.

### *Driving Simulator Case Study Findings*

The team examined the effectiveness of FCW to provide information for crash modification factors for the FCW systems for five different FCW system market penetration rates for the vehicle fleet from 10 to 50 percent in increments of 10 percent. Four scenarios were modeled in urban and rural settings in which participants were exposed to conditions deliberately designed to test their responses to unexpected events when they got warnings from FCW systems or when there was no system. To achieve this, two variables, lane position and speed, were used to measure conflict severity. This X was used as a proxy for crashes because researchers expected crashes to be extremely rare, given the relatively small number of drivers observed in the experiment. The conflict severity was determined from the TA and CS values of the road users. A total of 133 participants completed the experiment.

The results suggest that FCW systems have the potential to reduce incidences of crashes in vehicles that have them installed, and reduction in incidences of crashes will occur as market penetration rate goes up. The overall CMF values for all the scenarios conducted in the experiment ranged from 0.78 to 0.95, which is approximately a 10 to 22 percent reduction in crashes, with an R value of 0.57. The CMF values for scenario 1 and 2 ranged from 0.83 to 0.96 and 0.73 to 0.94, which is approximately a 4 to 17 percent and 6 to 27 percent reduction in crashes, respectively. R values were 0.67 and 0.46. In the rural scenario, the CMF values for scenario 3 and 4 were from 0.85 to 0.97 and 0.75 to 0.95, with R values of 0.70 and 0.50, respectively. These data indicate that incidences of crashes were reduced by 3 to 15 percent for rural plus animal and 5 to 10 percent for rural plus pedestrian. The data also reveals that FCW enhances driver responses over a range of velocities, strongly suggesting that FCW can enhance driver responses in scenarios that may lead to collisions.

This study shows that FCW systems have enormous benefits, even though some unanswered questions remain. Specifically, although the warning enhanced driver's awareness of potential conflict in the simulator, warnings might be perceived as a nuisance in an actual driving situation when they are too sensitive or too early. Further data collection might be required to assess drivers' responses to nuisance alarms generated by early warning or if dynamic adjustment of the warning is feasible. Also, a longer-term follow-up study might be helpful in understanding changes in driver attitudes and validating benefits observed in this study. The use of naturalistic studies may provide more insight into driver behavior during routine trips. Data from such potential studies can help validate the benefits of FCW and assist researchers in more completely understanding driver attitudes toward FCW systems.

# CHAPTER 7. SUMMARY AND CONCLUSIONS

This study created a system for generating RAD that can be used for safety analysis.

First, the team developed two independent RAD frameworks that can be adopted by other safety researchers interested in generating RAD for crashes of interest.

The macroscopic framework randomly generates a dataset of segments or intersections for a specified facility type (e.g., urban or rural two-lane, four-lane segment, or three-leg or four-leg signal or stop-controlled intersection), length of roadway or number of intersections, and number of years of crashes. The data are generated in two parts. First, physical and traffic characteristics of the segment or intersection are randomly generated according to distributions and sequential patterns (in the case of segments) observed in real datasets. Second, crashes are generated randomly using SPFs synthesized from distributions of parameter and CMF values found in a literature search of crash prediction models. Randomness is added to the resulting crash counts using a random mixture of several probability distributions to avoid producing a dataset with an easily discoverable common probability distribution.

The microscopic framework takes an alternative approach, generating vehicle trips for a specified region, each with an origin, destination, travel route, purpose, and time of day, and then the likelihood of each trip resulting in a crash, along with where on the route the crash occurs, is used to generate crashes. A rich trip dataset available from Argonne National Lab and the SHRP2 NDS was used to develop the model for predicting the likelihood of any trip resulting in a crash (Auld et al. 2016; Virginia Tech Transportation Institute 2020). The result is a database of crashes generated in a manner that attempts to replicate how crashes occur.

Second, the team developed an open-source software application that can be used to generate RAD for different combinations of inputs, including facility type, period of time, and geographic extent. A user desiring to generate a dataset can choose to use either the macroscopic or the microscopic approach. For the macroscopic approach, a user-friendly interface permits specification of facility type, miles of roadway segments or number of intersections, number of years of crashes, and a random number seed. The tool then generates a roadway file containing the roadway characteristics for each segment or intersection and total crash counts by injury severity and a crash file containing detailed crash severity and crash type information. For the microscopic approach, the interface has fewer parameters, requesting only the number of years of crashes. This tool generates annual crash data for the study region, providing crash, person, and vehicle files.

Third, the team conducted two case studies to demonstrate the feasibility and applicability of the software. They generated multiple datasets of various sizes using the resulting RAD tool for rural two-lane highways and urban four-leg signalized intersections. SPFs were estimated using statistical estimation software and then used to predict crashes on a different one of the RAD datasets. A variety of distributional assumptions and modeling approaches were considered for estimating the SPFs. The SPFs included a full complement of predictor variables. All variables used for generating the RAD were considered in each SPF. Statistical tests were used to compare the estimated parameter values from one model to another to evaluate dataset consistency.

Researchers found that, in general, the parameter values do not vary significantly from one dataset to another, even from the smallest to the largest datasets or among the modeling approaches. The resulting SPFs also demonstrated predictive transferability from one dataset to another.

Using driving simulator studies, the team explored the potential safety benefits of FCW systems. Participants were recruited to drive in one of four simulated scenarios in urban and rural areas. During the scenarios, an unexpected emergency stop was required due to either a pedestrian, animal, or other vehicle suddenly entering the vehicle's path. Half the drivers in each scenario were given an FCW system, which gave them an audible and/or visible warning about the hazard. Participants with the FCW system had a much higher TA than those without, and thus a much-reduced rate of serious conflicts and collisions with the hazard. The resulting reduced risk of conflict was used to estimate CMFs overall and for each scenario for a range of possible market penetration rates for vehicles with FCW systems.

In conclusion, the resulting RAD generation tool was proven to be reliable for generating datasets for testing new crash prediction approaches. The tool provides flexibility for generating datasets of any size desired, including many road facility types and crashes of different types and severity. The microscopic tool includes variables describing demographic and vehicle characteristics that can be used to investigate approaches for crash severity modeling. The macroscopic tool can also be used to generate crashes over an extended period of time, which is typically not possible with real crash data, in that road characteristics cannot be assumed to be constant. Having this tool available will permit further investigation of approaches for predicting crashes in scenarios where crashes are much too scarce to estimate models.

## REFERENCES


AASHTO. 2010. *The Highway Safety Manual*. Washington, DC: American Association of State Highway and Transportation Officials. http://www.highwaysafetymanual.org, last accessed June 20, 2023.

Abbas, K. A. 2004. "Traffic Safety Assessment and Development of Predictive Models for Accidents on Rural Roads in Egypt." *Accident Analysis and Prevention* 36, no. 2: 149−163. https://doi.org/10.1016/s0001–4575(02)00145-8, last accessed September 15, 2022.

Abdulhafedh, A. 2017. "Road Traffic Crash Data: An Overview on Sources, Problems, and Collection Methods." *Journal of Transportation Technologies* 7: 206−219. https://doi.org/10.4236/jtts.2017.72015, last accessed November 29, 2022.

Abe, G., and J. Richardson. 2006. "Alarm Timing, Trust and Driver Expectation for Forward Collision Warning Systems." *Applied Ergonomics* 37, no. 5: 577–586. https://doi.org/10.1016/j.apergo.2005.11.001, last accessed December 3, 2022.

Aguero-Valverde, J., and P. P. Jovanis. 2008. "Analysis of Road Crash Frequency with Spatial Models." *Transportation Research Record* 2061: 55–63. https://doi.org/10.3141/2061−07, last accessed December 28, 2022.

Ahmed M. M., M. Abdel-Aty, and J. Park. 2015. "Evaluation of the Safety Effectiveness of Divided Roadways: Bayesian Versus Empirical Bayes." *Transportation Research Record* 2515, 1: 41–49.

Al-Jabri, R. A. M. 2015. "Regression Analysis for Estimation of the Influencing Factors on Road Accident Injuries in Oman Poisson Regression Model and Poisson Alternatives." Master's thesis. University of Essex. https://repository.essex.ac.uk/16375/, last accessed November 4, 2022.

Amoros, E., J. L. Martin, and B. Laumon. 2003. "Comparison of Road Crashes Incidence and Severity Between Some French Countries." *Accident Analysis and Prevention* 35, 4: 537–547. https://doi.org/10.1016/s0001–4575(02)00031–3, last accessed August 20, 2022.

Anastasopoulos, P. C., and F. L. Mannering. 2009. "A Note on Modeling Vehicle Accident Frequencies with Random-Parameters Count Models." *Accident Analysis and Prevention* 41, 1: 153−159. https://doi.org/10.1016/j.aap.2008.10.005, last accessed November 29, 2022.

Aptech Systems. 2023. "Gauss 23" (software). http://www.aptech.com/, last accessed August 22, 2022.

Arnold, K., J. Gosling, and D. Holmes. 2005. *The Java Programming Language*. Boston, MA: Addison-Wesley .

Asano, M., T. Iryo, and M. Kuwahara. 2010. “Microscopic Pedestrian Simulation Model Combined with a Tactical Model for Route Choice Behaviour.” *Transportation Research Part C* 18, 6: 842–855. https://doi.org/10.1016/j.trc.2010.01.005, last accessed December 13, 2022.

Auld, J., M. Hope, H. Ley, V. Sokolov, B. Xu, and K. Zhang. 2016. “POLARIS: Agent-Based Modeling Framework Development and Implementation for Integrated Travel Demand and Network and Operations Simulations.” *Transportation Research Part C* 64: 101−116. https://sci-hub.se/10.1016/j.trc.2015.07.017, last accessed July 7, 2022.

Austin, M. P., L. Belbin, J. A. A. Meyers, M. D. Doherty, and M. Luoto. 2006. “Evaluation of Statistical Models Used for Predicting Plant Species Distributions: Role of Artificial Data and Theory.” *Ecological Modelling* 199, no. 2: 197−216. https://doi.org/10.1016/j.ecolmodel.2006.05.023, last accessed February 16, 2023.

Autey, J., T. Sayed, and M. H. Zaki. 2012. **“**Safety Evaluation of Right-Turn Smart Channels Using Automated Traffic Conflict Analysis.” *Accident Analysis & Prevention* 45: 120-130. https://doi.org/10.1016/j.aap.2011.11.015, last accessed December 14, 2022.

Bauer, K. M., and D. Harwood. 2000. “Statistical Models of At-Grade Intersection Accidents.” Report No. FHWA-RD-99-094. Washington, DC: Federal Highway Administration https://rosap.ntl.bts.gov/view/dot/35814, last accessed September 26, 2022.

Bhat, C. R. 2001. “Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model.” *Transportation Research Part B* 35: 677–693. https://doi.org/10.1016/S0191–2615(00)00014-X, last accessed December 8, 2022.

Bhat, C. R. 2003. “Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences.” *Transportation Research Part B* 37, no. 9: 837–855. https://repositories.lib.utexas.edu/bitstream/handle/2152/23925/Scrambled_Halton_Sequences.pdf?sequence=2&isAllowed=y, last accessed February 5, 2023.

Bhat, C. R., M. Castro, and M. Khan. 2013. “A New Estimation Approach for the Multiple Discrete-Continuous Probit (MDCP) Choice Model.” *Transportation Research Part B* 55: 1–22. https://www.sciencedirect.com/science/article/abs/pii/S0191261513000738, last accessed February 5, 2023.

Bhat, C. R., and R. Sidharthan. 2010. “A Simulation Evaluation of the Maximum Approximate Composite Marginal Likelihood (MACML) Estimator for Mixed Multinomial Probit Models.” *Transportation Research Part B* 45, No. 7: 940−953. https://repositories.lib.utexas.edu/bitstream/handle/2152/22649/TRPB_mixedmodelcml_simulation.pdf?sequence=1&isAllowed=y, last accessed October 15, 2022.

Bifulco, R., and S. Bretschneider, S. 2001. “Estimating School Efficiency: A Comparison of Methods Using Simulated Data.” *Economics of Education Review* 20: 417–429. https://www.sciencedirect.com/science/article/abs/pii/S027277570000025X, last accessed January 5, 2023.

Bonneson, J., and J. Ivan. 2013. “Theory, Explanation, and Prediction in Road Safety: Promising Directions.” *Transportation Research Circular*, No. E-C179. Washington, DC: Transportation Research Board of the National Academies. https://onlinepubs.trb.org/onlinepubs/circulars/ec179.pdf, last accessed January 13, 2023.

Brown, G. R. 1994. “Traffic Conflicts for Road User Safety Studies.” *Canadian Journal of Civil Engineering* 21, no. 1: 1–15. http://www.nrcresearchpress.com/loi/cjce, last accessed October 19, 2022.

Bzdusek, P. A., and E. R. Christensen. 2006. “Comparison of a New Variant of PMF with Other Receptor Modeling Methods Using Artificial and Real Sediment PCB Datasets.” *Environmetrics: The Official Journal of the International Environmetrics Society* 17, no. 4: 387−403. https://doi.org/10.1002/env.777, last accessed December 3, 2022.

Charly, A., and T. V. Mathew. 2019. “Estimation of Traffic Conflicts Using Precise Lateral Position and Width of Vehicles for Safety Assessment.” *Accident Analysis & Prevention* 132: 105264. https://doi.org/10.1016/j.aap.2019.105264, last accessed December 3, 2022.

Cicchino, J. B. 2017. “Effectiveness of Forward Collision Warning and Autonomous Emergency Braking Systems in Reducing Front-to-Rear Crash Rates.” *Accident Analysis & Prevention* 99, Part A: 142–152. https://doi.org/10.1016/j.aap.2016.11.009, last accessed January 5, 2023.

Crewson, P. 2006. *Applied Statistics: Desktop Reference*, 1st ed. https://www.acastat.com/Pub/Docs/AppliedStatistics.pdf, last accessed October 24, 2022.

NHTSA. n.d. “Crash Report Sampling System: Motor Vehicle Crash Data Collection” (dataset). https://www.nhtsa.gov/crash-data-systems/crash-report-sampling-system, last accessed August 23, 2023.

Cummings, P., B. McKnight, and N. S. Weiss. 2003. “Matched-Pair Cohort Methods in Traffic Crash Research.” *Accident Analysis & Prevention* 35, no. 1: 131−141. https://pubmed.ncbi.nlm.nih.gov/12479904/, last accessed January 5, 2023.

Dahmen, J., and D. Cook. 2019. “SynSys: A Synthetic Data Generation System for Healthcare Applications.” *Sensors*. 19, no. 5: 1181. https://pubmed.ncbi.nlm.nih.gov/30857130/, last accessed August 16, 2022.

Devroye, L., T. Felber, and M. Kohler. 2012. “Estimation of a Density Using Real and Artificial Data.” *IEEE Transactions on Information Theory* 59, no. 3: 1917–1928.

Donnell, E. T., R. J. Porter, and V. N. Shankar. 2010. “A Framework for Estimating the Safety Effects of Roadway Lighting at Intersections.” *Safety Science* 48, no. 10: 1436–1444. https://doi.org/10.1016/j.ssci.2010.06.008, last accessed September 24, 2022.

Eluru, N. 2013. “Evaluating Alternate Discrete Choice Frameworks for Modeling Ordinal Discrete Variables.” *Accident Analysis & Prevention* 55, no. 1: 1–11. https://pubmed.ncbi.nlm.nih.gov/23500025/, last accessed February 17, 2023.

Eluru, N., C. R. Bhat, and D. A. Hensher. 2008. “A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes.” *Accident Analysis & Prevention* 40, no. 3: 1033−1054. https://doi.org/10.1016/j.aap.2007.11.010, last accessed September 5, 2022.

Eluru, N., A. R. Pinjari, J. Y. Guo, I. N. Sener, S. Srinivasan, R. B. Copperman, and C. R. Bhat. 2008. “Population Updating System Structures and Models Embedded in the Comprehensive Econometric Microsimulator for Urban Systems.” *Transportation Research Record* 2076, no. 1: 171–182. https://repositories.lib.utexas.edu/bitstream/handle/2152/23823/TRR_CEMSELTS_1Apr08.pdf?sequence=2&isAllowed=y, last accessed February 17, 2023.

Essa, M., and T. Sayed. 2020. “Comparison Between Surrogate Safety Assessment Model and Real-Time Safety Models in Predicting Field-Measured Conflicts at Signalized Intersections.” *Transportation Research Record* 2674, no. 3: 100–112. https://doi.org/10.1177/0361198120907874, last accessed September 28, 2022.

Ferdous, N., N. Eluru, C. R. Bhat, and I. Meloni. 2010. “A Multivariate Ordered Response Model System for Adults’ Weekday Activity Episode Generation by Activity Purpose and Social Context,” *Transportation Research Part B* 44, no. 8–9: 922–943.

Gagniuc, P. A. 2017. *Markov Chains: From Theory to Implementation and Experimentation.* Hoboken, NJ: Wiley.

Gamel, J. W., and R. L. Vogel. 1997. “Comparison of Parametric and Non-Parametric Survival Methods Using Simulated Clinical Data.” *Statistics in Medicine* 16: 1629–1643. https://www.semanticscholar.org/paper/Non-parametric-comparison-of-relative-versus-in-and-Gamel-Vogel/25a1a50eba013d3845d6613471fb311e226704e1, last accessed January 5, 2023.

Garber, N. J., and G. Rivera. 2010. *Safety Performance Functions for Intersections on Highways Maintained by the Virginia Department of Transportation*. Report No. FHWA-VTRC-11-CR1. Richmond, VA: Virginia DOT. http://www.virginiadot.org/vtrc/main/online_reports/pdf/11-cr1.pdf, last accessed September 25, 2022.

Gates, T. J., P. T. Savolainen, R. E. Avelar, S. R. Geedipally, D. Lord, A. Ingle, and S. Y. Stapleton. 2018. “Safety Performance Functions for Rural Road Segments and Rural Intersections in Michigan.” *Journal of the Transportation Research Board* 2673, no. 10. https://doi.org/10.1177/0361198119850127, last accessed September 17, 2022.

Geedipally, S. R., Lord, D., and Dhavala, S. S. 2012. "The Negative Binomial—Lindley Generalized Linear Model: Characteristics and Application Using Crash Data.” *Accident Analysis & Prevention* 45: 258-265. https://pubmed.ncbi.nlm.nih.gov/22269508/, last accessed January 23, 2023.

Glassco, R. A., and D. S. Cohen. 2001. "Simulating Rear-End Collision Warnings Using Field Operational Test Data." *SAE Transactions* 110: 183–190. http://www.jstor.org/stable/44718322, last accessed February 5, 2023.

Greibe, P. 2003. "Accident Prediction Models for Urban Roads." *Accident Analysis & Prevention* 35, no. 2: 273−285. https://doi.org/10.1016/S0001–4575(02)00005−2, last accessed February 5, 2023.

Gross, F., and P. P. Jovanis. 2019. "Estimation of Safety Effectiveness of Changes in Shoulder Width with Case Control and Cohort Methods." *Transportation Research Record* 1: 237−245. https://doi.org/10.3141/2279−12, last accessed January 20, 2023.

Gross F., P. P. Jovanis, and K. Eccles. 2009. "Safety Effectiveness of Lane and Shoulder Width Combinations on Rural, Two-Lane, Undivided Roads." *Transportation Research Record* 2103, no. 1: 42−49. https://doi.org/10.3141/2103−06, last accessed November 8, 2022.

Hankey, J. M., M. A. Perez, and J. A. McClafferty. 2016. *Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Datasets*. Blacksburg, VA: Virginia Tech Transportation Institute. https://vtechworks.lib.vt.edu/handle/10919/70850 last accessed August 3, 2023.

Harwood, D. W., K. M. Bauer, I. B. Potts, D. J. Torbic, K. R. Richard, E. R. Rabbani, E. Hauer, L. Elefteriadou, and M. S. Griffith. 2003. "Safety Effectiveness of Intersection Left- and Right-Turn Lanes." Presented at the *82nd Transportation Research Board Annual Meeting*. Washington, DC: TRB.

Hazwani, R. A., N. Wahida, S. I. Shafikah, and P. N. Ellyza. 2016. "Automatic Artificial Data Generator: Framework and Implementation." In *2016 International Conference on Information and Communication Technology (ICICTM)*, 56-60. New York, NY: IEEE.

FHWA. n.d. "Highway Safety Information System" (database). https://highways.dot.gov/research/safety/hsis/overview, last accessed August 23, 2023.

Himes, S., F. Gross, M. Nichols, and M. Lockwood. 2018. "Safety Evaluation of Change in Posted Speed Limit From 65 to 70 mph on Rural Virginia Interstate System." *Transportation Research Record* 2672, no. 38: 35–45."

Hirst, W. M., L. J. Mountain, and M. J. Maher. 2004. "Sources of Error in Road Safety Scheme Evaluation: A Method to Deal with Outdated Accident Prediction Models." *Accident Analysis and Prevention* 36, no. 5: 717–727. https://doi.org/10.1016/j.aap.2003.05.00, last accessed January 4, 2023.

Hoover, L., T. Bhowmik, S. Yasmin, and N. Eluru. 2022. "Understanding Crash Risk Using a Multi-Level Random Parameter Binary Logit Mode: Application to Naturalistic Driving Study Data." *Transportation Research Record* 2676, no. 10: 737–745. https://doi.org/10.1177/03611981221090943, last accessed February 4, 2023.

Hovey, P. W., and M. Chowdhury. 2005. *Development of Crash Reduction Factors*. Report No. FHWA/OH-2005/12. Columbus, Ohio: Ohio Department of Transportation. https://www.researchgate.net/publication/292148719_Development_of_Crash_Reduction_Factors, last accessed November 14, 2022.

Ivan, J., S. Mamun, N. Ravishanker, B. Persaud, C. Lyon, R. Srinivasan, B. Lan, S. Smith, T. Saleem, M. Abdel-Aty, J. Lee, A. Farid, and J. Wang. 2021. *Improved Prediction Models for Crash Types and Crash Severities*. Report No. NCHRP 17-62. Washington, DC: Transportation Research Board.

Jermakian, J. S. 2011. "Crash Avoidance Potential of Four Passenger Vehicle Technologies." *Accident Analysis & Prevention* 43: 732–740. https://pubmed.ncbi.nlm.nih.gov/21376861/, last accessed October 15, 2022.

Jones, L., F. Janssen, and F. Mannering. 1991. "Analysis of the Frequency and Duration of Freeway Accidents in Seattle." *Accident Analysis and Prevention* 23, 2: 239−255. http://www.sciencedirect.com/science/journal/00014575, last accessed December 5, 2022.

Kamel, J., R. Vosooghi, J. Puchinger, F. Ksontini, and G. Sirin. 2019. "Exploring the Impact of User Preferences on Shared Autonomous Vehicle Modal Split: A Multi-Agent Simulation Approach." *Transportation Research Procedia* 37: 115–122. https://www.sciencedirect.com/science/article/pii/S235214651830588X, last accessed June 29, 2023.

Khattak, M. W., A. Pirdavani, P. De Winne, T. Brijs, and H. De Backer. 2021. "Estimation of Safety Performance Functions for Urban Intersections Using Various Functional Forms of the Negative Binomial Regression Model and a Generalized Poisson Regression Model." *Accident Analysis & Prevention* 151: 105964. https://doi.org/10.1016/j.aap.2020.105964, last accessed August 29, 2022.

Kitamura, R., C. Chen, R. M. Pendyala, and R. Narayanan. 2000. "Micro-Simulation of Daily Activity-Travel Patterns for Travel Demand Forecasting." *Transportation* 27, no. 1: 25-51. https://doi.org/10.1023/A:1005259324588, last accessed September 26, 2022.

Klebelsberg, D. 1964. "Traffic Conflict Characteristics: Accident Potential at Intersections." *Highway Research Record* 224: 35–43.

Konduri, K. C., D. You, V. M. Garikapati, and R. M. Pendyala. 2016. "Enhanced Synthetic Population Generator that Accommodates Control Variables at Multiple Geographic Resolutions." *Transportation Research Record* 2563, no. 1: 40–50.

Kusano, K. D., and H. C. Gabler. 2012. "Safety Benefits of Forward Collision Warning, Brake Assist, and Autonomous Braking Systems in Rear-End Collisions." *IEEE Transactions on Intelligent Transportation Systems* 13, no. 4: 1546–1555. https://www.researchgate.net/publication/258359207_Safety_Benefits_of_Forward_Collision_Warning_Brake_Assist_and_Autonomous_Braking_Systems_in_Rear-End_Collisions, last accessed January 13, 2023.

Labi, S. 2011. "Efficacies of Roadway Safety Improvements Across Functional Subclasses of Rural Two-Lane Highways." *Journal of Safety Research* 42, no. 4: 231–239. https://doi.org/10.1016/j.jsr.2011.01.008, last accessed October 28, 2022.

Laureshyn, A., and A. Varhelyi. 2020. *The Swedish Traffic Conflict Technique: Observers Manual*. Lund, Sweden: Lund University. https://lucris.lub.lu.se/ws/files/51195704/TCT_Manual_2018.pdf, last accessed February 15, 2023.

Lee, J. D., and H. C. Lee. 2007. "Technology and Teen Drivers." *Journal of Safety Research* 38: 203–213. https://doi.org/10.1016/j.jsr.2007.02.008, last accessed November 4, 2022.

Lee, J. D., D. V. McGehee, T. L. Brown, and M. L. Reyes. 2002. "Collision Warning Timing, Driver Distraction, and Driver Response to Imminent Rear-End Collisions in a High-Fidelity Driving Simulator." *Human Factors* 44, no. 2: 314–334. https://doi.org/10.1518/0018720024497844, last accessed February 2, 2023.

Li, H., M. Zhu, D. J. Graham, and Y. Zhang. 2020. "Are Multiple Speed Cameras More Effective Than One? Causal Analysis of the Safety Impacts of Multiple Speed Cameras." *Accident Analysis and Prevention* 139. https://doi.org/10.1016/j.aap.2020.105488, last accessed August 30, 2022.

Lord, D., and P. F. Kuo. 2012. "Examining the Effects of Site Selection Criteria for Evaluating the Effectiveness of Traffic Safety Countermeasures." *Accident Analysis & Prevention* 47: 52–63.

Lord, D., and F. Mannering. 2010. "The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives." *Transportation Research Part A* 44, no. 5: 291–305. https://doi.org/10.1016/j.tra.2010.02.001, last accessed September 14, 2022.

Lord, D., and F. Miranda-Moreno. 2008. "Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter of Poisson-Gamma Models for Modeling Motor Vehicle Crashes: A Bayesian Perspective." *Accident Analysis and Prevention* 46: 751–770. https://doi.org/10.1016/j.ssci.2007.03.005, last accessed December 19, 2022.

Lyon, C., P. Bhagwant, and K. Eccles. 2015. *Safety Evaluation of Centerline Plus Shoulder Rumble Strips*. Report No. FHWA-HRT-15-048. Washington, DC: Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/35718, last accessed October 5, 2022.

Mamun, S., F. J. Caraballo, J. N. Ivan, N. Ravishanker, R. M. Townsend, and Y. Zhang. 2020. "Identifying Association Between Pedestrian Safety Interventions and Street-Crossing Behavior Considering Demographics and Traffic Context." *Journal of Transportation Safety & Security* 12, no. 3: 441–462.

McLaughlin, S. B., J. M. Hankey, and T. A. Dingus. 2008. "A Method for Evaluating Collision Avoidance Systems Using Naturalistic Driving Data." *Accident Analysis & Prevention*

40, no. 1: 8−16. https://doi.org/10.1016/j.aap.2007.03.016, last accessed February 28, 2023.

Miaou, S. P. 1994. "The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions." *Accident Analysis and Prevention* 26, no. 4: 471–482. https://doi.org/10.1016/0001-4575(94)90038-8, last accessed January 6, 2023.

Miaou, S. P., and D. Lord. 2003. "Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods." *Transportation Research Record* 1840, no. 1: 31–40. https://doi.org/10.3141/1840−04, last accessed November 17, 2022.

Milton, J., V. Shankar, and F. L. Mannering. 2008. "Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis." *Accident Analysis & Prevention* 40, no. 1: 260−266. https://doi.org/10.1016/j.aap.2007.06.006, last accessed November 5, 2022.

Minchin, P.R. 1987. "Simulation of Multidimensional Community Patterns: Towards a Comprehensive Model." *Plant Ecology* 71: 145–156.

Najm, W. G., J. Koopmann, J. D. Smith, and J. Brewer. 2010. *Frequency of Target Crashes for 12 Intellidrive Safety Systems*. Report No. DOT HS 811 381. Washington, DC: National Highway Traffic Safety Administration. https://rosap.ntl.bts.gov/view/dot/12066, last accessed October 15, 2022.

Najm, W. G., and D. L. Smith. 2004. *Modeling Driver Response to Lead Vehicle Decelerating*. Report No. FHWA-JPO-04-091; 2004-01-0171. https://rosap.ntl.bts.gov/view/dot/4336, last accessed February 28, 2023.

National Transportation Safety Board. 2002. *Vehicle and Infrastructure-Based Technology for Prevention of Rear-End Collisions*. Report No. NTSB/SIR−01/01, PB2001-917003. Washington, DC: National Transportation Safety Board. https://www.ntsb.gov/safety/safety-studies/Documents/SIR0101.pdf, last accessed January 6, 2023.

N'Guessan, A. 2010. "Analytical Existence of Solutions to a System of Nonlinear Equations with Application." *Journal of Computational and Applied Mathematics* 234: 297−304. https://doi.org/10.1016/j.cam.2009.12.026, last accessed November 18, 2022.

N'Guessan, A., and C. Langrand. 2003. "A Covariance Components Estimation Procedure When Modelling A Road Safety Measure in Terms of Linear Constraints." *Statistics* 39, no. 4: 303−314. https://doi.org/10.1080/02331880500108544, last accessed October 15, 2022.

NHTSA. 2021. "NHTSA" (web site). https://www.nhtsa.gov/, last accessed August 23, 2023.

NHTSA. 2017. *MMUCC Guideline: Model Minimum Uniform Crash Criteria*, 5th ed. Washington, DC: National Highway Traffic Safety Administration.

https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812433, last accessed August 23, 2023.

Ogle, J., W. Sarasua, J. Dillon, V. Bendigieri, S. Anekar, and P. Alluri. 2009. *Support for the Elimination of Roadside Hazards: Evaluating Roadside Collision Data and Clear Zone Requirements*. Report No. FHWA-SC-09-01. Washington, DC: Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/25112, last accessed August 29, 2023.

Paez, A., and D. M. Scott. 2007. "Social Influence on Travel Behavior: A Simulation Example of the Decision to Telecommute." *Environment and Planning A* 39, no. 3: 647–665.

Paleti, R., and C. R. Bhat. 2013. "The Composite Marginal Likelihood (CML) Estimation of Panel Ordered-Response Models." *Journal of Choice Modelling* 7: 24–43. https://doi.org/10.1068/b3319t, last accessed August 17, 2022.

Papadoulis, A., M. Quddus, and M. Imprialou. 2019. "Evaluating the Safety Impact of Connected and Autonomous Vehicles on Motorways." *Accident Analysis & Prevention* 124: 12–22. https://doi.org/10.1016/j.aap.2018.12.019, last accessed January 12, 2023.

Park, B., K. Fitzpatrick, and M. Brewer. 2019. "Safety Effectiveness of Super 2 Highways in Texas." *Transportation Research Record: Journal of the Transportation Research Board* 2280: 38–50. https://doi.org/10.3141/2280−05, last accessed December 29, 2022.

Parker, M. R., Jr. 1997. *Effects of Raising and Lowering Speed Limits on Selected Roadway Sections*. Report No. FHWA-RD-92-084. Washington, DC: Federal Highway Administration. https://www.fhwa.dot.gov/publications/research/safety/97084/97084.pdf, last accessed September 19, 2022.

Peesapati, L., M. Hunter, and M. Rodgers. 2013. "Evaluation of Postencroachment Time as Surrogate for Opposing Left-Turn Crashes." *Transportation Research Record: Journal of the Transportation Research Board* 2386, no. 1: 42–51. https://doi.org/10.3141/2386−06, last accessed December 9, 2022.

Perkins, S. R., and J. I. Harris. 1968. "Traffic Conflict Characteristics—Accident Potential at Intersections." *Highway Research Record* 225: 35–43. http://onlinepubs.trb.org/Onlinepubs/hrr/1968/225/225−004.pdf, last accessed August 4, 2022.

Persaud, B., C. Lyon, K. Eccles, N. Lefler, D. Carter, and R. Amjadi. 2007. *Safety Evaluation of Installing Center Two-Way Left-Turn Lanes on Two-Lane Roads*. Report No. FHWA-HRT-08-042. Washington, DC: Federal Highway Administration. https://ntlrepository.blob.core.windows.net/lib/31000/31000/31089/FHWA-HRT−08−042.pdf, last accessed February 24, 2023.

Pinjari, A. R., and C. R. Bhat. 2010. "A Multiple Discrete-Continuous Nested Extreme Value (MDCNEV) Model: Formulation and Application to Nonworker Activity Time-Use and Timing Behavior on Weekdays." *Transportation Research Part B* 44, No. 4: 562-583. https://repositories.lib.utexas.edu/bitstream/handle/2152/23809/TRPB_Pinjari_

Bhat_MDCNEV_Revised_August4_09.pdf?sequence=1, last accessed November 16, 2022.

Pinjari, A., N. Eluru, S. Srinivasan, J. Y. Guo, R. Copperman, I. N. Sener, and C. R. Bhat. 2008. "Cemdap: Modeling and Microsimulation Frameworks, Software Development, and Verification." In *Proceedings of the Transportation Research Board 87th Annual Meeting*. Washington, DC: Transportation Research Board.

Pokorny, P., J. Jensen, F. Gross, and K. Pitera. 2020. "Safety Effects of Traffic Lane and Shoulder Widths on Two-Lane Undivided Rural Roads: A Matched Case-Control Study from Norway." *Accident Analysis & Prevention* 144: 105614. https://doi.org/10.1016/j.aap.2020.105614, last accessed September 5, 2022.

Potharst, R., A. Ben-David, and M. Van Wezel. 2009. "Two Algorithms for Generating Structured and Unstructured Monotone Ordinal Datasets." *Engineering Applications of Artificial Intelligence* 22, no. 4–5: 491–496. https://doi.org/10.1016/j.engappai.2009.02.004, last accessed August 9, 2022.

Python Software Foundation. 2023. *Python* (programming language software). https://www.python.org/, last accessed August 23, 2023.

Ranade, S., A. W. Sadek, and J. N. Ivan. 2007. "Decision Support System for Predicting Benefits of Left-Turn Lanes at Unsignalized Intersections." *Transportation Research Record* 2023, no. 1: 28−36. https://trid.trb.org/view/801420, last accessed October 19, 2022.

R Foundation. 2021. *R: A Language and Environment for Statistical Computing,* version 12.0 (software). https://www.R-project.org/, last accessed October 2, 2023.

Realtime Technologies. 2023. Realtime Driving Simulator (system), including SimCreator®, SimObserver®, SimCreator DX®, and Data Distillery® (softwares). https://www.faac.com/realtime-technologies/products/rds-modular-driving-simulator/, last accessed August 22, 2023.

Reinmueller, K., and M. Steinhauser. 2019. "Adaptive Forward Collision Warnings: The Impact of Imperfect Technology on Behavioral Adaptation, Warning Effectiveness and Acceptance." *Accident Analysis & Prevention* 128: 217−229. https://doi.org/10.1016/j.aap.2019.04.012, last accessed February 5, 2023.

Saleem, T., R. Srinivasan, D. Levitt, M. Vann, A. Worzella, and R. Storm. 2020. *Speed Limit Change (55 MPH to 60 MPH) Safety Evaluation*. Report No. MN 2020-06. Roseville, MN: Minnesota Department of Transportation. https://www.dot.state.mn.us/research/reports/2020/202006.pdf, last accessed September 26, 2022.

Salim, F. D., S. W. Loke, A. Rakotonirainy, and S. Krishnaswamy. 2007. "Simulated Intersection Environment and Learning of Collision and Traffic Data in the U&I Aware Framework." *Ubiquitous Intelligence and Computing: UIC 2007—Lecture Notes in*

*Computer Science* 4611. Berlin, Germany: Springer. https://doi.org/10.1007/978-3-540-73549-6_16, last accessed August 8, 2023.

Santamarina-Rubio, E., K. Perez, M. Olabarria, and A. M. Novoa. 2014. "Gender Differences in Road Traffic Injury Rate Using Time Travelled as a Measure of Exposure." *Accident Analysis & Prevention* 65: 1–7. https://doi.org/10.1016/j.aap.2013.11.015, last accessed January 14, 2023.

Scott, P. D., and Wilkins, E. 1999. "Evaluating Data Mining Procedures: Techniques for Generating Artificial Datasets." *Information and Software Technology* 41, 9: 579–587. https://doi.org/10.1016/S0950-5849(99)00021-X, last accessed August 24, 2022.

Shahzad, U., S. Shahzadi, N. Afshan, N. H. Al-Noor, D. Anekeya Alilah, M. Hanif, and M. M. Anas. 2021. "Poisson Regression-Based Mean Estimator." *Mathematical Problems in Engineering* 2021. https://doi.org/10.1155/2021/9769029, last accessed September 26, 2023.

Virginia Tech Transportation Institute. 2020. "InSight Data Access Website: SHRP2 Naturalistic Driving Study" (web page). https://insight.shrp2nds.us/home, last accessed August 22, 2023.

Spence, I. 1983. "Monte Carlo Simulation Studies." *Applied Psychological Measurement* 7, no. 4: 405-425 .

Srinivasan, R., D. Carter, C. Lyon, and M. Albee. 2018. "Before-After Evaluation of the Realignment of Horizontal Curves on Rural Two-Lane Roads." *Transportation Research Record 2672*, no. 30: 43–52. https://doi.org/10.1177/0361198118758011, last accessed October 14, 2022.

Srinivasan, R., B. Lan, and D. Carter. 2014. *Safety Evaluation of Signal Installation With and Without Left Turn Lanes on Two Lane Roads in Rural and Suburban Areas*. Report No. FHWA/NC/2013-11. Raleigh, North Carolina: North Carolina Department of Transportation. https://connect.ncdot.gov/projects/research/RNAProjDocs/2013-11finalreport.pdf, last accessed September 26, 2022.

StataCorp. 2023. Stata Statistical Software: Release 18.

Sullivan, J. 2019. "Calibration of the Highway Safety Manual Predictive Models for Rural Two-Lane Roads for Vermont." *Transportation Research Center Research Reports*. https://www.uvm.edu/sites/default/files/Transportation-Research-Center/Reports/2020%20and%20more/2019_-_Calibration_of_Highway_Safety_Manual.pdf?t=quk4im, last accessed September 17, 2022.

Tarko, A. P. 2018. "Estimating the Expected Number of Crashes With Traffic Conflicts and the Lomax Distribution—A Theoretical and Numerical Exploration." *Accident Analysis & Prevention* 113: 63–73. http://dx.doi.org/10.1016/j.aap.2018.01.008, last accessed October 1, 2022.

Tarko, A., G. Davis, N. Saunier, and T. Sayed. 2009. "Surrogate Measures of Safety." *Safe Mobility: Challenges, Methodology and Solutions* 11: 383–405. https://doi.org/10.1108/S2044-994120180000011019, last accessed October 13, 2023.

Transera. 1991. *Transera HπBasic Reference Manual*. Provo, UT: Transera.

Teoh, E. R. 2021. "Effectiveness of Front Crash Prevention Systems in Reducing Large Truck Real-World Crash Rates." *Traffic Injury Prevention* 22, no. 4: 284−289. https://trid.trb.org/view/1850298, last accessed November 18, 2022.

Van Rossum, G., and F. L. Drake. 1995. *Python Reference Manual* 111: 1–52. Amsterdam, Netherlands: Centrum voor Wiskunde en Informatica.

Wang, K., J. N. Ivan, N. Ravishanker, and E. Jackson. 2017. "Multivariate Poisson Lognormal Modeling of Crashes by Type and Severity on Rural Two-Lane Highways." *Accident Analysis and Prevention, Part A* 99: 6–19. https://doi.org/10.1016/j.aap.2016.11.006, last accessed September 25, 2022.

Wang, Z., L. Chanyoung, L. Pei-Sung, X. Chunfu, Y. Runan, K. Rama, and V.A. Abhijit. 2018. *Center for Urban Transportation Research: Study on Motorcycle Safety in Negotiation With Horizontal Curves in Florida and Development of Crash Modification Factors*. Report No. BDV25-977-21. Tampa, FL: Florida DOT. https://rosap.ntl.bts.gov/view/dot/63608, last accessed August 22, 2023.

Washington, S. P., M. G. Karlaftis, and F. L. Mannering. 2010. *Statistical and Econometric Methods for Transportation Data Analysis,* 2nd ed. Boca Raton, FL: Chapman Hall/CRC.

Whiting, M. A., J. Haack, and C. Varley. 2008. "Creating Realistic, Scenario-Based Synthetic Data for Test and Evaluation of Information Analytics Software." In *Proceedings of the 2008 Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization, pp. 1-9.* https://doi.org/10.1145/1377966.1377977, last accessed August 29, 2022.

Wu, L., D. Lord, and Y. Zou. 2015. "Validation of Crash Modification Factors Derived from Cross-Sectional Studies with Regression Models." *Transportation Research Record* 2514, no. 1: 88–96. https://doi.org/10.3141/2514−10, last accessed January 20, 2023.

Xiao, G., J. Lee, Q. Jiang, H. Huang, M. Abdel-Aty, and L. Wang. 2021. "Safety Improvements by Intelligent Connected Vehicle Technologies: A Meta-Analysis Considering Market Penetration Rates." *Accident Analysis & Prevention* 159: 106234. https://doi.org/10.1016/j.aap.2021.106234, last accessed on January 12, 2023.

Xie, K., C. Li, K. Ozbay, G. Dobler, H. Yang, A. T. Chiang, and M. Ghandehari. 2016. "Development of a Comprehensive Framework for Video-Based Safety Assessment." 2016 IEEE *19th International Conference on Intelligent Transportation Systems (ITSC)*: 2638−2643. https://doi.org/10.1109/ITSC.2016.7795980, last accessed December 13, 2022.

Xin, C., Z. Wang, C. Lee, P. S. Lin, T. Chen, R. Guo, and Q. Lu. 2019. "Development of Crash Modification Factors of Horizontal Curve Design Features for Single-Motorcycle Crashes on Rural Two-Lane Highways: A Matched Case-Control Study." *Accident Analysis and Prevention 123*: 51–59. https://doi.org/10.1016/j.aap.2018.11.008, last accessed September 8, 2022.

Ye, F., and D. Lord. 2011. "Investigation of Effects of Underreporting Crash Data on Three Commonly Used Traffic Crash Severity Models: Multinomial Logit, Ordered Probit, and Mixed Logit." *Transportation Research Record* 2241, no. 1: 51–58. https://doi.org/10.3141/2241−06, last accessed September 14, 2023.

Yu, R., and M. Abdel-Aty. 2014. "An Optimal Variable Speed Limits System To Ameliorate Traffic Safety Risk." *Transportation Research Part C* 46: 235–246. https://doi.org/10.1016/j.trc.2014.05.016, last accessed December 22, 2022.

Zegeer, C. V., J. Hummer, D. Reinfurt, L. Herf, and W. Hunter. 1986. *Safety Effects of Cross-Section Design for Two-Lane Roads*. Report No. FHWA-RD-87-008. Washington, DC: Federal Highway Administration.

Zheng, L., Ismail, K., and Meng, X. 2014. "Traffic Conflict Techniques for Road Safety Analysis: Open Questions and Some Insights." *Canadian Journal of Civil Engineering* 41, no. 7: 633–641. https://doi.org/10.1139/cjce-2013-0558, last accessed February 15, 2023.

Zimmermann, A. 2012. "Generating Diverse Realistic Datasets for Episode Mining." Presented at the *IEEE 12th International Conference on Data Mining Workshops*. Brussels, Belgium: IEEE.

## BIBLIOGRAPHY


Arvin, R., A. Khattak, and J. Rios-Torres. 2019. "Evaluating Safety With Automated Vehicles at Signalized Intersections: Application of Adaptive Cruise Control in Mixed Traffic." Presented at the *2019 Transportation Research Board Annual Meeting*. Washington, DC: Transportation Research Board. https://www.osti.gov/biblio/1493116, last accessed February 5, 2023.

Bagdade, J., A. Ceifetz, M. Myers, C. Redinger, B. N. Persaud, and C. Lyon. 2011. *Evaluating Performance and Making Best Use of Passing Relief Lanes*. Report No. RC-1565. West Bloomfield, MI: Michigan Department of Transportation. https://rosap.ntl.bts.gov/view/dot/24003, last accessed September 23, 2022.

Bhat, C. R. 2018. "New Matrix-Based Methods for the Analytic Evaluation of the Multivariate Cumulative Normal Distribution Function." *Transportation Research Part B* 109: 238−256. https://www.sciencedirect.com/science/article/abs/pii/S0191261517306847, last accessed February 5, 2023.

Graham, J. L., D. W. Harwood, K. R. Richard, M. K. O'Laughlin, E. T. Donnell, and S. N. Brennan. 2014. Median Cross Section Design for Rural Divided Highway. National Cooperative Highway Research Program Report 794. Washington DC .http://kls-eng.com/nchrp_rpt_794.pdf , last accessed December 13, 2022.

Jamson, A. H., C. H. Lai, and O. M. J. Carsten. 2007. "Potential Benefits of an Adaptive Forward Collision Warning System." *Transportation Research Part C* 16, no. 4: 471-484. https://doi.org/10.1016/j.trc.2007.09.003, last accessed January 5, 2023.

Mannering, F., and C. Bhat. 2014. "Analytic Methods in Accident Research: Methodological Frontier and Future Directions." *Analytic Methods in Accident Research* 1: 1–22. https://doi.org/10.1016/j.amar.2013.09.001m, last accessed December 16, 2022.

Najm, W. G., and A. L. Burgett. 1997. "Benefits Estimation for Selected Collision Avoidance Systems." In *Mobility for Everyone 4th World Congress on Intelligent Transport Systems, 21–24 October 1997, Berlin*. Report No. 1022. https://trid.trb.org/view/541511, last accessed November 18, 2022. Washington, DC: ITS America.

Pendyala, R. M., K. C. Konduri, Y. C. Chiu, M. Hickman, H. Noh, P. Waddell, L. Wang, D. You, and B. Gardner. 2012. "Integrated Land Use—Transport Model System With Dynamic Time-Dependent Activity-Travel Microsimulation." *Transportation Research Record* 2303, no. 1: 19−27. https://core.ac.uk/download/pdf/16695972.pdf, last accessed November 8, 2022.

Rahman, M. S., M. Abdel-Aty, J. Lee, and M. H. Rahman. 2019. "Safety Benefits of Arterials' Crash Risk Under Connected and Automated Vehicles." *Transportation Research Part C* 100: 354−371. https://doi.org/10.1016/j.trc.2019.01.029, last accessed November 18, 2022.

Schumaker, L., M. M. Ahmed, and K. Ksaibati. 2016. “Policy Considerations for Evaluating the Safety Effectiveness of Passing Lanes on Rural Two-Lane Highways With Lower Traffic Volumes: Wyoming 59 Case Study.” *Journal of Transportation Safety & Security* 9, no. 1: 1−19. http://dx.doi.org/10.1080/19439962.2015.1055415, last accessed January 25, 2022.

Recycled
Recyclable

Recommended citation: Federal Highway Administration,
*DREDGE (Disaggregate Realistic Artificial Data Generator)—Design, Development, and Application for Crash Safety Analysis, Volume I*
(Washington, DC: 2024) https://doi.org/10.21949/1521453

HRSO-02/01-24(WEB)E