Understanding the Causative, Precipitating, and Predisposing Factors in Rural Two-Lane Crashes

PUBLICATION NO. FHWA-HRT-17-079 JULY 2018

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

The Federal Highway Administration (FHWA) is exploring the use of alternate, non-traditional data sets. Access to relevant data is critical to understanding when traditional data sets seem to have been exhausted and to finding novel ways to mitigate the consequences of crashes. This study examines alternate data sources to increase understanding of precipitating events and predisposing crash factors. Models were developed for crashes occurring on horizontal curves and unsignalized intersections along rural two-lane roads to create crash profiles.

This technical report forms part of the series of low-cost in-house efforts to make full use of existing data resources to understand the roadway departure problem. The first report, *Photographic Data Extraction Feasibility and Pilot Study in Support of Roadside Safety and Roadway Departure Research*, sought to understand the roadway departure problem by repurposing existing data and blending it with emerging data sources.(1) Its purpose was to provide additional detail on roadside hardware and potentially identify previously unreported roadside conditions in the repurposed data. The intended audience is comprised of current data scientists, transportation researchers, and decisionmakers from State and Federal departments of transportation.

Monique R. Evans, P.E., CPM
Director, Office of Safety Research
and Development

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-HRT-17-079	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Understanding the Causative, Precipitating, and Predisposing Factors in Rural Two-Lane Crashes		5. Report Date July 2018
		6. Performing Organization Code:
7. Author(s) Richard Porter, Scott Himes, Anusha Musunuru, Thanh Le, Kim Eccles, Kara Peach, Ivana Tasic, Milan Zlatkovic, Kailash Tatineni, and Brendan Duffy		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Utah 201 Presidents Circle Salt Lake City, UT 84112	VHB 4000 West Chase Blvd, Ste. 530 Raleigh, NC 27607	10. Work Unit No.
		11. Contract or Grant No. DTFH61-12-C-00033
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Highway Administration 1200 New Jersey Avenue SE Washington, DC 20590		13. Type of Report and Period Covered Technical Report; October 2012–September 2017
		14. Sponsoring Agency Code HRDS-20

15. Supplementary Notes
The Federal Highway Administration Contracting Officer's Technical Manager for this project was Dr. Ana Eigen (HRDS-20).

16. Abstract
The overall objectives of this study were to (1) identify and explore alternative safety data sources and analysis perspectives and (2) demonstrate the potential utility of these alternative approaches in increasing understanding of precipitating events and predisposing factors for crashes occurring on horizontal curves and at unsignalized intersections along rural two-lane roads. Generalized conceptual crash model frameworks were developed, informed by a review of supporting published literature on conceptual crash models and contributing factors, alternative approaches to accident analysis, and the role of constraints in systemic approaches to accident analysis. The frameworks proved useful from several perspectives, including (1) identifying and organizing all factors that influence the likelihood of a crash and defining the event sequences that lead to a crash, (2) providing terminology that will encourage clear communication across accident analysis disciplines as research on crash causation continues, (3) visualizing the nature by which a certain factor influences the likelihood of a crash or by which an event directly causes a crash, and (4) identifying data needs (versus data availability) for studying the precipitating events, system constraints, predisposing factors, and target groups associated with a specific crash type. After marrying the conceptual crash model framework with available data, a study was conducted to determine whether crash causal types, or similar crashes grouped together based on their key precipitating events, could be developed from data, photographs, and narratives developed from detailed, on-scene crash investigations available in the National Motor Vehicle Crash Causation Survey. This was followed by a set of three additional studies primarily focused on alternative safety data sources and analysis perspectives related to predisposing factors. Enhanced data collection and subsequent analysis were demonstrated for three high-priority crash scenarios on rural two-lane roads: "straight crossing path crashes" at unsignalized intersections, combination "control loss/no vehicle action" and "road edge departure/no maneuver" single-vehicle crashes on horizontal curves, and "opposite direction/no maneuver crashes" on horizontal curves. Findings demonstrate that expanding beyond traditional databases used for crash-based evaluations can provide further insight into these crashes. One follow-on analysis in the final part of the study indicated that the alternative approaches to estimating disaggregate measures of exposure, kriging, and quasi-induced demand techniques show some promise and should be considered in future research.

17. Key Words conceptual crash models, rural two-lane, single-vehicle crashes, multi-vehicle crashes	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. http://www.ntis.gov

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 275	22. Price 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 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
		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 (2000 lb)	T
		TEMPERATURE (exact degrees)		
^oC	Celsius	1.8C+32	Fahrenheit	^oF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the 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	average annual daily traffic
AASHTO	American Association of State Highway and Transportation Officials
AR	alcohol related
ATR	automatic traffic recorder
BAC	blood alcohol concentration
CDS	Crashworthiness Data System
CIREN	Crash Injury Research and Engineering Network
CMF	crash modification factors
CSV	comma separated value
CURE	cumulative residual
EMS	emergency medical service
ER	encroachment
FA	drowsy and fell asleep
FARS	Fatality Analysis Reporting System
FD	Failed to look/detect
FHWA	Federal Highway Administration
GES	General Estimates System
GIS	geographic information system
GPS	global positioning system
HSIS	Highway Safety Information System
ISD	intersection sight distance
IVI	Intelligent Vehicle Initiative
KML	Keyhole Markup Language
LA	looked away
LiDAR	Light Detection and Ranging
LTAP	Local Technical Assistance Program
LVS	lead vehicle stopped
MCMIS	Motor Carrier Management Information System
MS	misjudgment of speed
MUTCD	Manual on Uniform Traffic Control Devices
NASS	National Automotive Sampling System

NCDOT	North Carolina Department of Transportation
NDS	Naturalistic Driving Study
NHTSA	National Highway Traffic Safety Administration
NMVCCS	National Motor Vehicle Crash Causation Survey
NOAA	National Oceanic Atmospheric Administration
NPS	National Park Service
PC	point of curvatures
PT	point of tangent
RFIP	Roadside Features Inventory Program
RSAP	Roadside Safety Analysis Program
RSS	residual sum of squares
SCP	straight crossing path
SHRP2	Second Strategic Highway Research Program
SPF	safety performance functions
STARS	Service-wide Traffic Accident Report System
TIA	Traffic Improvement Association
TTAP	Tribal Technical Assistance Program
UDA	unsafe driving acts
UDOT	Utah Department of Transportation
USDOT	United States Department of Transportation
VBA	Visual Basic for Applications
VP	vehicle problems
VTTI	Virginia Tech Transportation Institute
WSDOT	Washington State Department of Transportation

EXECUTIVE SUMMARY

The overall objectives of this research were to identify and explore alternative safety data sources and analysis perspectives, and demonstrate the potential utility of these alternative approaches in increasing understanding of precipitating events and predisposing factors for crashes occurring on horizontal curves and at unsignalized intersections along rural two-lane roads. The project team executed the research through a series of nine tasks occurring across two project phases. The team used three different work plans and a white paper at selected points in the project to identify alternative research directions and select priorities. Following a review of published literature that they used to identify and define relevant terminology and assess the state of knowledge related to models of crash causation, the project team carried out research activities in five main areas:

- Conceptual crash models.
- Safety data availability and quality assessments.
- Precipitating events and causal type groupings.
- Enhanced datasets of predisposing factors.
- Estimating exposure.

The remainder of this executive summary provides key highlights and findings of these research activities, with more detail in the body of this final research report.

CONCEPTUAL CRASH MODELS

The project team developed a generalized conceptual crash model framework, informed by a review of supporting published literature on conceptual crash models and contributing factors, alternative approaches to accident analysis, and the role of constraints in systemic approaches to accident analysis. The generalized crash model framework developed in this research consists of the following components:

- System elements—pedestrians, bicyclists, drivers, vehicles, roadway, traffic, and environment.

- Predisposing factors—specific characteristics corresponding to each system element that have some influence on whether a driving/walking/biking task will be carried out successfully (or unsuccessfully).

- System constraints—policies, restrictions, technologies, and other features related to each system element that guide, warn, or protect a driver/pedestrian/bicyclist and/or enable crash avoidance.

- Precipitating events—types and nature of events and event sequences prior to a crash that start with a “collision course” and can ultimately end with either a collision or collision avoidance.

- Event visualization timelines—visualization of the time and space over which the user and vehicle actions corresponding to precipitating events take place for one or more road users on a collision course.

The project team then extended the generalized conceptual crash model framework to additional layers of detail for three crash types on rural two-lane roads: "straight crossing path" crashes at unsignalized intersections (two-vehicles); "control loss/no vehicle action" and "road edge departure/no maneuver" crashes on horizontal curves (one vehicle); and "opposite direction/no maneuver" crashes on horizontal curves (two vehicles). The project team selected these three crash types, adapted from previous published work on "pre-crash scenarios," following an analysis of 4 years of General Estimates System (GES) data to identify crash type priorities for this research.

The crash model frameworks proved useful from several perspectives, including the following:

- Identifying and organizing all factors that influence the likelihood of a crash and defining the event sequences that lead to a crash.

- Providing terminology that will encourage clear communication across accident analysis disciplines as research on crash causation continues.

- Visualizing the nature by which a certain factor influences the likelihood of a crash or by which an event directly causes a crash.

- Identifying data needs (versus data availability) for studying the precipitating events, system constraints, and predisposing factors associated with a specific crash type.

Future applications of the conceptual crash model framework in studies exploring crashes at a more microscopic level seems promising, and is now possible with the availability of naturalistic driving study data. At least one such Second Strategic Highway Research Program (SHRP2) implementation study is incorporating the framework offered in this report to explore driver behavior and crash causation in the vicinity of closely spaced interchange ramps.

SAFETY DATA AVAILABILITY AND QUALITY ASSESSMENTS

The project team assessed the availability and quality of data on driver, vehicle, road, traffic, and environmental characteristics corresponding to the predisposing factors in the conceptual crash model framework, as well as availability and quality of data on system constraints and precipitating events, across a number of alternative datasets at different levels of detail. First, the team conducted a high-level review of the following databases:

- Fatality Analysis Reporting System (FARS).

- National Automotive Sampling System General Estimates System (NASS GES).

- National Automotive Sampling System Crashworthiness Data System (NASS CDS).

- National Motor Vehicle Crash Causation Survey (NMVCCS).

- Current Highway Safety Information System (HSIS) States (California, Illinois, Maine, Minnesota, North Carolina, Ohio, and Washington).
- Crash Injury Research and Engineering Network (CIREN).
- Motor Carrier Management Information System (MCMIS).
- National Park Service's (NPS's) Service-wide Traffic Accident Report System (STARS).
- Crash and Roadway Data from State of Michigan.
- Crash and Roadway Data from State of Kansas.
- Crash and Roadway Data from Oakland County, Michigan.
- Crash and Roadway Data from State, Local, and Tribal Roads in Wyoming.

The high-level review covered general availability and maintenance of the databases; spatial and temporal coverage; data collection and coding; sampling; injury severity definitions; vehicle coverage; and the ability to either geo-locate crashes in the database or determine in some other way whether or not each crash occurred at an unsignalized intersection or along a horizontal curve on a rural two-lane road. The project team then completed a more detailed review of data from NMVCCS, GES, four HSIS States (Illinois, Minnesota, North Carolina, and Washington), Utah, and NPS. The detailed review covered availability, accuracy, and completeness of data for 14 driver-related variables, 10 vehicle-related variables, 28 road variables relevant to horizontal curves, 23 road variables relevant to unsignalized intersections, 6 environmental variables, and 12 traffic variables. The review also identified approximately 25 additional variables for which data were not available in any of the databases reviewed but that were likely have some influence on safety at unsignalized intersections and/or horizontal curves along rural two-lane roads. It then listed alternative tools, techniques, and data sources for gathering information related to each of these additional variables. The 25 additional variables included available sight distance, access density, cross slope, barrier presence and offset, general roadside character, presence/type/condition of signing and markings, rumble strip presence, presence of queues on intersection approaches, intersection skew, and friction supply, among others. Alternative tools, techniques, and data sources for gathering information related to these variables included laser scanning, field collection, project/contract records, model estimation (based on other available data), spatial databases, satellite imagery, street views/video logs, safety improvement databases, asset specific databases, and others.

PRECIPITATING EVENTS AND CAUSAL TYPE GROUPINGS

As one possible "non-traditional" analysis approach, there was interest in whether it was possible to group crashes occurring at unsignalized intersections and/or horizontal curves along rural two-lane roads together based on common "precipitating events." The idea was based on previous efforts that suggested that summarizing or tabulating data from more than one crash case results in a loss of information, and that drawing causal conclusions prior to creating summaries is more beneficial. In other words, the interactions of circumstances and "causes" are necessary to group crashes and identify countermeasures; circumstances (i.e., predisposing factors) alone are not enough. If successful, results of this type of analysis approach could lead to more microscopic interpretations of known empirical associations (e.g., crash modification factors (CMFs)) and inform more effective development of treatments or countermeasures.

It became clear, however, that information on the type and nature of events and event sequences prior to a crash was not available in any of the traditional or readily available alternative data sources (the timing of this research made the SHRP2 naturalistic driving data an impractical option to include, even though it would have been very promising for this purpose). A preliminary review of narratives and diagrams from hard copy police crash reports indicated that a significant level of subjectivity was necessary to characterize precipitating events, and that even then it was not always possible to do so. The project team therefore conducted a study to determine whether it was possible to develop crash "causal types," or similar crashes grouped together based on their key precipitating events, using NMVCCS data, which contains photographs and narratives developed from detailed, on-scene crash investigations. The team expected the NMVCCS database to provide detailed information on the crash, manner of collision, drivers and vehicles involved, and the crash location. Each case report came with detailed descriptions, sketches, and photographs.

It proved possible to combine location information with detailed crash data from NMVCCS and identify crashes for a specific area and site type. To the best of the project team's knowledge, this study was the first to utilize location information for NMVCCS crashes. The team developed seven causal type groupings for rural two-lane, horizontal curve crashes and three causal type groupings for unsignalized intersection crashes. Most causal type groupings focused on one or more types of driver error, and a significant level of subjectivity was still necessary, particularly for cases where drivers and occupants declined interviews with crash investigators. From a highway and traffic engineering perspective, the role that traffic and roadway factors played in the crash event sequence was of particular interest, but was difficult to determine using NMVCCS data. Therefore, the overall utility and practicality of this approach remained unclear, particularly with emerging datasets and data sources such as naturalistic driving data and other emerging technologies. The remainder of this research project, therefore, focused primarily on alternative safety data sources and analysis perspectives related to predisposing factors. Precipitating events received limited attention in the following way: wherever possible, safety analysis conducted during this research compared results and findings using more traditional crash type groupings (e.g., single-vehicle, multiple-vehicle) with results and findings using pre-crash scenario definitions that take into account specific pre-crash vehicle maneuvers.

ENHANCED DATASETS OF PREDISPOSING FACTORS

The project team executed three studies on enhanced datasets, each assessing the potential of a specific enhanced and robust dataset for increasing what the team could learn about the influence of various predisposing factors on safety. Specifically, the studies estimated and examined parameters that quantify the relationships between expected number of crashes (by type) at a location during some defined time period, and the predisposing roadway, traffic, and weather factors at that location. Based on the previously discussed conceptual crash models, the presence of one or more specific predisposing factors over time (e.g., rainy weather) does not in itself "cause" a crash. It does however, by its presence, have some level of influence on whether or not the driving task will be carried out successfully (or unsuccessfully) driver by driver. Therefore, one would expect that influential predisposing factors are associated with the expected number of crashes at an aggregate level. The focus of the methodology and interpretation was not on the specific parameter estimates themselves, but on demonstrating the construction and/or use of an enhanced dataset of predisposing factors and exploring whether the variables not typically collected appear to have an influence on the expected number of crashes and act as a confounder for variables that are typically collected (and therefore result in possible over- or under-statements of such a traditional variable's influence on safety).

The three enhanced datasets used for this portion of the analysis were as follows:

- Unsignalized intersections along rural two-lane highways in North Carolina and Ohio, built using a combination of State and local crash, traffic, and roadway inventory files, Google® Earth™, Google® Street View™, field measurements, and National Oceanic and Atmospheric Administration (NOAA) data. Predisposing factors in this dataset that are not traditionally available include intersection sight distance; vertical grade; intersection angle; pavement quality; weather patterns; and the presence, type, and condition of various types of traffic control devices.

- Horizontal curves on rural two-lane highways in Washington, built by combining State crash, traffic, and roadway inventory files with a detailed roadside features inventory and NOAA data. Predisposing factors in this dataset that are not traditionally available include horizontal curve characteristics; vertical grade; and the presence, type, and location of various roadside features.

- Horizontal curves on rural two-lane highways in Utah, built using data from an extensive effort by the Utah Department of Transportation (UDOT) to gather, identify, and process detailed information on all above-ground assets and road characteristics along State routes using light detection and ranging (LiDAR). The data collection effort appeared to be the first of its kind executed by a State DOT, and the dataset holds significant potential for safety analysis.

The following three subsections summarize the key findings from the exploratory analysis of building and analyzing these data sources.

Unsignalized Intersections in North Carolina and Ohio: Key Findings

The analysis of the expected number of crashes at unsignalized intersections compared the results of an analysis of "straight crossing path" (SCP) crashes (a pre-crash scenario adapted from previous research) utilizing traditional data supplemented by other sources of enhanced data. Additionally, the analysis compared models linking predisposing factors to SCP crashes to models linking predisposing factors to the expected number of all multi-vehicle crashes, a more traditional crash type grouping. Results showed that variables available from traditional data sources did not fully describe the predisposing factors associated with the expected frequency of the SCP crashes or the multi-vehicle crashes. Parameter estimates for variables only available from non-traditional data sources were influential and statistically significant, indicating that models that do not consider these variables likely suffer from omitted variable bias (i.e., they act as confounders to variables that are traditionally available). These variables included detailed information on traffic control devices (e.g., stop sign number/placement, stop ahead warning sign presence, speed advisory sign presence), geometrics (e.g., approach grade, intersection sight distance, horizontal alignment on intersection approaches), and weather (e.g., frequency of below freezing temperatures, snowfall frequency/amount). Variables with statistically significant parameters included in model specifications developed using the traditional databases were also present and statistically significant in model specifications developed using the enhanced databases, but the magnitude of their estimated effect was different (e.g., the estimated safety benefits of increasing lane widths were larger in the models specified with the enhanced dataset than with the traditional dataset). In addition, the estimated dispersion parameter was nearly 70 percent smaller for the models specified with the enhanced dataset than with the traditional dataset. These findings indicated that researchers should consider developing recommendations or research protocols (similar to what is available for creating high-quality safety performance functions (SPFs) and CMFs) that identify “minimum data elements” for different crash types. The generalized conceptual crash modeling frameworks presented in chapter 3 and implemented in chapter 4 would be useful for informing these protocols. Several positive outcomes of implementing these types of data protocols would be likely:

- Improve the reliability of results for both before–after and cross-sectional road safety studies conducted using traditional analysis methods.
- Improve the effectiveness of emerging analysis methods intended to increase the repeatability of observational road safety study results.
- Allow researchers to consider more refined crash types definitions based on countermeasures of interest.

A gain in model precision offset the reduced sample size that comes with looking at SCP crashes (versus all multi-vehicle crashes). While the sample size was smaller, the project team still observed similar levels of statistically significant variables for this more defined crash type versus all multi-vehicle crashes. This is likely due to a removal of “noise” in the data that comes from looking at crash types that have similar crash-generating processes. Generally speaking, parameter estimates for predisposing factors were consistent in direction and magnitude for models of all multiple-vehicle crashes and SCP crashes.

The project team used several supplemental databases and field collection to create the enhanced datasets for the unsignalized intersection analyses. Field measurements included traffic volumes on the minor routes, intersection sight distances, and vertical grades. Collecting each of these three elements differed in the amount of time and effort required. Obtaining the traffic volumes for minor routes with no average annual daily traffic (AADT) data was the most time-consuming effort, while obtaining intersection sight distance required the most manual effort. The project team used driver vantage imagery (e.g., Google® Street View™, video logs) and Google® Earth™ extensively to obtain supplemental data such as sign locations in reference to intersections and messages on signs. The team obtained these data quickly and accurately, in a desktop environment. Further, driver vantage imagery was available from 2007 to 2014 for most locations, allowing for observation of changes over time. Analysts were able to collect desktop-based data for a significant number of intersections in one day, whereas field-based data collection would have allowed for one or two sites per day. NOAA weather data supplemented the traditional data and the team used an extensive number of land-based stations by relative location to study intersections. The unsignalized intersections were mostly located in central and eastern North Carolina, as well as northwestern Ohio, where the terrain is relatively flat or rolling. For these locations, using the nearest land-based station data was justifiable. However, for locations with mountains and large hills, the terrain may impact the weather patterns such that the study site may not have similar weather to the nearest station.

Horizontal Curves in Washington: Key Findings

Washington undertook an effort to collect roadside features as part of a Roadside Features Inventory Program (RFIP). The State collected the data between 2006 and 2012 and focused heavily on rural two-lane roads. In this portion of the research, the project team used the Washington State Department of Transportation (WSDOT) RFIP data in combination with other roadway databases available through HSIS and WSDOT. The hypothesis was that detailed roadside information is important to study horizontal curve crashes because it directly impacts available sight distance when roadside features are on the inside of a horizontal curve. Roadside design may also influence the sequence of events leading to certain crash outcomes. Without addressing these features, analysis results may falsely “over-attribute” the frequency of certain crash outcomes to characteristics of the roadway. In addition to the roadside inventory, the team also incorporated detailed weather data from the NOAA into the dataset. The team used geographic information system (GIS) mapping software to combine coordinate-based roadside inventories with linear-referenced roadway inventories. The team used the same software to compute average offsets from the centerline to continuous roadside features (e.g., barrier, fence) as well as individual counts of and offsets to other roadside features (e.g., trees) efficiently.

For single-vehicle crashes, analysis results showed very stable findings when comparing parameters estimated using traditional databases, advanced databases, and databases enhanced with non-traditional variables. Statistically significant variables in the model specification developed using the traditional dataset included AADT, lane width, shoulder width, and posted speed limit. Statistically significant variables in the model specification developed using more advanced datasets included AADT, lane width, posted speed limit, and horizontal curve radius. Finally, statistically significant variables in the model specification developed using enhanced datasets included AADT, posted speed limit, horizontal curve radius, vertical curve presence, number of days at 90 °F or more, number of days at 32 °F or less, number of days with more than 1 inch of rainfall, and average tree diameter. The estimated dispersion parameter was only 6 percent smaller for the traditional single-vehicle models specified with the enhanced dataset than with the traditional dataset; it was nearly 30 percent smaller for the more refined single-vehicle crash type definition (i.e., combination “control loss/no vehicle action” and “road edge departure/no maneuver” adapted from the previously published pre-crash scenarios). The magnitude of the estimated effects of lane width, shoulder width, and horizontal curvature did decrease as the project team incorporated additional, non-traditional variables into the model, indicating that the effects of these variables on the expected number of single-vehicle crashes may be overestimated in more limited model specifications. Parameters associated with variables built from the WSDOT RFIP (e.g., guardrail presence and length, tree count, tree diameters, fixed object count) were generally in the direction expected for single-vehicle crashes, and were statistically significant for the more refined single-vehicle crash type definition. However, the sample sizes and analysis did not support any direction-specific conclusions related to sight distance restrictions from these roadside features when located on the inside of a horizontal curve.

For multi-vehicle crashes, the model parameters were less stable and more difficult to explain, which is not completely unexpected on lower-volume rural roads when the chances of two vehicles traversing a horizontal curve at the same time are small. The expected number of multi-vehicle crashes was shown to increase as lane width and shoulder width increased. This estimated effect was larger in magnitude as additional variables were included in the model specification. The direction of the regression parameter quantifying the speed limit effect was negative for the multi-vehicle models, indicating a decrease in the expected number of crashes as speed limit increases. It was, however, positive and statistically significant at a higher level of confidence for the more refined multiple-vehicle crash type definition (i.e., opposite direction/no maneuver crashes), indicating the opposite effect of an increase in the expected number of crashes as speed limit increases. The same general findings and recommendations from the unsignalized intersection studies seemed to hold for this study as well:

- Parameter estimates for variables only available from non-traditional data sources were influential and statistically significant, indicating that models that do not consider these variables suffer from omitted variable bias (i.e., they act as confounders to variables that are traditionally available). Similar to results for the unsignalized intersection analysis, these variables included detailed information on geometrics (e.g., curve radius, guardrail presence/length, roadside object counts), and weather (e.g., frequency of below freezing temperatures, snowfall frequency/amount, rainfall frequency/amount).

- Gains in precision offset the reduced sample size that comes with looking at the more refined crash type definitions. While the sample size was smaller, the project team still observed similar levels or higher levels of statistically significant variables for the more refined crash type models.

Horizontal Curves in Utah: Key Findings

The LiDAR-based dataset of horizontal curves from Utah differed the most from other electronically coded datasets that most safety studies have traditionally used (e.g., HSIS, State-maintained inventory databases). These differences originated from the fact that the project team employed a new technology for collecting and processing the data. The project team also incorporated geolocations into every element of this dataset. This offered the opportunities of experimenting, transitioning, and adopting spatial referencing systems instead of the linear referencing approaches (i.e., route and milepost) for safety analysis on a more widespread basis. The transition from linear-based referencing to geo-coding for building safety analysis datasets is likely to become more prevalent as datasets become more robust. It was important to distinguish that the primary design of the dataset from Utah was for asset management purposes, with input and funding from a number of UDOT divisions who also plan to use the data (e.g., traffic and safety). Many important data elements were not directly or readily available. As State transportation agencies gain experience with these datasets, safety-specific processing of original point cloud data (in terms of features needed and level of accuracy) is likely necessary to realize the benefit of LiDAR data fully from a safety perspective.

Implications of Exploratory Research on Enhanced Datasets of Predisposing Factors

This research on building enhanced datasets of predisposing factors has potential implications for researchers and safety practitioners, including those in State and local agencies. It has demonstrated that there is value to using data beyond traditional data sources for safety evaluations. The implication of this for researchers is that crash-based studies should attempt to look beyond the readily available data and consider other sources of data. Relying solely on traditional sources may lead to inappropriately attributing causation to the limited set of variables in crash, roadway, and traffic databases (i.e., causation between predisposing factors and expected crash frequencies). Although the variables in these databases can be explanatory, they are not exhaustive in their ability to explain expected crash frequencies, and may not be reliable predictors of future crash occurrence. Some supplemental data could already be available (with some post processing) in databases that were collected and built for other purposes (e.g., asset management). While it is likely that only a small number of data elements from these supplemental or alternative data sources are useful for road safety study, they could potentially have a significant impact on the ability to predict crash frequencies and the reliability and transferability of such predictions.

The implications for this effort for State and local agencies relate to the data that the team collected. Although agencies have limited funds to expand the data in their data collection, an agency may want to consider collecting additional elements (e.g., skew), as efforts such as this and others demonstrate the value of these elements to informed decisionmaking. Potentially even more critical than the expansion of elements is the ability to integrate the data that agencies have already collected with other sources of data, or to leverage it for other uses. Communication between different organizational units within public agencies about how data are or will be used could, for example, maximize the ultimate utility of data-related investments.

What this project does not completely answer is how one assesses the “value added” by this additional data or the value of integrating these data with other sources. The enhanced data collected in this effort required a reasonable amount of effort to collect, particularly given the additional insight that the project team gained. However, there is a need for additional research to quantify explicitly the modeling benefits and related decisionmaking benefits, as well as the costs and “points of diminishing returns” associated with using different non-traditional data sources. This could lead to effective recommendations or research protocols, similar to what is available for creating high-quality SPFs and CMFs, which identify “minimum data elements” for different crash types. The current state of this research does, however, demonstrate a need for those researching the safety impacts of countermeasures or exploring the precipitating, predisposing, or causative factors of other crash scenarios to be thoughtful in considering expanding their data set to include non-traditional sources of data.

ESTIMATING EXPOSURE

After exploring and assessing multiple traditional and non-traditional data sources, as outlined in the previous chapters of this report, it still proved difficult to find data on traffic patterns at more disaggregate levels than estimates of daily traffic. Commonly, this is true for rural areas, such as the ones this study explores. This lack of information is troublesome because traffic patterns are a key indicator of user exposure to crashes at various times of day. For example, two sites with the same AADT may have significantly different safety performance due to differences in day/night volume distributions. The lack of information on daily travel patterns, as well as suspected uncertainty in daily volume estimates by time of day, remains the “elephant in the room” when analyzing rural road safety. Therefore, the project team executed a final set of studies as part of this research to do the following:

- Demonstrate and assess the use of kriging techniques to estimate day and night traffic volumes at rural, horizontal curve locations in Utah and rural, unsignalized intersection locations in North Carolina.

- Investigate the use of the quasi-induced demand approach, which uses data on crash history and “not-at-fault” drivers to estimate more disaggregate measures of exposure based on traffic volumes at rural, unsignalized intersection locations in North Carolina.

- Explore the use of socioeconomic data as a surrogate for typically unobserved characteristics related to crashes.

The kriging techniques implemented in this study to estimate average annual day and night traffic volumes in rural locations where permanent counters are not available showed promise. The approach was successful in Utah and unsuccessful in North Carolina, the latter likely due to inadequate automatic traffic recorder coverage. There is a tested hypothesis that horizontal curves with higher proportions of traffic at night are expected to experience more crashes than similar curves with higher proportions of traffic during the day. The project team verified this with a “positive parameter” for night-to-day volume ratio in a negative binomial regression model of total expected crash frequency. The parameter estimate, however, was “noisy” and statistically insignificant, most likely attributable to the errors in the kriging predictions. The team offered additional modifications and extensions to the kriging and safety modeling approaches to improve the kriging predictions, and therefore reduce the standard error associated with the predicted night and day volumes. These included the following:

- Incorporating variables in addition to spatial proximity (e.g., functional classification, surrounding area characteristics) into the kriging model.

- Predicting night-to-day ratios directly, instead of night and day traffic volumes.

- Developing separate models for day and night crashes as a function of day and night volumes, respectively.

The quasi-induced demand approach was successful in estimating the percentage of daily volumes by driver age group as a function of either roadway class, day of week, time of day in terms of light conditions (i.e., day/night), and time of day in terms of 6-hour time intervals.

The project team tested the use of these more disaggregated measures of traffic volumes in statistical road safety models of unsignalized intersections using the following specifications of traffic volumes on the major and minor intersection approaches:

- AADT by year for a 5-year period.

- AADT averaged across a 5-year period.

- AADT broken into 6-hour periods by proportions developed from the quasi-induced demand methodology (annual average 6-hour traffic volume). Crashes were aggregated across the 5-year period for the 6-hour window at each intersection (e.g., average AADT for all crashes occurring from 12:00 a.m. to 5:59 a.m.).

- AADT broken into weekday versus weekend by proportions developed from the quasi-induced demand methodology. Crashes were aggregated across the 5-year period for weekdays or weekends at each intersection (e.g., average AADT for all weekend crashes).

Due to differing sample sizes, it was difficult to make generalizations by aggregation level across models. Keeping in mind that sample size can influence model fits, the results indicated the following:

- In all cases, the total entering volume and proportion of traffic on the minor road provided a better model fit than independent AADT values for the major and minor road.

- Aggregation across years provided one observation for each site, which provided minor benefit in terms of model fit. The pseudo R^2 is somewhat improved (larger) and the dispersion parameter is somewhat improved (smaller). The coefficient for total entering volume is consistent and the coefficient for proportion of traffic on the minor road increased substantially.

- Disaggregation of traffic volumes into 6-hour increments provided a much-improved pseudo R^2 of the safety model over the averaged annual model, but the dispersion parameter got larger. It is likely that the sample size negatively affected both the pseudo R^2 and the dispersion parameter.

- Aggregation into 6-hour increments resulted in the proportion of minor road traffic variable being statistically significant (with 95-percent confidence).

- Aggregation of weekday versus weekend AADT resulted in model improvement over the averaged annual model, but was not as good as the 6-hour increment model.

The research was unsuccessful at quantitatively linking socioeconomic data to minor road traffic volumes at unsignalized intersections, which is often missing in traditional datasets. Analysis also indicated that socioeconomic variables were not associated with expected crash frequencies at unsignalized intersections. The project team discontinued any further exploration of socioeconomic data following these findings. In the end, kriging and quasi-induced demand techniques both showed some promise based in these exploratory studies, and future research should consider them in future research, given the importance of traffic volume and exposure in most types of safety analysis.

FUTURE RESEARCH

The conclusion of this report offers ideas for future research that build on this work and cover the following areas: expansion to other crash types and situations, traffic volume data, study design and analysis methodologies, spatial and temporal resolution, and new and emerging sources of data.

CHAPTER 1. INTRODUCTION

BACKGROUND

Nearly 400,000 people have died as a result of traffic crashes in the United States during the last decade. Another 23 million people have been injured. The toll on society is more acute as many of the victims were young and healthy prior to their crash.(2) A societal problem of this magnitude begs for scientific and systematic inquiry. Indeed we have seen the study and practice of road safety steadily transition and advance from decisionmaking based on history and judgment toward decisionmaking based on research, evidence, and technology.(3) Shinar noted the rapid growth in published road safety work, with half as many books on highway safety and driver behavior published in the first 5 years of this century as in all of the previous century.(4)

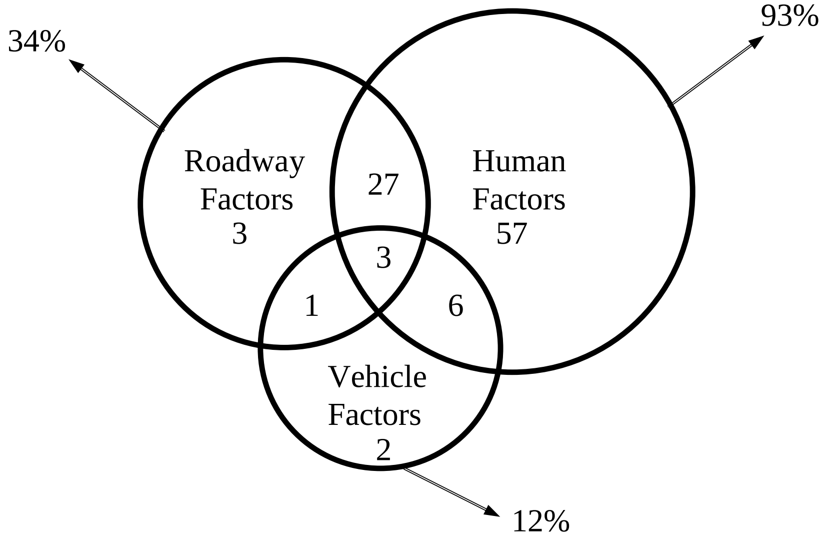
Road safety researchers seek to identify and understand the factors that contribute to the occurrence of traffic crashes in a variety of contexts, as well as to identify related "treatments" (e.g., policies, actions, decisions, design changes, countermeasures, and operational strategies) that reduce both the number of crashes and/or the severity of injuries that result from crashes. The challenge is immense. Crashes are relatively rare events defined by complex sequences of actions, and a range of interacting elements affect their likelihood of occurrence.

The difficulty and complexity of determining why traffic crashes occur, as well as understanding what leads to the severity of injury outcomes, becomes evident when observing the Haddon Matrix in table 1. The first dimension of the matrix, represented in rows, divides a traffic crash into three temporal crash phases: pre-crash, crash, and post-crash. The second dimension, represented by columns, identifies factors contributing both to crash occurrence and to injury outcomes resulting from a crash. First developed in 1980 by William Haddon, the matrix is still widely referenced and considered very useful as it provides information on the range of possible safety issues and provides potential appropriate solutions.(5)

Table 1. Haddon matrix structure.

Crash Phase	Human	Vehicle	Environment
Pre-crash	Age, alcohol speed, drugs, experience, distraction, etc.	Size and weight, brakes, tires, vehicle type, safety condition, etc.	Visibility, lighting, traffic control devices, cross section, etc.
Crash	Belt use, size, seating position, helmet use, human tolerance, etc.	Crash-worthiness, passenger restraints, air bags, etc.	Guardrails, median barriers, maintenance of structures, etc.
Post-crash	Age, physical condition, access to health care, insurance status, etc.	Post-crash fires, fuel leakage, hazardous materials, etc.	Emergency Medical Services (EMS) system, bystander care, traffic management, etc.

In looking at pre-crash events, Treat et al. showed the relative contributions of human, vehicle, and roadway (i.e., environment) factors and their interactions to crash occurrence, as shown in figure 1.(6) The significant role of driver-related factors becomes immediately evident, with “human factors” alone identified as the key contributing factor 57 percent of the time. Human factors interact with roadway and vehicle elements another 36 percent of the time.



Source: FHWA.

Figure 1. Chart. Contributing factors to vehicle crashes based on data from Treat et al.[(6)]

Highway and traffic engineers working for Federal, State, and local agencies make decisions about the appearance of the road infrastructure, the use of traffic control devices, and the implementation of countermeasures and operational strategies (i.e., the “road environment”). Safety is an important consideration in these decisions—many agencies are likely to say the most important consideration—but both table 1 and figure 1 demonstrate the difficulty of accurately and reliably relating policies and decisions about the road environment to likely safety outcomes. This challenge stems from the complex nature of numerous factors that contribute to the occurrence of traffic crashes, as well as from the multiple factor interactions and the difficult-to-observe pre-crash event sequences that lead to a crash, of which the road environment generally plays some part.

Much of the current knowledge about the relationship between the road environment and safety is based on analyses of police-reported crash data and agency-coded roadway inventory data. Researchers tend to use this data, often coded in an electronic format, “as-is,” assuming it is accurate and complete (i.e., the roadway variables readily available in the electronically-coded dataset for a particular jurisdiction are the only variables considered during the analysis). Researchers then use a number of observational study designs and statistical analysis techniques to estimate relationships between the road variables and the expected number of crashes (e.g., see FHWA’s *A Guide to Developing Quality Crash Modification Factors*).[(7)] The result of this general approach to road safety research is a wealth of published statistical associations between road characteristics and safety. Knowledge on crash causation, including the road environment as one of many factors in the causal chain, is limited. Unlike fundamental, causal relationships, the signs and magnitudes of statistical associations may be an artifact of the database used for analysis. Data accuracy and completeness vary between jurisdictions, and few existing roadway databases contain all measurable variables. Therefore, the predicted outcomes of statistically estimated models do not always transfer from location to location. One of the more telling examples of this is evident in a review of literature with statistical associations between crash frequency and speed. Two separate literature reviews on speed and safety, citing 73 and 65 published sources, respectively, concluded that "although evidence tended to support the notion that accident risk increased with speed, more study was needed to determine when changes in speed limit affect accidents or to predict the sizes of these effects."[(8–11)]

Another example is roadway cross-section design, where there have been numerous studies on the safety effects of lane and shoulder width. Some earlier studies indicated that increased lane widths and shoulder widths were associated with higher crash frequencies; however, there was little control for confounders, and results likely reflected the effects of other variables.[(12–14)] Perkins concluded that lane and shoulder width have no effect on the expected crash frequency on rural two-lane roads, but did not exclude intersection crashes from the analysis.[(15)] While many studies showed that crashes tend to decrease as lane width and shoulder width increase, they did not account for all potential confounding factors or the interaction effects of lane and shoulder width.[(16,17)] More recent studies have attempted to control for additional variables, confounding effects, and interactions, and concluded that the effect of lane width is not independent of shoulder width and may not be linear.[(18,19)] After much work, the relationship between cross-section design and safety still seems elusive.

New data sources and emerging data collection technologies are becoming increasingly available. Both have the potential to effectively support the building of more robust road safety databases and the modeling of the "crash-generation process," including what role the road environment played in the process. These datasets will make possible research approaches that hold considerable promise for increasing what is known about how, and to what level, driver, vehicle, and roadway factors influence crash occurrence, as well as the nature of specific events and event sequences that lead to crashes. However, these new data sources and corresponding research approaches will require significant time investments to explore and implement fully.

POTENTIAL IN NEW DATA SOURCES

Crash-based traffic safety research has traditionally focused on analyzing readily available crash, roadway, and traffic data, generally obtained from one or more State databases. The emergence of new data sources and the general availability and accessibility of the data provide a potential opportunity to uncover additional safety insights by constructing more enhanced and robust databases. Several data sources are available as potential alternatives to fill some of the data element needs identified in the conceptual models. Promising alternative data sources include, but are not limited to, publicly available images, project-level and asset-specific databases, enhanced roadway inventories, and naturalistic driving study (NDS) data. The following is a discussion of each of these alternatives as well as other relevant data sources, focusing on the opportunities. Note that there are limitations to each of the data sources as well.

Publicly Available Images

Consumer geoinformatics and other tools that provide online images—such as the United States Geological Survey Seamless Server, Google® Earth™, Bing™ Maps, Google® Street View™, and transportation agency video logs—are publicly available and are free to use. Quality imagery, from multiple vantage points, allows agencies and researchers to visit sites virtually from their desktops and to collect relevant information more efficiently and more cost effectively than visiting each location in the field. This is particularly useful for national studies that may include hundreds or thousands of locations. Aerial and satellite images provide a plain view of locations, allowing the user to identify the presence or absence of features (e.g., turn lanes at intersections) or to assess the number of features within a given space (e.g., number of driveways along a roadway segment). Additionally, aerial images allow for measurement, with relative accuracy, of features such as travel lane width or distance to an object (e.g., distance from an intersection to an upstream intersection warning sign), and horizontal curvature. Driver vantage imagery—e.g., Google® Street View™ or agency video logs—allows the user to virtually visit the site from the standpoint of driving in a vehicle. This provides users with ways to identify the presence of small objects such as raised reflective pavement markers, or to identify messages on traffic signs such as the posted speed limit. These images allow the researcher both to verify the accuracy of the data in the roadway inventory and collect missing elements. Additionally, systems such as Google® Earth™ provide a historical record by allowing access to multiple years of imagery from the same location. This is important for studies that involve a time component such as before/after evaluations or a cross section observed over time (e.g., panel data). Researchers can use this imagery to ensure that there have been no additional changes to the site in the time period of interest.

Project-Level and Asset-Specific Databases

State and local agencies regularly utilize other databases that safety analyses have not traditionally incorporated. Project and contract records, asset-specific databases, and earthmine are examples of non-traditional datasets that some agencies collect and maintain. Project and contract records allow an analyst to identify when certain features were changed, upgraded, or removed. Current safety research methods do not always explicitly consider and incorporate changes—other than the treatment of interest—when estimating the effect of a safety treatment, particularly if those changes are not readily available in existing datasets. For example, projects adding superelevation to a curve or realigning a horizontal curve may change roadside slopes. If analyzed in isolation, estimating the effects of adding horizontal curve superelevation may be misleading.

Asset-specific databases allow cross-sectional studies to consider safety-related assets. For example, several States maintain traffic sign inventories. Matching chevron signs to horizontal curves allows an analysis to consider the safety effect of chevrons. Additionally, some agencies are collecting data through private companies, such as earthmine, to manage assets and create inventories. However, safety analyses can also use the high-resolution photos, allowing for accurate measurements in three dimensions from a street view perspective.

Enhanced Roadway and Roadside Inventories

Transportation agencies use roadway inventory databases to provide basic information they need for effective road and system planning, management, and operations. Features included in these inventories differ by agency, along with data collection methodologies. It is not uncommon for many features to be manually inspected, measured, and coded with varying levels of quality control checks. Some agencies have invested in collecting more detailed inventory information than is typically available in traditional roadway inventory databases to better inform their asset and safety management activities. Examples of these enhanced inventories include data collected with mobile Light Detection and Ranging (LiDAR) units, horizontal curve inventories, intersection inventories, and roadside inventories.

LiDAR

Researchers are exploring mobile LiDAR, combined with a variety of other sensors, as a way to collect detailed roadway and roadside data to support asset management. The Utah Department of Transportation (UDOT), for example, is executing a Roadway Imaging and Inventory program. Data collection vehicles equipped with a LiDAR sensor, a laser road imaging system, a laser rut measurement system, a road surface profiler, and a position orientation system are driving the entire 6,000+ center-lane mi of State routes and interstates in Utah in both directions. Contractors are executing data gathering, post-processing—which includes calibration and measurement of specific data elements (e.g., total paved width, horizontal curvature)—and data delivery. The data are openly available to the public through UDOT's UPLAN planning network platform.[20]

Curve Inventories

The prevalence of crashes on horizontal curves, particularly more severe crashes, has focused attention on this roadway feature. Horizontal curve properties, however, are usually not available in most roadway inventory databases. A selected number of agencies have developed inventories of their curve locations, including information such as curve radius, degree of curvature, and the presence and type of traffic control devices. In some States, the development of these curve inventories occurs during an analysis of curve crashes through efforts such as the FHWA Focused Approach to Safety for Lane Departure Crashes, and the resulting inventory is a subset of the curves in the State. In a few States such as Iowa, the State has undertaken an effort to inventory every curve by scanning geographic information systems (GISs) and aerial imagery. The inventory provides more of a census of curves in this case. In Illinois, Ohio, and Washington, a curve inventory is a component of their existing roadway inventories and the data were developed from construction drawings and straight-line diagrams. These inventories have been the only readily available source of horizontal curve information for quite some time and as such, researchers have used them on many occasions while studying horizontal curve crashes. All three States are included in the Highway Safety Information System (HSIS).

Intersection Inventories

Similar to horizontal curve inventories, several agencies have been working on developing detailed at-grade intersection inventories. In 2012, as part of the Model Inventory of Roadway Elements Management Information System effort, FHWA sponsored the development of an intersection inventory for the States of New Hampshire and Washington. The project demonstrated that it was possible to develop an inventory inexpensively by building on the node system in each State's GIS mapping. Intersection inventories built in this way can include basic geometry (e.g., number of legs, number of approach lanes, presence of turn lanes, and skew) collected from aerials, and existing traffic control information (e.g., approach speed limit, advance warning signs, number of signal heads, or stop signs) collected from driver vantage imagery. The inventories can also include more detailed information that is important to the driving task, such as intersection sight distance and vertical grade. Currently, it is necessary to collect both of these in the field for inclusion in the inventory. Intersection inventories can also be an important component of an asset management system, helping an agency to manage the maintenance of their infrastructure. This is particularly helpful for signalized intersections, which often include thousands of dollars of equipment and require continuing power and frequent maintenance costs.

Roadside Inventories

Approximately 20 percent of traffic fatalities occur when a vehicle leaves the road and strikes a fixed object such as a tree, utility pole, or traffic barrier.(21) The types, sizes, and positions of these different roadside objects play a significant role in the safety performance of a road segment, but traditional safety databases rarely include this information. Roadside hazard rating—a seven-point qualitative scale that characterizes the road environment—has been used in some cases, and it continues to be one option with the widespread availability of driver vantage imagery discussed earlier.(16) However, analyst judgment is still involved, and it is possible for experienced researchers to assign different hazard ratings for similar roadside environments. Washington State Department of Transportation (WSDOT) collected roadside features as part of a Roadside Features Inventory Program (RFIP). The State collected the data between 2006 and 2012, and focused more heavily on rural two-lane roads than on areas with higher numbers of collisions. WSDOT collected the features at a high level of disaggregation, and included cable barrier, concrete barrier bridge rail, ditches, guardrail, trees, walls, and other fixed objects. Offsets from the roadway are not readily available, however, and this may explain what appears to be the relative underutilization of this rich dataset from a safety analysis perspective to-date.

Naturalistic Driving Study

The Strategic Highway Research Program (SHRP2) NDS was designed to research the role of driver behavior and performance in traffic safety.(22) The SHRP2 NDS data consist of more than 3,500 drivers making more than 5.4 million individual trips, with more than 1,400 crashes and 2,700 near crashes among other event categories (e.g., crash-relevant event, non-conflict, and baseline driving). The dataset provides driver data, vehicle data, roadway data, and driving performance data. Driving performance data includes data collected from vehicle sensors, video data, global positioning system (GPS) location data, and driver-initiated audio recordings. Data accessibility varies for non-identifying data, potentially identifying data, and personally identifying data. However, all data included in the dataset will allow researchers to consider precipitating events in addition to predisposing factors at the individual trip level. Data collection activities are concluding. Recently identified SHRP2 safety implementation projects, sponsored through partnerships between FHWA and the American Association of State Highway and Transportation Officials (AASHTO), will explore the use of NDS data to develop and implement promising new safety countermeasures.(23)

Other Relevant Data Sources

Weather and climate data are readily available and can be integrated for safety analyses. The National Oceanic and Atmospheric Administration (NOAA) maintains land-based data sources for minimum, maximum, and mean temperatures as well as precipitation data. Additionally, NOAA maintains the National Climatic Data Center, which provides the land-based data among other data elements for regions in the U.S. Regional data available include temperature, precipitation, degree days, crop moisture index, residential energy demand temperature index, air stagnation index, U.S. wildfires, U.S. wind climatology, apparent temperature, heat stress index, drought index, snowfall, and snow depth amounts. These data allow crash-based analyses to consider weather and overall climate at an aggregate level. They do not provide researchers with detailed knowledge about weather at an individual event level.

Social media has vast amounts of user-generated data that can be made available for data mining. Data mining of social media can identify textual representations of safety-specific concepts as the first step toward the semantic interpretation of crash causality. Many of these platforms provide some level of built-in statistics of account holders, including age, gender, general location, and other useful characteristics. “Tweets” or other messages can contain information about the nature of witnessed crashes, traffic, driver mood/mental state prior to a crash, and any type of driver disability or impairment. Phrase recognition algorithms are available for application to this type of data mining.

OBJECTIVE

The overall objectives of this effort were as follows:

- Identify and explore alternative safety data sources and analysis perspectives.
- Demonstrate the potential utility of these alternative approaches in increasing understanding of precipitating events and predisposing factors for crashes occurring on horizontal curves and at unsignalized intersections along rural two-lane roads.

Specifically, the project team explored methods to understand how, and to what level, human, vehicle, roadway, and environmental elements contribute to crash occurrence and what opportunities exist to identify and understand more effectively the direct and indirect effects of these elements using enhanced data from various traditional and non-traditional data sources.

CHAPTER 2. LITERATURE REVIEW

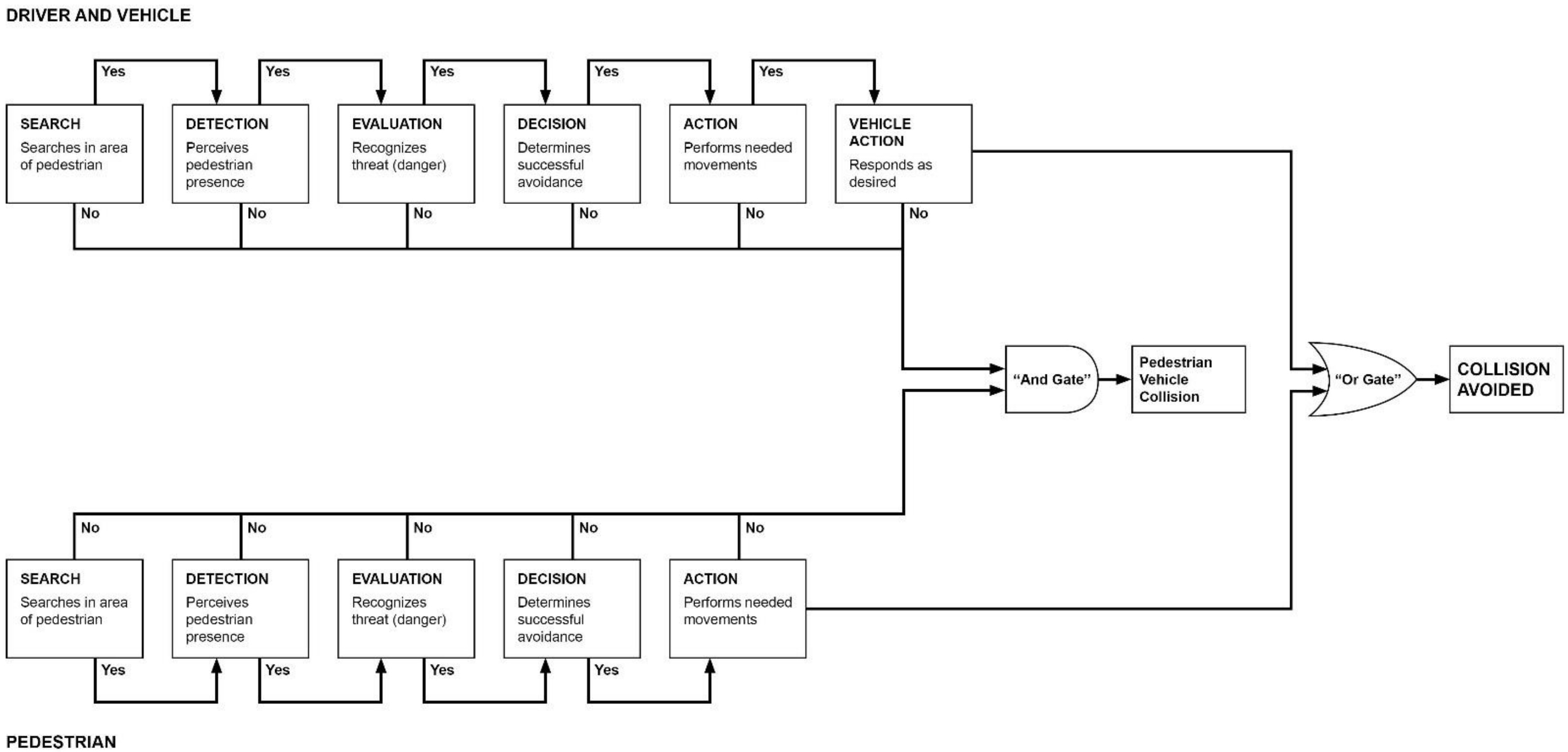
The project team conducted a literature review to evaluate the state of related published information, define relevant terminology, and review supporting literature on conceptual crash models and contributing factors. This chapter presents the results of that review.

SUPPORTING LITERATURE FOR DETERMINING CRASH CAUSAL TYPES

Snyder and Knoblauch identified causes and countermeasures relevant to pedestrian accidents in urban areas.(24) The analysis approach developed for their study stretched beyond the traditional, descriptive approaches to address the following two key needs:

- Establishing crash causation.
- Classifying crash cases for countermeasure identification.

The approach represented a combination of “clinical” and “data association” approaches. The researchers developed a generalized behavioral sequence model to examine the events leading to vehicle–pedestrian crashes and the factors influencing those events. The model includes basic functions (search, detection, evaluation, decision, human/operator action, and vehicle action) that begin after a vehicle and pedestrian are on a collision course, as shown in figure 2. Once either the driver and vehicle or the pedestrian has not performed one function in “the sequence” adequately, they will not perform the following functions adequately either. The driver and pedestrian only avoid the collision when either the vehicle or pedestrian sequence is completed adequately. The author considers the failure of any function or poor performance resulting in too much time delay to complete the sequence a precipitating factor. The behavioral sequence model developed in the study also incorporated predisposing factors: driver, pedestrian, vehicle, and environmental factors that influence the function/event sequence.



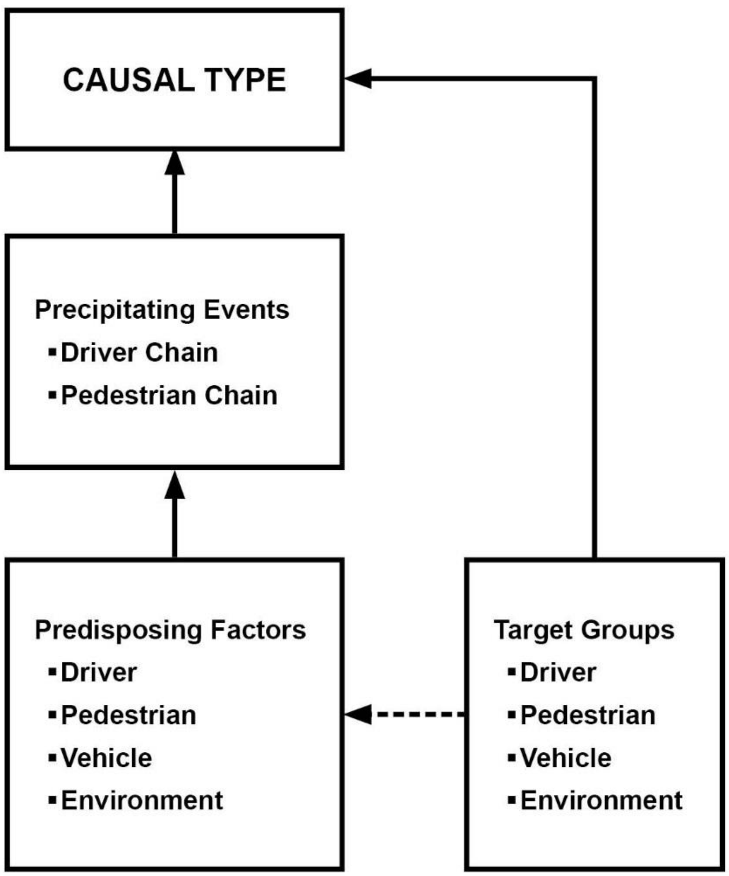
Source: USDOT.

Figure 2. Flow chart. Generalized function/event sequence.[(24)]

The author also developed a causes and countermeasures model to identify causes at a detailed and specific level. The model assigns each pedestrian crash a causal type based on three sets of factors, as shown in figure 2 and figure 3:

- Precipitating events—the specific nature of the failure in the function/event sequence that led to the collision.
- Predisposing factors—specific environmental, human, or vehicle variables that influenced the function failure.
- Target groups—human populations and/or kinds of physical locations involved in the crash type.

Predisposing factors are distinguished from target groups by actually leading to the function failure; target groups only have an association with the involvement.

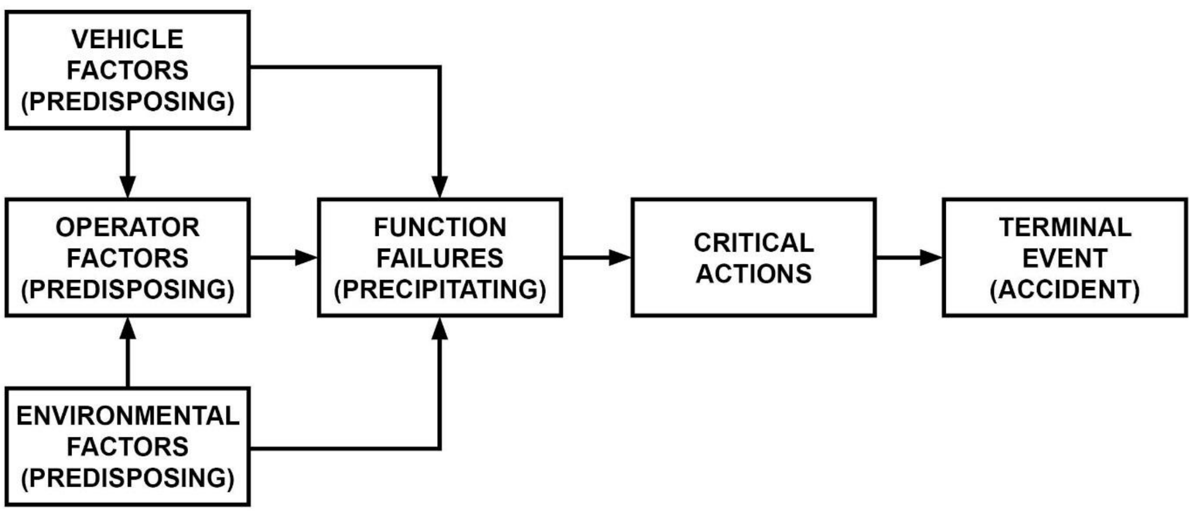


Source: USDOT.

Figure 3. Flow chart. Aspects of causal type.[(24)]

Cross and Fisher conducted a study to determine the causes of bicycle/motor vehicle crashes.[(24)] The work of Snyder and Knoblauch had a significant influence on the study methodology.[(23)] The objective was to identify and study vehicle–bicycle crash problem types, with the idea that crashes of the same problem type should be amendable to the same, specific countermeasures. The study developed conceptual representations of the crash process, including a model of the crash-generation process and a behavior sequence model. The study developed the latter because the authors expected a large portion of functional failures to be behavioral. The framework was based heavily on the behavioral sequence model developed by Snyder and Knoblauch, with more emphasis on collision course selection, and with the addition of anchor points to consider the time dimension more explicitly.[24] The study divided a trip into three phases to gain insights to the collision course: preparatory, anticipatory, and reactive. It defined anchor points at five points along each operator/vehicle's path to assess operator timeliness of performing the functions.

The study developed conceptual representations of the crash process, including a model of the crash-generation process and a behavior sequence model. The model of the crash-generation process developed for this study appears in figure 4. It includes the terminal event, critical actions, function failures, and predisposing vehicle, operator, and environmental factors.



Source: USDOT.

Figure 4. Flow chart. Conceptual model of the crash-generation process.[24]

The terminal event is the crash. It gives no insight about crash causation, but details of the crash are necessary to assess crash consequences.

Critical actions are vehicle actions and movement patterns that led directly to the crash. Critical actions are the ultimate target of countermeasures. Contextual data are necessary to understand why the actions proved critical in terms of the crash.

Functional failures are when operators, vehicles, and environmental elements do not perform as expected. The functional failures have a causal relation to the critical actions. Operator failures are behavioral acts that the operator should not have performed, or should have performed but did not. Vehicle failures occur when vehicles do not perform in a typical way. Similarly, an environmental failure occurs when some part of the environment, such as a traffic signal, does not perform as expected. It is important to distinguish between functional failures and predisposing factors. For example, operator characteristics that directly influence operator behavior are predisposing operator factors. Vehicle design and roadway design characteristics that directly influence the vehicle and environmental factors also fall under predisposing factors. The operator often mediates the influence of the vehicle and environmental factors on functional failures.

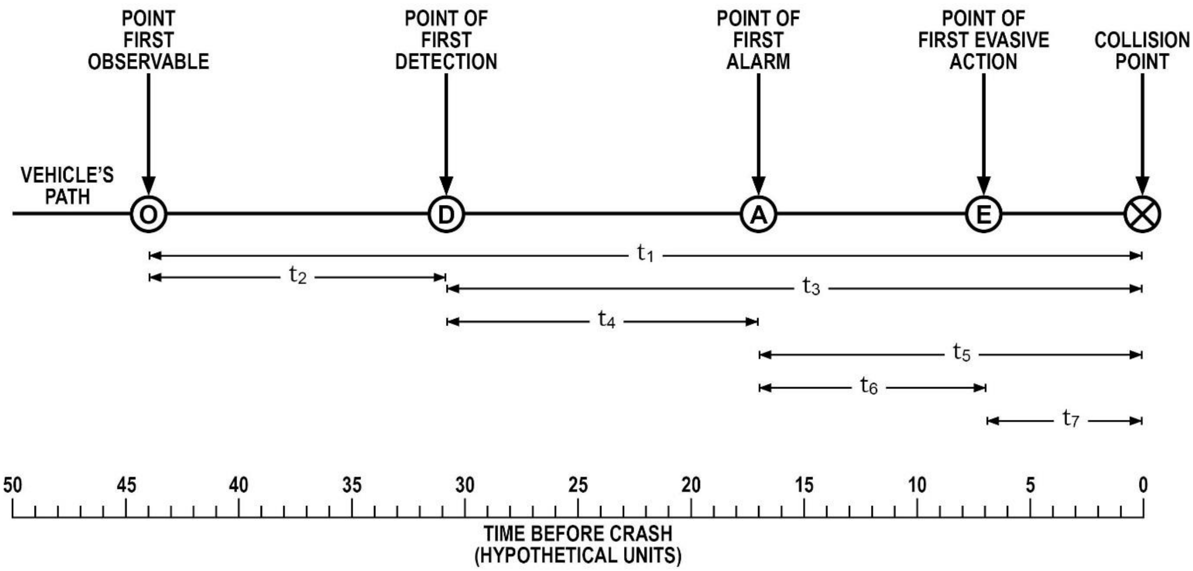
The study developed a framework to identify and define behaviors that are operator failures as part of a behavioral sequence model, because the authors expected a large portion of functional failures to be behavioral. The framework was based heavily on the behavioral sequence model developed by Snyder and Knoblauch, with more emphasis on collision course selection and with the addition of anchor points to consider the time dimension more explicitly.[(24)] The study divided a trip into three phases to gain insights to the collision course: preparatory, anticipatory, and reactive.

- The preparatory phase spans the actions from the time the operator decides to execute a trip to the point where the operator selects a course (i.e., a vehicle path and speed) through the crash area. Evaluation failures and decision failures may occur during this phase.

- The anticipatory phase begins when the operator selects a course through the crash area and ends at the point where they first could have observed the other vehicle. The anticipatory phase includes search, detection, evaluation, decision, operator action, and vehicle action. When the selected course is suboptimal, the anticipatory phase failure is either predisposing (if an operator reaction to avoid a collision is possible) or precipitating (if an operator reaction to avoid a collision is not possible).

- The reactive phase begins when the operator first could have observed the other vehicle and terminates at the collision point. This phase is closely modeled after the behavioral sequence model developed by Snyder and Knoblauch, and also includes search, detection, evaluation, decision, operator action, and vehicle action.[(24)] The two operators avoid the crash if at least one of them performs these functions adequately.

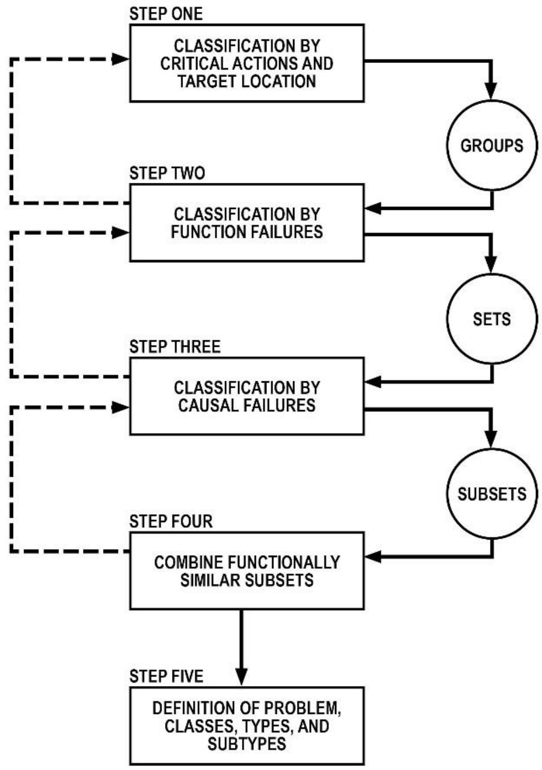
Anchor points at five points along each operator/vehicle’s path, collision point, point of evasive action, point of first alarm, point of first detection, and point other vehicle is first observable, were defined to assess operator timeliness of performing the functions, as shown in figure 5. The author compares times between different anchor points for the operator of interest to times needed for a normal operator to identify possible function failures in the time dimension.



Source: USDOT.

Figure 5. Illustration. Anchor points for assessing time available and time used to perform behavioral sequence functions.[(25)]

The author classified individual accidents into a set of mutually exclusive problem types. The classification is based on variables causally related to the crash (figure 4). Crashes grouped only by descriptive variables (e.g., vehicle type, road type, operator's age) may not be common in terms of the crash-generation process (and ultimately countermeasures). The study investigated each crash in detail. For each crash case, the study identified variables causally related to the crash (i.e., factors and function failures). It grouped crashes that show common causal variables together. Field investigators and the principal investigator identified the causal variables. They developed groups based on commonality in one or more of the following: traffic context, operator function failures, and combinations of factors causally related to the function failures. The degree of commonality needed was based on the specificity of countermeasures. The grouping process is illustrated in figure 6.



Source: USDOT.

Figure 6. Flow chart. Crash classification procedure.[(25)]

Hendricks et al. conducted a study to determine specific unsafe driving acts (UDAs) that lead to crashes and the situational, driver, and vehicle characteristics associated with these behaviors.[(26)] The study classified behaviorally-caused crashes into problem types based on common characteristics. It used data from the National Automotive Sampling System Crashworthiness Data System (NASS CDS) with supplemental variables that the NASS field investigators and members of the project team added for analysis. For each different crash type, the study used the collected variables to develop a series of profiles of driver actions, attributes, and crash conditions. The study computed the relative involvement index for each level of each profile variable to assess over- or under-representation for the crash type relative to all crash types combined. These results served as inputs to the subjective determinations of causal factors and UDA assessments.

The six most commonly assigned causal factors were as follows:

- Driver inattention.
- Vehicle speed.
- Alcohol impairment.
- Perceptual error.
- Decision errors.
- Incapacitation.

The study identified problem types following this general analysis of all crashes. It redefined crashes into seven crash type classes with operational differences likely to be associated with driver behavior. Disaggregating these crash types based on six key profile variables then created problem types. The key profile variables were crash cause, driver impairment, primary behavior source, necessary UDA, travel speed, and first UDA in sequence. The study then identified demographic and behavioral characteristics, as well as crash descriptors associated with these problem types, using the relative involvement index.

The study created profiles of driver actions, attributes, and crash conditions for each crash type. It then defined problem types by disaggregating the crash types based on six key profile variables: crash cause, driver impairment, primary behavior source, necessary UDA, travel speed, and first UDA in sequence. Then, it identified demographic and behavioral characteristics, as well as crash descriptors associated with these problem types. The study utilized a quantitative tool, the relative involvement index, during several of these data analysis steps to assess over- or under-representation of specific characteristics for the crash type relative to all crash types combined.

Davis argued that causal models are necessary for effective engineering decisions.[(27)] Three methods to estimate the safety effect of a countermeasure were discussed: statistical analysis, simulation, and crash reconstruction.

Statistical Analysis

The analysis showed a crash reduction factor to represent a property of a countermeasure, but not an expected causal effect of a countermeasure. In this context, the study revisited a property of statistical analysis: it allows identification of associations, with additional conditions necessary to establish causation. The study defined the probability of necessity in this context as the probability that a crash would not have occurred had a countermeasure been present, given that a crash did occur with the countermeasure absent. The study showed the probability of necessity to be the more fundamental quality that a crash reduction factor attempts to estimate. Statistical studies assume little about the underlying accident generation process (i.e., sequence of events).

Simulation

It may be possible to identify a characteristic sequence of events in certain crashes and represent those events by equations and variables (e.g., equations of motion). Simulation assumes the variables are random outcomes and specifies a probability distribution for the variable combinations. The probabilities of events (i.e., crash, no crash) are estimated by integrating across the distribution. The results of simulation are useful in determining aggregate properties of populations of potential accidents. ROADSIDE software, presented in appendix A of the 1996 Roadside Design Guide, and the Roadside Safety Analysis Program (RSAP) are identified as two examples of the simulation approach.[(28,29)] Structural models and simulation demonstrated the process for estimating the reduction factor (i.e., probability of necessity). Results showed the variation in countermeasure across locations depending on the distributions of variables most related to the countermeasure. In other words, the distributions of these variables and their role in the crash-generating process are important in estimating countermeasure effectiveness. Knowing the crash type alone is not enough.

Crash Reconstruction

The objective of crash reconstruction is to determine how a crash occurred by collecting information about an individual crash. An additional objective may be to determine what researchers can regard as one or more causes of the crash. Crash reconstruction is often a part of legal proceedings, and sometimes of safety research, with the Tri-Level study one of the more famous cases of the latter. A Bayesian view of crash reconstruction is presented because there is often uncertainty about the value of some variables studied during a reconstruction. The crash reconstruction expert is the Bayesian agent, with a prior distribution describing his/her initial uncertainty concerning the values of the crash variables. The probabilities that are an outcome of Bayesian crash reconstruction represent a "degree of belief." Structural models and Bayesian accident reconstruction again serve to demonstrate the reduction factor (i.e., probability of necessity) for individual accidents. One of the main crash variables had a close relation to the possible countermeasure in some cases. The Bayesian approach resulted in a high probability that those individual crashes would not have occurred had the countermeasure been in place. No crash variables had a close relation to the possible countermeasure in other cases, even though the crash type was the same. Results showed a low probability that the individual crash would not have occurred had the countermeasure been in place in those cases.

Davis proposed the individual crash as the basic level for causal modeling in safety research. The author explored a measure of causation—the probability of necessity—in the context of estimating the safety effect of a countermeasure.(27) The author showed it was possible to identify a characteristic sequence of events in certain crashes and represent those events by equations and variables (e.g., equations of motion). Structural models and simulation demonstrated the process for estimating the reduction factor (i.e., probability of necessity). The author identified AASHTO's ROADSIDE and RSAP as two examples of similar simulation approaches. Results showed the variation in countermeasure effectiveness across locations depending on the distributions of variables most related to the countermeasure at those locations. In other words, the distributions of these variables and their role in the crash-generating process are important in estimating countermeasure effectiveness. Knowing the crash type alone is not enough.

Davis also presented a Bayesian view of crash reconstruction that accounts for uncertainty in the value of some variables studied during a crash reconstruction.(27) He used an example of vehicle–pedestrian collisions in which crash reconstruction served to estimate speeds of vehicles involved in pedestrian collisions. He then asked how many vehicle–pedestrian crashes would have been avoided had all vehicles obeyed the existing speed limit. Davis first defines an impending vehicle–pedestrian collision scenario with key variables (e.g., initial speed of vehicle, distance driver detects pedestrian, perception/reaction time, braking transition times, skid mark length). He then combines prior distributions for these variables with crash-specific measurements to estimate posterior estimates of initial vehicle speed, detection distances, and impact speeds. Results quantitatively showed how the effectiveness of a countermeasure on the same crash types varied, depending on whether or not variables that have a close relation to the possible countermeasure were present in the crash sequence.

DIFFERENT APPROACHES TO ACCIDENT ANALYSIS

Often, researchers use one of three approaches to analyze and diagnose accidents clinically: sequential, epidemiological, and systemic.[(30)] Sequential accident modeling—sometimes called linear or "root cause" modeling—was the first approach used in the evolution of accident analysis. Sequential accident models are based on linear thinking. There is the assumption that events happen one after another with the cause–effect relationships among them. A straightforward way to describe unexpected outcomes is that, if an accident occurs, there must be a reason for it somewhere in a chain of events. Researchers usually compare root cause modeling to the "domino effect" concept, because it focuses on a single component failure that takes the blame for the collapse of the entire system. This way of thinking is not unreasonable for certain types of crashes when the cause is straightforward to identify and understand. Driver–vehicle–road systems are, however, not straightforward to understand, so the linear way of thinking may no longer be applicable.

Researchers developed epidemiological crash models to overcome the shortcomings of linear thinking in the late 1970s. This approach recognized that serious events may involve multiple event sequences occurring in series and in parallel. Sequences can be interactive and complex, so instead of causal series, this approach represented crashes as causal nets. While linear models applied the "everything happens for a reason" logic, they did not separate the events' order in time from the assumption of a causal relationship between them (i.e., if one event followed another event, the latter must be an effect of the former). Epidemiological models instead describe an accident using the "spreading of a disease" logic, where an outcome results from the combination of factors related to the human factors, technology, and the environment. Epidemiological models recognize that system failure could occur, not only due to human error, but because of performance deviations that come naturally for every system. Even though these models use causal networks rather than an event series, they still use the same principle of propagation of effects from the beginning to the crash outcome.

Systemic accident models observe a system as a whole in order to describe its performance. These models analyze crashes as emergencies, rather than structurally decomposed into components and their associated functions. Crashes still happen for a reason, but not with the simplistic progression or serial/parallel progression suggested in sequential models or by the epidemiological approach, respectively. The crash timeline still remains, since events do occur in order and it is possible to trace their development in time. However, the concept of emergence is based on impacts with more than one direction. Crashes occur due to system performance variability, which is different from human error in sequential models or inherited deviation in epidemiological models. The advantage of systemic models is that they look to understand crashes through the functional characteristics of the system, rather than assuming internal mechanisms or failure pathways. A crash is neither a causal series nor a causal net. It is an emergent result of dynamic interactions of nonlinear effects. This is why these models are very difficult to represent graphically.

The approach used to develop conceptual crash models in this section was based on the systemic approach to accident analysis. The researchers modified the systemic approach to make visualization of the models more practical. They used them to identify different combinations of events that might lead to similar crash outcomes.

THE ROLE OF CONSTRAINTS IN SYSTEMIC APPROACHES TO CRASH ANALYSIS

Crash analysis based on the systemic approach accounts for system inputs, resources, controls, preconditions, outcomes, and time. The systemic approach analyzes crashes looking for the set of probable causes that together constitute a satisfactory explanation. Another definition of a crash is a set of constraints that have failed. Crashes are viewed as a control problem, rather than an event problem.[(31)]

Control in the systemic approach is associated with the imposition of constraints, and preventing future crashes requires designing a control structure that will enforce the necessary constraints. Crashes result from interactions among components that violate the safety constraints or the lack of appropriate control actions to enforce those constraints. The failure of constraints alone cannot in itself be a cause of a crash, because the constraints exist to prevent the crash from happening, but it may have an effect on the further development thereof.[(30)]

Related to the conceptual models presented in this paper, constraints should exist for each element involved in the crash: driver, vehicle, roadway, traffic, and environment. Constraints coming from driver control could include types of enforcement, education, and training. Vehicle-related controls could include safety checks, vehicle maintenance procedures, and automatic safety related performance systems (e.g., collision warning, braking). Constraints imposed by roadway controls could be related to roadway maintenance traffic control devices (e.g., posted speed limit signs, other signs and markings) and other safety-related countermeasures (e.g., rumble strips, high-friction pavement). Environment related controls could include available pre-trip information and the quality of during/post-crash response.

Predisposing factors (e.g., age, income, driving experience) and controls (e.g., signage presence) can affect precipitating events (e.g., fail to detect the threat or respond to it in time), but it will be the inability of driver vehicle environment together to cope with the difficulties (due to lack or violation of constraints) that leads to the crashes, not simply bad weather, or distracted driving or any factor alone. The presence of posted speed limit signs, for example, is not a predisposing factor related to roadway or traffic; it is a constraint placed on the roadway to communicate appropriate maximum driving speed and prevent crashes. While predisposing factors usually have ranges of values, controls are usually there or are not; they are usually in good condition and visible or are not.

Predisposing factors can either increase or decrease the likelihood of a crash, while the purpose of implementing constraints in various forms is to prevent crashes. While there is no way to change or fully control predisposing factors, it is possible to remove ineffective constraints in the short term. Constraints can also be methods used to limit the exposure of certain predisposing factors; examples include minimum driving age, DUI enforcement, and graduated driver licensing laws. This is why researchers should observe constraints separately from predisposing factors, and the following models show separate categories for these two groups of factors.

Terminology

For this framework, crashes are classified into causal types based on three sets of factors:

- Precipitating events—the specific nature of the failure in the function/event sequence that led to the collision.
- Predisposing factors—specific environmental, human, or vehicle variables that influenced the function failure.
- Target groups—human populations and/or kinds of physical locations involved in the crash type.

Predisposing factors are distinguished from target groups by actually leading to (i.e., causing) the function failure; target groups only have an association with the crash involvement.

Data Collection

Data collection for most of the previously reviewed studies on crash causation utilized some form of travel and onsite investigation. The researchers also conducted interviews of people involved in the crash and eyewitnesses. The field investigators supported these activities. For example, Knoblauch hired and trained 40 field investigators for his study of rural pedestrian accidents.(24) Hendricks et al. expanded the number of variables in their datasets with the assistance of specially trained NASS researchers because of a decision to integrate their research into the NASS program as a special study.(26) Directly interviewing people involved in a crash has become less practical since the work in the 1970s due to privacy concerns and human participant rights. For the purpose of this study, the project team gathered variables that enhance existing databases through non-traditional data sources, as further described in chapter 4. These data sources were available in more traditional systems and consisted of tools that are non-proprietary, and which Federal, State, and local agencies accessed.

Conceptual Event Sequence Models

Previous studies demonstrated the usefulness of defining conceptual models of the crash-generation process to determine data needs. The most useful framework was a combination of the general models proposed by Snyder and Knoblauch and Cross and Fisher, that includes the following elements:(24,25)

- Basic functions of search, detection, evaluation, decision, human/operator action, and vehicle action.
- Three trip phases to gain insights into the collision course selection: preparatory, anticipatory, and reactive.
- Anchor points at points along each operator/vehicle's path to assess operator timeliness of performing the functions.

These conceptual models were utilized for the high-priority pre-crash scenarios to guide the determination of data needs.

Determining Crash Cause

The methodologies of previous work ultimately had to incorporate some level of investigator judgment in determining crash causes. This is more consistent with a "degree-of-belief" approach to crash probability than a purely data-driven, frequentist approach to crash probability. Ultimately, a credible approach to identifying precipitating events, crash causes, and "causal types" was necessary for this study. Snyder and Knoblauch classified crashes into causal types based on precipitating events, predisposing factors, and target groups.(24) This framework is utilized in this study. Other ideas drawn from the literature were also useful for determining crash cause. The project team utilized a series of pre-crash scenarios as a starting point for arriving at a causal type (see chapter 5). Hendricks et al. demonstrated the application of a relative involvement index to a large database to assess over- or under-representation of crash profile variables in specific crash types.(26) Davis demonstrated the use of structural models in crash simulation and Bayesian crash reconstruction to link quantitatively variables that are part of a crash-generating process.(27)

PRE-CRASH SCENARIOS

The pre-crash scenario typology developed by the John A. Volpe National Transportation Systems Center (Volpe) describes 37 pre-crash scenarios that represent about 99.4 percent of all light-vehicle crashes.(31) Since these scenarios would become one key part of this research, a brief background summary of their development and use is provided in this section.

The Volpe pre-crash scenarios are based on the 2004 General Estimates System (GES) crash database; it consists of pre-crash scenarios depicting vehicle movements and dynamics, and the critical event immediately prior to a crash. The idea behind this typology is to develop a common vehicle safety research foundation that would help researchers and practitioners in developing crash avoidance systems.

The objectives of this pre-crash scenario typology are as follows:

- Identify all common pre-crash scenarios of all police-reported crashes involving at least one light vehicle.
- Quantify pre-crash scenarios' severity in terms of frequency of occurrence, economic cost, and functional years lost.
- Portray each scenario by crash contributing factors and circumstances in terms of the driving environment, driver, and vehicle.
- Provide nationally representative crash statistics that can be annually updated using national crash databases.

The 37 pre-crash scenarios (including "Other") defined in this typology are given in table 2. The definitions of Vehicle Action and Vehicle Maneuver in the pre-crash scenarios are as follows:

- Vehicle Action refers to a vehicle decelerating, accelerating, starting, passing, parking, turning, backing up, changing lanes, merging, and successful corrective action to a previous critical event.

- Vehicle Maneuver denotes passing, parking, turning, changing lanes, merging, and successful corrective action to a previous critical event.

Table 2. Volpe pre-crash scenarios.

Number	Title
1	Vehicle Failure
2	Control Loss without Prior Vehicle Action
3	Control Loss with Prior Vehicle Action
4	Running Red Light
5	Running Stop Sign
6	Road Edge Departure with Prior Vehicle Movement
7	Road Edge Departure without Prior Vehicle Movement
8	Road Edge Departure While Backing Up
9	Animal Crash with Prior Vehicle Maneuver
10	Animal Crash Without Prior Vehicle Maneuver
11	Pedestrian Crash with Prior Vehicle Maneuver
12	Pedestrian Crash without Prior Vehicle Maneuver
13	Pedalcyclist Crash with Prior Vehicle Maneuver
14	Pedalcyclist Crash without Prior Vehicle Maneuver
15	Backing Up into Another Vehicle
16	Vehicle(s) Turning—Same Direction
17	Vehicle(s) Parking—Same Direction
18	Vehicle(s) Changing Lanes—Same Direction
19	Vehicle(s) Drifting—Same Direction
20	Vehicle(s) Making a Maneuver—Opposite Direction
21	Vehicle(s) Not Making a Maneuver—Opposite Direction
22	Following Vehicle Making a Maneuver
23	Lead Vehicle Accelerating
24	Lead Vehicle Moving at Lower Constant Speed
25	Lead Vehicle Decelerating
26	Lead Vehicle Stopped
27	Left Turn Across Path From Opposite Directions at Signalized Junctions
28	Vehicle Turning Right at Signalized Junctions
29	Left Turn Across Path From Opposite Directions at Non-Signalized Junctions
30	Straight Crossing Path at Non-Signalized Junctions
31	Vehicle(s) Turning at Non-Signalized Junctions
32	Evasive Action With Prior Vehicle Maneuver
33	Evasive Action Without Prior Vehicle Maneuver

Number	Title
34	Non-Collision Incident
35	Object Crash With Prior Vehicle Maneuver
36	Object Crash Without Prior Vehicle Maneuver
37	Other

CHAPTER 3. CONCEPTUAL CRASH MODELS

The hope of road safety researchers and managers is that a more complete understanding of precipitating events and predisposing factors of traffic crashes will lead to identifying why crashes occur (i.e., the crash causes). By knowing these causes, researchers can identify, develop, and implement corresponding "treatments" in more cost-effective ways than if only "noisy" associations are known. Whether it is possible to determine cause–effect relationships from observational studies, the usual study design employed in road safety research, remains an ongoing and important debate. In the context of discussing cause–effect relationships and observational road safety studies, Hauer referred to an anecdote about Sir Ronald Fisher as reported by Cochran:

> About 20 years ago, when asked in a meeting what can be done in observational studies to clarify the step from association to causation, Sir Ronald Fisher replied: 'Make your theories elaborate.' This reply puzzled me at first, since by Occam's Razor, the advice usually given is to make theories as simple as is consistent with known data. What Sir Ronald meant, as subsequent discussion showed, was that when constructing a causal hypothesis one should envisage as many different consequences of its truth as possible... He thought this attitude to be one of the most potent weapons in observational studies.[(33)] (p. 252)

Elaborate theories about the crash-generation process are difficult to construct. As Hauer noted in this same paper, basic research in advancing fundamental road safety theories is not typically sponsored.[(32)] It is possible, however, to construct generalized conceptual crash models, with a goal of identifying all possible known elements that likely interact to either lead to, contribute to, or prevent a crash from occurring. It is possible to use conceptual crash models to clarify where different driver, vehicle, and environment factors "fit into" various stages of the crash-generating process. They could also serve as a framework to identify what data related to the conceptual model elements are needed, where those data can (or cannot) be obtained, and at what level of quality are the data likely to have.

One promising approach to designing conceptual crash models is to base those models on a modified systemic approach to accident analysis that treats accidents as emergent events rather than searching for failures of specific elements. In following this approach, the principles outlined below offered a framework for developing generalized conceptual crash models for any crash type:

- Each scenario was defined by the following system elements or inputs:
    - Pedestrians or bicyclists (if the crash involves pedestrians or bicyclists).
    - Driver(s).
    - Vehicle(s).
    - Roadway(s).
    - Traffic.
    - Environment.

- Each element was characterized as follows:
  - Predisposing factors, or specific characteristics of each element that have some level of influence on whether or not the driving task will be carried out successfully (or unsuccessfully).
  - System constraints, or policies, restrictions, technologies, and other features related to each system element that guide/warn/protect/prevent/enable crash avoidance (e.g., signs, markings, rumble strips, self-correcting vehicle technologies).
- Interactions of predisposing factors and system constraints lead to certain precipitating events that take place in time and space, with the following elements:
  - Precipitating events—the specific nature of events and event sequences that start with a "collision course" and can ultimately lead to a collision. Precipitating events correspond to the following user actions: search, detect, evaluate, decide, and act. The definition of a final "vehicle action" also captures situations where the driver acts, but the vehicle does not respond accordingly.
  - Time and space—can either be a resource (i.e., given driver performance capabilities, enough time following initiation of a collision path to avoid a crash and/or enough space to decelerate/stop or change paths to avoid a collision) or a restriction (i.e., not enough time and/or space following initiation of a collision path to avoid a crash, even if that collision path is detected).
- Nature of user and vehicle actions that define the precipitating events was dual: they either occurred or they did not.
- Nature of the user and vehicle actions and their temporal and spatial characteristics (i.e., time/space requirements versus the time/space available) directly lead to the system outcome.
- Possible system outcomes are as follows:
  - Crash does not occur.
  - Crash occurs.
- Conceptual crash models that are defined using points "A" through "F" have the following dimensions:
  - System elements consisting of predisposing factors and system constraints.
  - Precipitating events that correspond to specific user and vehicle actions and the event outcomes.
  - Time and space dimensions to capture required versus available time and space for events.
- Target groups are as follows:
  - Human characteristics associated with different event outcomes (e.g., age, gender).
  - Physical locations where events are occurring (e.g., signalized intersections, unsignalized intersections, and horizontal curves).

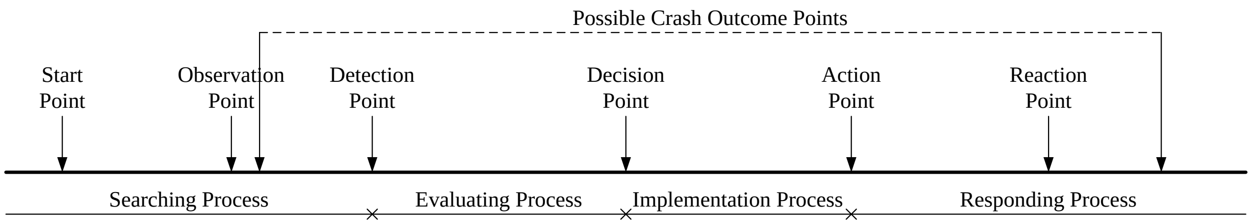
Part C of the generalized conceptual crash models was central to determining crash causes. The project team used event visualization timelines to visualize the time and space over which the user and vehicle actions corresponding to precipitating events take place for one or more road users on a collision course. An example is provided in figure 4.[25] The following definitions correspond to the major search, decision, and evaluation points displayed in figure 7:

- Start point—where the event or combination of events that makes a crash possible occurs (i.e., the time of the initiation of a potential "collision course").
- Observation point—where the driver's (or drivers') observation of the collision course is possible considering object and vehicle characteristics (e.g., driver eye height, vehicle height), roadway geometric features, and other visibility-related characteristics.
- Detection point—where the driver actually detects the collision course.
- Decision point—where the driver decides what the response to the detected collision course will be.
- Action point—where the driver completes the implementation of the decision previously made at the decision point.
- Reaction point—where the driver's vehicle starts to respond to the driver's actions.

The following driver and vehicle processes take place between some these critical points:

- Searching takes place prior to detection.
- Evaluating occurs between detection and making a decision.
- Implementation (of the selected action) occurs between the decision and actual driver action.
- Responding (of the vehicle) after the driver's action.

The nature of the potential collision course can vary significantly depending on the crash type and location of interest. For example, the collision course for a single vehicle approaching a horizontal curve occurs when the current vehicle trajectory/path is different than what will be needed to traverse the horizontal curve (i.e., it begins when the vehicle exits the previous, upstream curve). The collision course for two vehicles on a straight, tangent section of road occurs when the two vehicle paths are such that they would collide if neither operator makes an additional maneuver.



Source: FHWA.

Figure 7. Graph. General example of a scenario timeline with the most important points and processes.

It is clear that some points on the scenario timeline may not even occur due to time constraints (e.g., detection, decision, action, reaction). This study assumed that crash outcome points could occur at any time after the start and observation points, as observation is always feasible for typical vehicle, roadway, and environment conditions, even if only for a split second before the collision. It is possible to find potential observation points for the given location and conditions, depending on the object characteristics, roadway features, and environmental characteristics that impact visibility.

The project team put conceptual crash models into practice for this study to clarify where different driver, vehicle, and environment factors “fit into” various stages of the crash-generating process. They also served as a framework to identify what data related to the conceptual model elements are necessary, where it is (or is not) possible to obtain those data, and at what level of quality are the data likely to have. Chapter 6 provides additional detail.

CHAPTER 4. INITIAL DATABASE ASSESSMENT

The project team conducted a review of national, State, and local databases with potential utility for this study. The following databases were included in the review:

- FARS.
- NASS GES.
- NASS CDS.
- NMVCCS.
- Current HSIS States (California, Illinois, Maine, Minnesota, North Carolina, Ohio, and Washington).
- CIREN.
- MCMIS.
- NPS's STARS.
- Crash and Roadway Data from State of Michigan.
- Crash and Roadway Data from State of Kansas.
- Crash and Roadway Data from Oakland County, Michigan.
- Crash and Roadway Data from State, Local, and Tribal Roads in Wyoming.

Michigan and Kansas were selected as two representative "non-HSIS" States with a significant amount of rural road mileage and an efficient data collection process (the research team has established contacts in previous studies that utilized these data).

The project team reviewed and evaluated each database against the following questions:

- Who "houses" and maintains the data?
- What is the spatial coverage of the data?
- What years of data are in the database?
- What is the general availability of the data? (i.e., online, formal request, not available)
- How are the data collected? How are the data coded?

- Does the database include all crashes for the coverage area (i.e., the population) or just a portion of the crashes (i.e., a sample)?

- How are crash severity levels defined?

- What is the vehicle type coverage of the data?

- If just a sample, how was the sampling done?

- If just a sample, what (if any) guidance is given to incorporate the sampling procedure into data analysis?

- Are variables included to identify horizontal curve and unsignalized intersection crashes on rural two-lane roads?

- Are crash coordinates or other location references to directly or indirectly geo-locate the crash available?

The findings of this review informed the research team's approach to the analysis.

DATABASE SELECTION

The project team developed criteria to select 8 to 10 databases from those identified in the previous section for further analysis. The main selection criteria included:

- Ability to identify horizontal curve and unsignalized intersection crashes on rural two-lane roads.

- Ability to geo-locate crashes (either directly or indirectly).

- Availability of variables for grouping crashes into the Volpe pre-crash scenarios.

- Extent of additional human, vehicle, and environmental variables.

- Coverage of local and/or tribal roads.

- Overall resource intensiveness of data collection.

Some databases need to be "in hand" to address all six of these criteria. Time or budget for the project did not allow for formal data requests, data delivery, and exploration of all actual databases. As noted, the review was based on readily available information. The project team applied judgment on the potential utility of each database combined with specific project needs (e.g., consider local and tribal data) in these cases to make a recommendation. As such, the team recommended the following databases for further review and analysis:

- GES.

- CDS.

- NMVCCS.
- HSIS—Illinois.
- HSIS—Minnesota.
- HSIS—North Carolina.
- HSIS—Washington.
- Oakland County, Michigan—Traffic Crash Analysis Tool.
- Wyoming DOT/Local Technical Assistance Program (LTAP) /Tribal Technical Assistance Program (TTAP).

GES served as the starting point for the analysis methodologies described in subsequent sections. The project team selected CDS and NMVCCS based on the extent of additional human, vehicle, and environmental variables. Over 600 elements are coded in CDS; NMVCCS includes more than 500, including a critical pre-crash event and a critical reason for the crash event. The team recommends HSIS data, due to the capability to specifically locate the crashes and link roadway inventory data, exposure data, and crash data for a large sample of primary route mileage. North Carolina Department of Transportation (NCDOT) owns and maintains all roads in the State; the North Carolina database includes local road coverage as a result. Washington data includes information on horizontal and vertical alignment. WSDOT is also completing an effort to collect detailed roadside inventory data. At one time, WSDOT was scheduled to complete the roadside inventory data collection by late 2015 or early 2016, but the program was subsequently stopped. Data are still available for road segments where the agency inventoried the roadside prior to the program ending. Additional information relevant to this project was also readily available through Washington DOT’s website (e.g., video logs). For this effort, the project team leveraged an ongoing FHWA project, a portion of which the University of Utah is conducting, to code horizontal alignment and cross-section data for rural roads in Minnesota and Illinois. In addition, Minnesota data also includes a specific intersection file. Oakland County, Michigan, and Wyoming DOT/LTAP/TTAP data needed additional exploration. The team recommended these for further study to increase the chances of meaningful findings on local and tribal roads. All databases remained under consideration until the project team identified priorities for the remainder of project. The team also made additional database assessments in the context of this project’s budget and schedule. The remaining sections of this chapter briefly summarize these assessments; they were key inputs to establishing the work plans for the remainder of the research.

GES Coding Changes

The project team used GES data to identify a set of high-priority pre-crash scenarios for crashes occurring on horizontal curves and at unsignalized intersections along rural two-lane roads. GES variable coding underwent significant changes beginning in 2010. Horizontal curve and intersection variables were discontinued. Analysis of GES data for this project therefore focused on years earlier than 2010.

Geo-Locating Crashes in CDS and NMVCCS Databases

The project team explored the ability to geo-locate crashes in the CDS and NMVCCS datasets. They determined that CDS did not have any location information available. This made it difficult to link CDS to other spatially-coded datasets; the team removed CDS from further consideration.

Location information was available for the NMVCCS and, at the time, no other studies had used it. FHWA worked directly with the National Highway Traffic Safety Administration (NHTSA) to request the NMVCCS location data officially. They adopted the following protocol:

- The project team will provide a list of crash case numbers to FHWA.
- FHWA will pass this information onto NHTSA.
- The location information for those crashes will be released.

As part of Task 5, the project team developed SAS codes to extract the high-priority Volpe pre-crash scenario types that are the focus of the detailed data analysis for this demonstration. Additional information on the GPS coordinate request and detailed NMVCCS analysis are described further in chapter 6 and chapter 7.

Obtaining Local and Tribal Road Data

The project team originally proposed moving forward with three datasets in order to cover local and tribal roads, a key aspect of the study: North Carolina, Wyoming DOT/LTAP/TTAP, and Oakland County, Michigan. NCDOT owns and maintains all roads in the State; the North Carolina database includes local road coverage as a result. It was relatively easy to access and analyze North Carolina data, given its HSIS status.

Wyoming DOT/LTAP/TTAP and Oakland County, Michigan, data needed additional exploration, but the project team recommended them for further study to increase the chances of meaningful findings on local and tribal roads. Early in the study, the project team learned that the Wyoming databases are far behind schedule, and location data on local and tribal roads would not be available in the near term. As a result, the team did not move forward with this database. Obtaining local data from Oakland County took longer than expected due to a requested $1,000 annual data access fee that the Traffic Improvement Association (TIA) requested. TIA agreed to waive the fee and provide access to raw data to the project team. The project team explored the Oakland County data, but variables beyond what the team could find in Michigan's State databases were limited. The team did not recommend the dataset for additional analysis.

The project team also explored the possibility of incorporating NPS data into future analysis. A sample (2001–2005) of crash and roadway data for Zion National Park was provided by NPS and FHWA. Crash data variables were inadequate for this study. The data files did not contain information on the number of lanes or the type of traffic control device at the location of a crash. This made it difficult to determine if a crash happened at an unsignalized intersection or on a two-lane road. The dataset had variables describing the types of collisions between vehicles or between a vehicle and a fixed object, but it did not have anything to describe the critical events prior to the collision. Reliably identifying the Volpe pre-crash scenarios did not appear possible because these key pieces of "pre-crash" information were missing. The team did not recommend the dataset for additional analysis.

Naturalistic Driving Data

The project team identified the SHRP2 NDS data as a potential alternative data source for studying the sequence of events leading to crashes or near crashes, particularly those events related to driver behaviors and reactions. NDS data collection efforts were still underway throughout most of this project. Since the Virginia Tech Transportation Institute (VTTI) operated and maintained the NDS data, there was discussion regarding the process for researchers outside of VTTI to access the data for their own research at the Eighth SHRP2 Safety Symposium, held in Washington, DC, on July 11, 2013. Accessing a sample of "raw" data and videos (e.g., trips along rural two-lane horizontal curves) required a subcontract and data sharing agreement with VTTI. Based on ongoing SHRP2 research, VTTI provided the project team an unofficial subcontract estimate of $60–100K to obtain a sample of raw NDS data for rural two-lane horizontal curves. VTTI also noted at the time that their current focus was on collecting the remainder of the NDS data, and that processing new data requests was a lower priority. They would likely not process any new data requests until April 2014 or later. Based on these findings, the NDS data was not a practical data alternative for this specific study. NDS data has since become more widely available, including to project teams conducting work as part of SHRP2's "Concept to Countermeasure – Research to Deployment Using the SHRP2 Safety Data."

Utah LiDAR Data

Utah undertook an extensive data collection effort to gather, identify, and process detailed information on all above-ground assets and road characteristics along State routes using LiDAR. The effort appeared to be the first of its kind executed by a State DOT. The database was expected to provide access to detailed road information that is traditionally not available for safety analysis, including grade, superelevation, roadside objects and offsets, roadside slopes, sign and pole presence, and pavement condition, among others. The FHWA technical advisory committee for this research effort expressed a high level of interest in incorporating the Utah LiDAR data into this project as a way to demonstrate how this type of detailed information can improve road safety analysis. The project team gained access to a sample of non-freeway data, but UDOT was still in the process of reducing data to quantify key road elements of interest (e.g., roadway and roadside features, horizontal curvature, cross slope/superelevation, grade, vertical curvature, shoulder slope) throughout much of this project. The project team continued to communicate with UDOT and kept flexibility in this project to incorporate the LiDAR data when it was fully available, which it ultimately was. The horizontal data did not arrive in a form that the project team could immediately use for analysis. This project, as well as a parallel effort conducted at Brigham Young University, explored ways to make full use of the horizontal curve data through additional post-processing and algorithms. Additional information is provided in chapters 6, 8, and 9.

CHAPTER 5. SELECTING RESEARCH PRIORITIES

As previously noted, the objective of this study was to increase understanding of causative, precipitating, and predisposing factors of crashes occurring on horizontal curves and at unsignalized intersections along rural two-lane roads, demonstrating the use of enhanced data from various traditional and non-traditional data sources. This chapter discusses additional prioritization of crash types and study designs for future research efforts to pursue in more detail.

HIGH-PRIORTY CRASH SCENARIOS

The project team identified "high-priority" crash scenarios for crashes occurring on horizontal curves and at unsignalized intersections along rural two-lane roads. The team selected these crash scenarios to ensure that project resources would be spent on crash types considered most problematic at the rural two-lane locations of interest. The team used the Volpe pre-crash scenarios to group and prioritize crashes for analysis. Volpe developed a classification structure that utilized GES data to define 38 single-vehicle pre-crash scenarios and 46 multi-vehicle pre-crash scenarios.[(32)] The scenarios grouped crashes that were similar based on a set of pre-crash events.

The project team considered the Volpe groupings superior to more traditional methods of grouping crashes by characteristics, such as number of involved vehicles and crash location. For example, "single-vehicle, run-off-road" is a commonly used crash category for a Highway-Safety-Manual style of analysis comprised of SPFs and CMFs. However, a driver/vehicle may run off the road for multiple (and very different) reasons, such as avoiding an oncoming vehicle that had encroached into their lane, avoiding an animal, overcorrecting after driving on a pavement edge drop-off or rumble strip, or simply losing control. The project team considered these types of detailed pre-crash explanations a more effective starting point for identifying causative, precipitating, and predisposing crash factors when compared to the more traditional crash types. The team also retained the Volpe terminology for each pre-crash scenario at this stage of the project for consistency with other related efforts that utilize these scenarios. In the case of run-off-road crashes, the project team retained Volpe's term, "road edge departure."

Methodology to Select Scenarios and Results

The project team identified the "high-priority" crash types, grouped by the Volpe pre-crash scenarios, using 2005–2008 GES data. As noted in the previous chapter, the team selected these years to avoid significant changes made to GES variable coding in more recent years, allowing more direct application of previous coding work by Dr. Wassim Najm at Volpe. Dr. Najm provided SAS programs that identify pre-crash scenarios from the GES. The SAS codes represented a "third generation" crash typology that combines information from the "GM 44-crashes typology" and the United States Department of Transportation (USDOT) "pre-crash scenarios typology" in support of the Intelligent Vehicle Initiative (IVI). The pre-crash scenarios were then ranked using economic crash costs, and prioritized lists were developed for each of the following rural two-lane subgroups: single-vehicle, horizontal curve; multi-vehicle, horizontal curve; single-vehicle, unsignalized intersection; and multi-vehicle, unsignalized intersection. The project team estimated the economic crash costs as part of the Volpe SAS codes and accounted for goods and services that must be purchased or productivity that is lost as a result of the crashes. They do not consider the intangible pain, suffering, and loss of life consequences.[31] The top three crash types for each of the four subgroups, ranked according to annualized crash costs over the 4-year (2005–2008) analysis period, are shown in table 3 through table 6.

Table 3. High-priority, single-vehicle pre-crash scenarios on rural two-lane horizontal curves identified using overall crash cost.

Single-Vehicle Pre-Crash Scenario	Curve Crash Cost (Millions of Dollars)
4. Control loss/no vehicle action	16,524
8. Road edge departure/no maneuver	8,323
38. Object contacted/no maneuver	515

Table 4. High-priority, multi-vehicle pre-crash scenarios on rural two-lane horizontal curves identified using overall crash cost.

Multi-Vehicle Pre-Crash Scenario	Curve Crash Cost (Millions of Dollars)
22. Opposite direction/no maneuver	4,402
4. Control loss/no vehicle action	1,006
31. SCP @ non-signal	653

Table 5. High-priority, single-vehicle pre-crash scenarios on rural two-lane unsignalized intersections identified using overall crash cost.

Single-Vehicle Pre-Crash Scenario	Intersection Crash Cost (Millions of Dollars)
3. Control loss/vehicle action	857
8. Road edge departure/no maneuver	709
13. Pedestrian/no maneuver	701

Table 6. High-priority, multi-vehicle pre-crash scenarios on rural two-lane unsignalized intersections identified using overall crash cost.

Multi-Vehicle Pre-Crash Scenario	Intersection Crash Cost (Millions of Dollars)
31. SCP @ non-signal	6,942
30. Left turn across path/opposite direction at non-signal	1,073
27. Rear-end/lead vehicle stopped (LVS)	796

Prioritized lists based on frequency instead of economic crash costs were very similar to those presented in table 3 through table 6, with the following differences:

- "38. Object contacted/no maneuver" is replaced by "11. Animal/no maneuver" in table 3.
- The order of "4. Control loss/no vehicle action" and "31. SCP @ non-signal" is reversed in table 4, but their frequencies are approximately equal.

- “13. Pedestrian/no maneuver” is replaced by “7. Road edge departure/maneuver” in table 5.

- The order of “30. Left turn across path/opposite direction at non-signal” and “27. Rear-end/LVS” is reversed in table 6.

A more detailed summary of the pre-crash scenario ranking results is provided in the appendix.

The project team made the following additional observations in an effort aimed at further refining the scope of future analysis activities for this project:

- Horizontal curves:
  - The pre-crash scenario SAS codes were very similar for the single-vehicle “4. Control loss/no vehicle action” and the single-vehicle “8. Road edge departure/no maneuver” scenarios; both accounted for 98 percent of the crash costs from the top three single-vehicle scenarios.
  - The multi-vehicle “22. Opposite direction/no maneuver scenario” accounted for more than four times the crash cost as the “4. Control loss/no vehicle action scenario” and more than six times the crash cost as the “31. SCP @ non-signal scenario.”
  - The multi-vehicle “31. SCP @ non-signal scenario” captured crashes at unsignalized intersections that also happened to be along horizontal curves.

- Unsignalized intersections:
  - The multi-vehicle “31. SCP @ non-signal scenario” accounted for between 6 and 10 times the total crash cost as any other unsignalized intersection scenario, whether single- or multiple-vehicle.

Based on these findings, the project team recommended the following three pre-crash scenarios as the focus of the project:

- Combination “control loss/no vehicle action” and “road edge departure/no maneuver” pre-crash scenarios for single-vehicle crashes on horizontal curves.

- “Opposite direction/no maneuver” pre-crash scenario for multi-vehicle crashes on horizontal curves.

- “SCP 2 non-signal” pre-crash scenario for multi-vehicle crashes at unsignalized intersections (on both tangent sections and on horizontal curves).

These three recommended scenarios account for approximately 51 percent of the total crash frequency and 72 percent of the total crash cost for all crashes occurring on horizontal curves and at unsignalized intersections along rural two-lane roads. The remaining seven scenarios that this report does not address are possible research alternatives for the future analysis. These include the following pre-crash scenarios:

- Horizontal curves:
    - Single-vehicle “38. Object contacted/no maneuver.”
    - Multi-vehicle “4. Control loss/no vehicle action.”

- Unsignalized intersections:
    - Single-vehicle “3. Control loss/vehicle action.”
    - Single-vehicle “8. Road edge departure/no maneuver.”
    - Single-vehicle “13. Pedestrian/no maneuver.”
    - Multi-vehicle “30. Left turn across path/opposite direction at non-signal.”
    - Multi-vehicle “27. Rear-end/LVS.”

PROOF OF CONCEPT: IDENTIFYING HIGH-PRIORITY SCENARIOS IN DATABASES OTHER THAN GES

The SAS codes used to identify the high-priority, pre-crash scenarios were developed specifically for GES data. However, this research project utilized crash data from a number of databases, including State and local data, as well as databases that are created using on-scene crash investigations, such as NHTSA’s NMVCCS data. This study identified crashes that match the pre-crash scenarios of interest in these other, non-GES databases. Each database is different in terms of variable names, coding, and availability of pre-crash events. Whether or not the team could reliably identify the Volpe pre-crash scenarios in other, non-GES databases remained a lingering question. This section describes how the team achieved this.

Najm et al. includes a detailed list of codes for identifying single-vehicle and multi-vehicle pre-crash scenarios.[(32)] These codes served as the starting point for identifying the same pre-crash scenarios in other, non-GES databases. The next step was to identify similar variables and codes in the other databases. The project team documented all the possible combinations of the identified variables that contribute to a crash. This became the first draft of what the team called the “pre-crash scenario functional specifications.”

The team then developed SAS codes to implement the first draft of the functional specifications. Findings showed that the net cast by the first draft was too wide; the functional specifications were capturing crashes that did not necessarily belong to a pre-crash scenario of interest. The project team refined the functional specifications, eliminating redundant variables as well as variables that did not capture the intention of the pre-crash scenario. The SAS code underwent revision to match the refined specifications.

Finalizing the functional specifications in SAS was the final step of the proof of concept for identifying high-priority, pre-crash scenarios in other non-GES datasets. The SAS codes for each dataset consisted of three key parts: importing the data, processing and merging the data, and identifying and assigning crash scenarios.

Importing Data into SAS

The raw data come in various structures and formats. Some datasets have all necessary variables in two or three separate files (e.g., HSIS data with accident, vehicle, and person files). Some other datasets may come in a much larger number of data files (e.g., GES or NMVCCS data with many different data files) and the information necessary to identify crash scenarios is stored across those files. The datasets do not only vary in their structures but they also come in different formats. Some are SAS data format; others might be Microsoft® Excel spreadsheets (XLS and XLSX) or comma separated values (CSV). This first key step of SAS code implementation was to import the data from various files and in various formats into SAS, store them in similarly managed data library structures, and prepare them for the next steps of the process.

Processing and Merging Data

After the project team imported them into SAS libraries, the data were in SAS format. However, the relevant information was still spread across different files. For example, the crash event sequence information may be stored in a vehicle data file for a particular dataset, but the equivalent variables might be in a separate event data file for another dataset. The primary goal of this step was to process the data, transform the file typology into a similar structure, and then merge the data into one single data file where every piece of information available to identify crash scenarios was in the same file. All datasets had an accident file in which each row represented one crash. The other data files (e.g., vehicle or person/occupant files) all had one row of data for each vehicle or person. That means they had multiple rows of data associated with each crash if such crash had more than one vehicle or person involved. The project team needed to transform these multiple rows of data associated with each crash into a single row to merge them with each row in the accident file. The team used SAS data arrays to convert rows into columns of data. After transforming all data files from multiple rows for each crash into one single row with multiple columns, the team merged the data with accident files using case numbers as the matching identification.

Identifying and Assigning Crash Scenarios

In this step, the project team coded the functional specification for each crash scenario into SAS. The SAS codes run through the dataset and test a series of conditions for each crash. If all conditions for a particular crash scenario are met, a crash scenario code is assigned to that crash. The team considered a total of five different scenarios in this step:

- Opposite direction with no maneuver (multi-vehicle crash) on horizontal curve.
- Straight crossing path (multi-vehicle crash) at stop-controlled intersection.
- Left-turn cross path (multi-vehicle crash) at stop-controlled intersection.
- Control loss/road-edge departure (single-vehicle crash) on horizontal curve.
- Pedestrian with no vehicle maneuver at stop-controlled intersection.

The first, second, and fourth scenarios correspond to the high-priority scenarios recommended for this study. The project team developed the SAS code for third and fifth scenario as a starting point for possible future analysis.

Example Results

The project team implemented the above process for three different datasets: 1 year of North Carolina (HSIS database), 1 year of Utah, and complete NMVCCS data.

A summary of the results is presented in table 7, table 8, and table 9. The NMVCCS numbers in table 9 represent the raw numbers of crashes in the dataset, not a weighted count.

Table 7. Number of crashes identified for pre-crash scenarios coded in SAS—North Carolina (2010 data).

Pre-Crash Scenario	SV/MV	Location	Number of Crashes
Opposite direction/no maneuver	Multi-vehicle	Horizontal curve	953
SCP	Multi-vehicle	Stop-controlled intersection	1,000
Left turn cross path/no maneuver	Multi-vehicle	Stop-controlled intersection	75
Control loss/road edge departure/no maneuver	Single-vehicle	Horizontal curve	701
Pedestrian/no maneuver	Single-vehicle	Stop-controlled intersection	2

Table 8. Number of crashes identified for pre-crash scenarios coded in SAS—Utah (2008 data).

Pre-Crash Scenario	SV/MV	Location	Number of Crashes
Opposite direction/no maneuver	Multi-vehicle	Horizontal curve	63
SCP	Multi-vehicle	Stop-controlled intersection	6
Left turn cross path/no maneuver	Multi-vehicle	Stop-controlled intersection	0
Control loss/road edge departure/no maneuver	Single-vehicle	Horizontal curve	499
Pedestrian/no maneuver	Single-vehicle	Stop-controlled intersection	0

Table 9. Number of crashes identified for pre-crash scenarios coded in SAS—NMVCCS.

Pre-Crash Scenario	SV/MV	Location	Number of Crashes
Opposite direction/no maneuver	Multi-vehicle	Horizontal curve	71
SCP	Multi-vehicle	Stop-controlled intersection	339
Left turn cross path/no maneuver	Multi-vehicle	Stop-controlled intersection	4
Control loss/road edge departure/no maneuver	Single-vehicle	Horizontal curve	335
Pedestrian/no maneuver	Single-vehicle	Stop-controlled intersection	0

RECOMMENDED STUDY DESIGNS TO DEMONSTRATE DATA CONCEPTS

The project team initially defined the specific focus for this project as crashes occurring on horizontal curves and at unsignalized intersections along rural two-lane roads. The previous sections of this chapter identified high-priority crash scenarios for crashes occurring on horizontal curves and at unsignalized intersections along rural two-lane roads using GES. The project team selected the following three pre-crash scenarios as the focus of more detailed data mining and analysis:

- Combination control loss/no vehicle action and road edge departure/no maneuver pre-crash scenarios for single-vehicle crashes on horizontal curves.
- Opposite direction/no maneuver pre-crash scenario for multi-vehicle crashes on horizontal curves.
- SCP @ non-signal pre-crash scenario for multi-vehicle crashes at unsignalized intersections (on both tangent sections and on horizontal curves).

Chapter 6 will demonstrate conceptual crash models developed for the three selected high-priority, pre-crash scenarios, "filling in" the generalized conceptual crash model framework presented in chapter 3 of this report with more specific data elements. The team then used the models to identify data needs related to each crash type and to assess related data availability, accuracy, and reliability in both traditional and non-traditional data sources. Chapter 6 also describes this process. Finally, the project team proposed the following two sets of studies based on the project's schedule and remaining budget:

- The Benefits and Trade-Offs of Enhanced Data and Refined Crash Type Definitions on Models Relating Expected Crash Frequency to Predisposing Roadway, Traffic, and Weather Factors.
- Developing Crash Causal Types Using Data from Detailed, On-Scene Crash Investigations.

Methodologies and results associated with these studies are fully documented in chapters 7, 8, and 9.

CHAPTER 6. MARRIAGE OF CONCEPTUAL CRASH MODELS WITH SAFETY DATA AVAILABILITY AND QUALITY ASSESSMENTS

This chapter is oriented toward identifying the data collection needs related to the three critical crash scenarios defined in the previous sections. The conceptual models for the critical crash scenarios identify potential predisposing factors and precipitating events. The next section builds on the conceptual crash models by beginning to “fill in” a framework that captures where data related to the conceptual model elements can (or cannot) be obtained. The framework also addresses the quality of data for the conceptual model elements. This approach results in a more general approach to determining data needs, availability, and quality because it utilizes the “universe” of applicable data elements (as defined in the conceptual crash models) as opposed to being constrained or biased by known limitations in existing databases.

This section addresses predisposing factors, constraints, and target groups separately from precipitating events. The reason for this is because the nature of the data is different. Most of the possible target groups and predisposing factors represent at least partially available information. However, precipitating events represent cognitive abilities, decisionmaking under pressure and uncertainty, motoric reactions of drivers, and vehicular responses to the driver’s actions. While the nature of a vehicle’s responsiveness might be determined from the post-crash vehicle check, human behavior-related data present more of a challenge.

The first part of this section discusses predisposing factors and target groups needs identified from conceptual models. For the exploratory data analysis, the project team categorized data needs in five groups: vehicle, driver, environment, roadway, and traffic. The team identified data needs through the analysis of databases used to develop functional specifications for the three high-priority, pre-crash scenarios. Other alternatives for data collection are also discussed. The second part of this section is a more general discussion of potential opportunities for data collection related to precipitating events. Data needs related to precipitating events fall into two groups: vehicle and operator data. Data reliability is also covered. The final section describes the actual data collection and sources of data for the critical crash scenarios.

CONCEPTUAL CRASH MODELS FOR HIGH-PRIORITY, PRE-CRASH SCENARIOS

This first part of this chapter presents conceptual crash models for the three selected high-priority, pre-crash scenarios on horizontal curves and unsignalized intersections. The goal of the conceptual crash models is to try to identify all possible elements that interact to either lead to or prevent a crash from occurring. The project team developed a conceptual model for each of the three selected scenarios. The conceptual models helped identify the categories of data that need to be collected for each crash scenario during the Task 5 data mining and analysis activities, and provided insights to the possible combinations of events that could lead to a crash outcome.

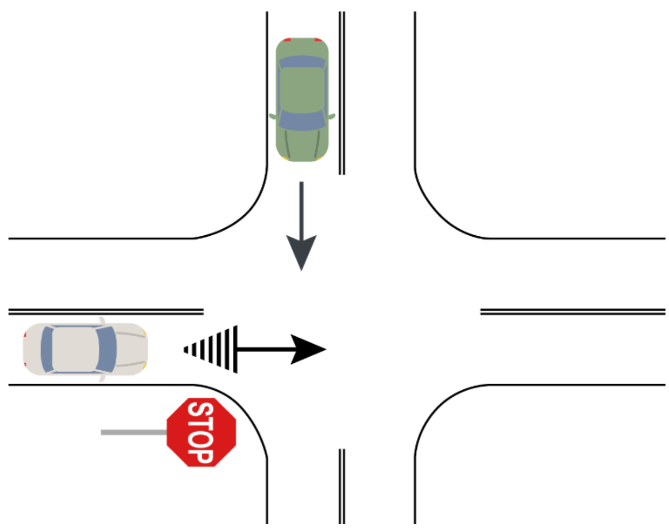
This section presents the models in a flexible, tabular way based on what the team learned during the Task 5 research, data mining and analysis. In addition to predisposing factors and precipitating events, the conceptual models included what have been identified as constraints, countermeasures that are (or are not) implemented somewhere in the driver–vehicle–road system to influence behavior of different elements involved in a crash scenario. The project team also developed hypothetical timelines for each scenario. They demonstrated allocation of most important time points and processes in the pre-crash event sequence, accounting for the time-sequence and time-sensitive nature of the involved elements and behaviors.

This study first identified guidelines used to develop conceptual crash models for this research. It then presented conceptual models for the three high-priority, pre-crash scenarios, including crash descriptions, elements present during a crash, main questions that need to be answered to determine the crash causality, and hypothetical event timelines. This led to defining the data needs and providing the framework for establishing causal relationships for the selected crashes.

CONCEPTUAL CRASH MODEL FOR SCP AT UNSIGNALIZED INTERSECTIONS

Figure 8 provides a graphical representation of an SCP crash at an unsignalized intersection. This is a multi-vehicle crash at an unsignalized intersection with stop control on the minor street. The crash scenario describes two vehicles, one with the right of way and one without, traveling through the intersection and crossing paths at nearly 90 degrees (depending on the intersection approach alignment). A conceptual crash model framework for this scenario is provided in table 10, table 11, and figure 9. The following elements are included in this crash conceptual model:

- Vehicle with the right of way (Vehicle 1).
- Driver of Vehicle 1 (Driver 1).
- Vehicle without the right of way on the stop-controlled approach (Vehicle 2).
- Driver of Vehicle 2 (Driver 2).
- Roadway.
- Traffic.
- Environment.



Source: USDOT.

Figure 8. Graphic. Critical crash scenario 31.[32]

The roadway and traffic elements provided in table 10 capture characteristics of each intersection approach as well as the intersection as a whole. Known variables for the conceptual model include location (unsignalized intersection), facility type (rural two-lane highway), and traffic control (stop sign on the minor approaches). It is also possible to assume potential maneuvers of drivers on both approaches (e.g., going straight, slowing, stopping, passing, changing lanes, merging, turning). Other details about the crash are unknown and should be established from the crash data about each of the elements mentioned above. In order to answer what other data to collect to describe this crash, the conceptual framework should cover the information that could later answer the following questions:

- Why is Driver 1/Vehicle 1 crossing paths with Driver 2/Vehicle 2 at the same point in space and time?
    - The Driver 1/Vehicle1 is on the approach with the right of way but still should proceed through the intersection aware of potential conflicting vehicles. Predisposing driver factors include experience, mental and physical condition, distraction, and risk-taking characteristics. Even if the driver acts in a timely manner to avoid a potential collision, the vehicle's steering and braking system should be able to respond appropriately.
- Why is Driver 2/Vehicle 2 crossing paths with Driver1/Vehicle 1 at the same point in space and time?
    - The Driver 2/Vehicle 2 is on the minor approach with the stop sign. This driver needs to be aware of both intersection and signage, come to a stop, detect the presence of traffic on the major street, and identify appropriate gaps to cross the intersection. The same driver and vehicle related factors related to Driver 1/Vehicle 1 are also applicable to Driver 2/Vehicle 2. Vehicle 2's acceleration capabilities are also important, as Driver 2 needs to accelerate through the intersection.
- Do any roadway features contribute to the crash?
    - Probably the most important roadway features that could prevent/contribute to this type of crash are related to sight distance (e.g., stopping, intersection, decision), roadway alignment, and the types and conditions of the traffic control devices.
- Do traffic conditions contribute to the crash?
    - Higher traffic volumes increase exposure and driver workload, but can also make a driver more aware of conflicting vehicles. As gap availability on the major road decreases, drivers on the minor street may accept smaller gaps. This may be particularly true if a queue forms behind the minor approach driver.
- Does the environment contribute to the crash?
    - Anything from the environment that could affect the driver's ability to detect roadway and traffic conditions and respond in time are worth considering in this crash scenario. Potential contributing factors are light conditions, weather conditions, and overall visibility.
- What combination of elements and their features contributed to the crash/caused two vehicles to reach the crash point at the same time?
    - Both vehicles involved in this crash could either strike or be struck by the other vehicle, and vehicles crossing paths could result from both drivers' behavioral characteristics. Constraints imposed by the environment, roadway, and traffic,

especially related to signage location and maintenance, could have their own impact on the crash occurrence, and this is why the composite impact of different elements is considered.

These and other related questions are captured in the part of the conceptual crash model covering search, detection, evaluation, decision, and actions, defined in table 11.

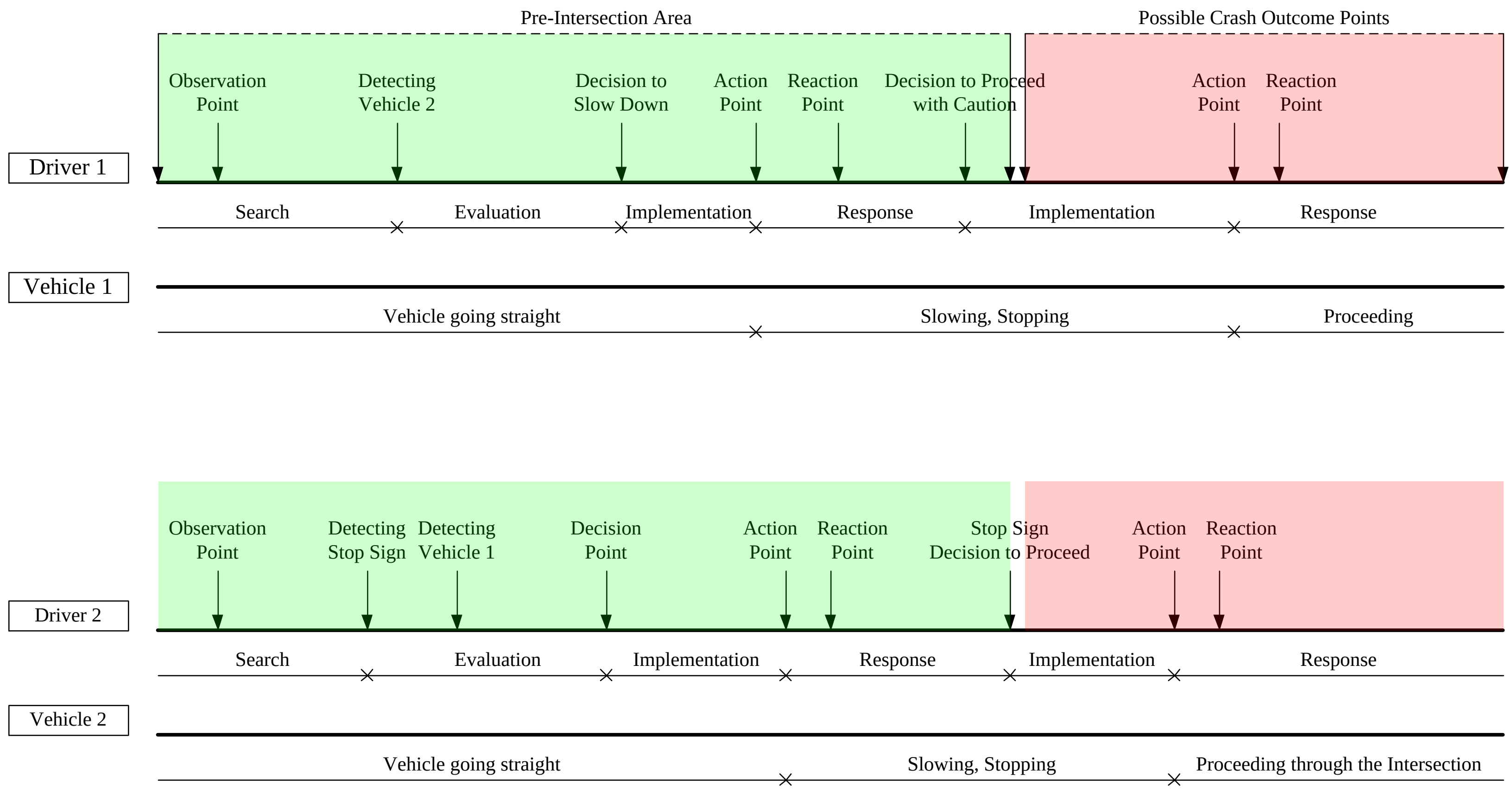
The timeline developed for this crash scenario, shown in figure 9, separates pre-intersection from the intersection area. The crash occurs in the central area of the intersection where vehicles from lateral directions cross paths (i.e., a crossing conflict point). The number of detection, decision, action, and reaction points is also higher than in the previous scenarios, because these points may relate to multiple decisions and movements, including deceleration, stopping, turning, and/or accelerating. The crash points may occur at any time after Driver 2/Vehicle 2 proceeds from the stop sign on the minor approach and before Driver 2 clears the intersection.

Table 10. Conceptual Crash Models: Predisposing Factors and Controls.

Inputs: System Elements	Resources: Predisposing Factors	Controls: Conditions
Driver 1, 2	License, age, gender, training and experience, income, sensory and motor capabilities (vision, hearing, reaction time), familiarity with vehicle, familiarity with site, physical condition (fatigue, illness, injury), mental condition (inattention, bad judgement, influence of drugs or alcohol), distraction, speed choice, risk taking characteristics (failure to wear seat belt)	Education, law, enforcement, communications and outreach, penalties and sanctions, treatment/monitoring/control, and other programs and actions
Vehicle 1, 2	Vehicle type and condition, year and model, seat belts, child restraints, airbags, braking system, steering, tires, bumper heights, energy absorption, windshields, headlights and taillights, mirror presence and condition, side and rear windows condition, ease of removal or injured passengers	Vehicle maintenance and vehicle safety inspection
Roadway	Type and classification, surface type and condition, pavement friction, design speed, horizontal alignment on approaches, superelevation, intersection skew, cross-section design, auxiliary lanes (presence, type, terminals, transitions), islands presence and type, turning radii, roadside design (slopes, ditches, objects, hazard rating), available sight distance, lighting presence and condition, expectancy violations (unconventional design, work zones)	Required design criteria, pavement markings (presence, type, condition), delineators (presence, type, condition), rumble strip presence and condition, roadside barrier (presence, type), crash cushions (presence, type, conditions), traffic calming treatments, expectancy violations, pavement maintenance (availability, response time), and other safety countermeasures
Traffic	Approach volumes, turning movements, capacity and Level of Service, operating and running speed, directional split, queue presence, peak hour factor, traffic composition (percent of trucks and buses), pedestrian presence, bicyclist presence, transit presence, parking presence	Traffic control devices (presence, location, visibility), posted/advisory speed, speed detectors (presence, location), pedestrian accommodation, transit accommodation, ITS information, and amount of information displayed
Environment	Daytime, weather conditions, visibility, noise level, incident conditions, animals, and other hazards	Pre-trip information (available or not), emergency response quality (availability, response time)

Table 11. Conceptual Crash Models: Driver Processes and Events Corresponding to Crash Scenario 31.

Process	Driver/Vehicle 1	Driver/Vehicle 2
Search: scanning; perpetual process	Driver actively searches for potential threats on the road, being aware of the intersection and the right of way	Driver actively searches for potential threats on the road, being aware of the intersection and the right of way
Detection: actual perception; awareness	Driver detects the traffic coming from the lateral direction	Driver detects the stop sign and traffic coming from the lateral direction
Evaluation: recognition of threat; evaluate the need to avoid that threat	Driver evaluates the potential threat from crossing paths with the vehicles coming from another approach and assessed their speed, acceleration, and distance	Driver evaluates the possibility of striking/being struck by the vehicles coming from the lateral direction, and the need to stop and proceed with caution
Decision: determination of the action	Driver determines the action: no response, slowing, stopping, warning another driver	Driver determines the action: no response, slowing, stopping, and proceeding through the intersection
Human action: implementing decision	Driver implements previously determined action: in time, with a delay, has no time to implement decision	Driver implements the previously determined action: in time, with a delay, has no time to implement decision
Vehicle action: vehicle responding to driver's action	Vehicle responds to driver's action with possible maneuvers: going straight, slowing, stopping, passing, parking, leaving parked position, backing, changing lanes, merging	Vehicle responds to driver's action with possible maneuvers: going straight, slowing, stopping, passing, parking, leaving parked position, backing, changing lanes, merging



Source: FHWA.

Figure 9. Graphic. Example anchor points corresponding to crash scenario 31 for assessing time available and time used by drivers to perform functions.

Conceptual Crash Model for Single-Vehicle Crashes on Horizontal Curves

Figure 10 presents a graphical representation of control loss/no vehicle actions, which is combined with road edge departure/ no maneuver. This is a single-vehicle crash on a horizontal curve. The pre-crash scenario describes a vehicle traversing a horizontal curve and the driver losing control of the vehicle without an obvious reason (i.e., evasive maneuver, overcorrection). This scenario appears to be overrepresented under adverse weather conditions, where the driver loses control due to wet and slippery conditions and runs off the road. The predisposing factors and controls that are part of a conceptual crash model framework for this scenario remain the same as those provided in table 10. The model considers the characteristics of the following elements:

- Driver.
- Vehicle.
- Roadway.
- Traffic.
- Environment.

Source: USDOT.

Figure 10. Graphic. Critical crash scenario 4.[(32)]

Known variables in this scenario are roadway type (rural two-lane) and roadway segment type (horizontal curve). Some of the potential crash contributing factors are environmental characteristics (e.g., weather and light conditions), roadway conditions (e.g., wet or slippery road), and speed. Driver's actions include no steering, improper steering, proper steering, or corrective actions combined with different types of braking and acceleration. The crash type in the case of a crash is single-vehicle, run-off-road (right or left). The developed conceptual framework includes factors that should answer the following questions after the data collection:

- Did the driver's speed choice contribute to the control loss of the vehicle?
    - Driver's speed choice depends on driver characteristics, vehicle capabilities, traffic volumes, and perceptual cues from the roadway and environment. In order to answer this question, driver's experience, physical and mental condition, and risk-taking characteristics could be the most relevant predisposing factors. Constraints like legislation, education, or police surveillance could either decrease or enhance the driver's predisposition to select a certain speed.
- What vehicle characteristics could contribute to this crash?
    - If the vehicle did not pass a regular safety check or if it is poorly maintained and the steering or braking system does not respond properly when the driver takes

action to adjust the speed or path, then vehicle factors could contribute to this crash type.

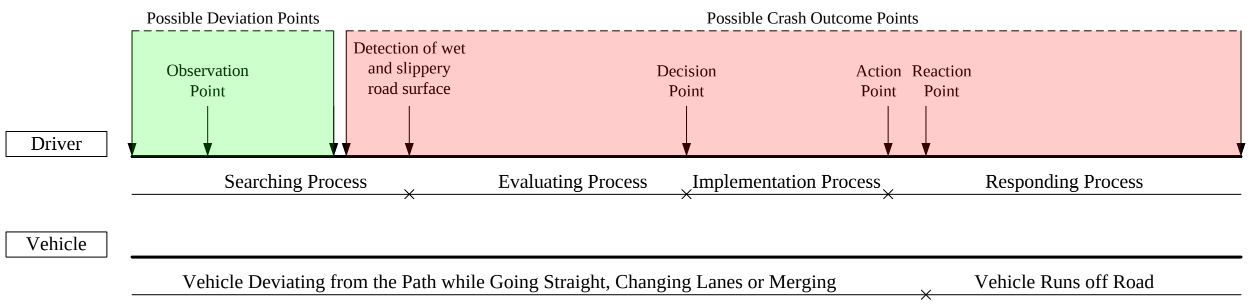
- What roadway features contribute to the crash?
    - Here the question revolves around whether the control loss occurs due to wet and slippery roadway conditions caused by the weather, or perhaps some other related features like friction supply, which is sensitive to roadway maintenance practices. In terms of constraints, the presence or absence of features such as edge rumble strips, edge drop-off treatments, and roadside barrier that would prevent the vehicle from running off the road also requires examination.
- Do traffic conditions contribute to the crash?
    - It is more likely that these crashes will occur in a lower traffic volume environment because as density increases the vehicle that loses control is more likely to be involved in a multi-vehicle crash. Traffic characteristics may also contribute to the propensity of this crash type through the presence or lack of car-following phenomena in terms of speed selection, and vehicle paths.
- Does the environment contribute to a crash?
    - General descriptions of this crash usually refer to adverse weather conditions and light conditions, but other potential causes of control loss coming from the environment should not be neglected (e.g., sun glare, wind).
- What combination of elements and their features contributed to the crash?
    - The idea behind these questions is that each element could contribute to a crash with both its predisposing factors and relevant constraints. However, it is important to examine the possibility of several elements together interacting and contributing to the crash.

These and other related questions are captured in the part of the conceptual crash model covering search, detection, evaluation, decision, and actions, defined in table 12.

The timeline for this crash scenario (figure 11) is the simplest among the scenarios described here because it only includes one driver's behavior and the vehicle's movements. Time periods are hypothetical. A vehicle might start to deviate from its intended path before or after it is feasible for the driver to observe the potential threat. The placement of the detection point comes after the point of feasible threat observation, and the distance between these two points depends on driver, roadway, and environmental predisposing factors. A crash could possibly take place even before the driver detects the threat, especially if the time between observation and detection point is too short. If the crash does occur, it stops the driver's behavioral processes, which means that the entire process from detection to vehicle's reaction point does not have to be completed.

Table 12. Conceptual Crash Models: Driver processes and events corresponding to crash scenario 4.

Process	Driver/Vehicle
Search—scanning, perpetual process	Driver actively searches for potential threats on the road, being aware of curvature, weather conditions, traffic
Detection—actual perception, awareness	Driver detects slippery or wed road or/and excessive speed as a threat in a timely manner
Evaluation—recognition of threat, evaluate the need to avoid that threat	Driver evaluates the situation on the road, considering the potential threat, traffic and weather conditions, and the need for potential speed and movement adjustments
Decision—determination of the action	Driver determines what action, considering system conditions, will enable crash avoidance
Human action—implementing decision	Driver performs necessary maneuvers to move the vehicle away from the potential threat
Vehicle action—vehicle responding to driver's action	Vehicle successfully responds to driver's actions, avoiding the crash



Source: FHWA.

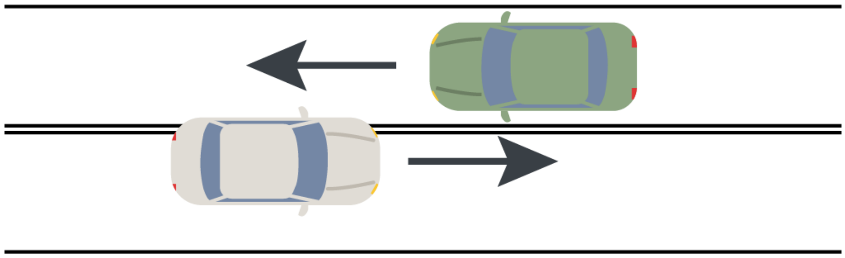
Figure 11. Graphic. Example anchor points corresponding to crash scenario 4 for assessing time available and time used by driver to perform functions.

Conceptual Crash Model for Multi-Vehicle Crashes on Horizontal Curves

Figure 12 presents a graphical representation of a multi-vehicle crash scenario involving two vehicles on a horizontal curve. The crash scenario describes a vehicle traversing a horizontal curve on a rural two-lane road, drifting left, encroaching into the opposing lane of traffic, and striking another vehicle traveling in the opposite direction. The predisposing factors and controls that are part of a conceptual crash model framework for this scenario remain the same as those provided in table 10. The model considers the characteristics of the following elements:

- Driver of the vehicle encroaching (Driver 1).
- Vehicle encroaching (Vehicle 1).
- Driver of the vehicle coming from the opposite direction (Driver 2).
- Vehicle coming from the opposite direction (Vehicle 2).
- Roadway.

- Traffic.
- Environment.



Source: USDOT.

Figure 12. Graphic. Critical crash scenario 22.[(32)]

Known variables in this scenario are roadway type (rural two-lane) and roadway segment type (horizontal curve). It is also possible to identify the potential maneuvers of drivers in both vehicles. The driver of the vehicle that is encroaching into the opposite direction travel lane would either continue on a path where at least part of the vehicle is in the opposing lane, or recover and return to the correct lane. The other driver, coming from the opposite direction, would likely perform some type of speed adjustment and avoidance maneuvers, depending on the perceived path of the encroaching vehicle. The type of crash outcome, including whether the crash is a head-on, angle, or sideswipe opposite direction crash, will also become apparent. Other details about the crash will be unknown, and researchers should establish these from the crash data about each of the elements mentioned above. The conceptual framework should cover the information that could answer the following questions:

- Why is Driver 1/Vehicle 1 encroaching into the opposite direction travel lane of the road?
    - This could be due to many of the same driver, vehicle, roadway, traffic, and environmental factors identified in pre-crash scenario 4, control loss/no vehicle action, as well as due to the absence or ineffectiveness of related system constraints (e.g., visible centerlines, centerline rumble strips).
- Why is Driver 2/Vehicle 2 unable to avoid Driver 1/Vehicle 1?
    - The answers to this question will revolve around the ability of Driver 2 to detect the encroaching vehicle as well as the avoidance maneuver options and available time to execute the avoidance maneuvers. Roadway design features, including lane width, shoulder width, barrier presence, roadside slope characteristics, and others will have an influence on avoidance maneuver options. This is why State agencies sometimes widen the shoulders in the area of a sight distance design exception.
- Do any roadway features contribute to the crash?
    - Apart from the fact that both vehicles are traveling on a horizontal curve, other roadway features can contribute to the crash together with constraints. For example, tighter curve radii, narrower cross-section dimensions, worn pavement markings, and the absence of centerline rumble strips may contribute to the initial encroachment of Driver 1/Vehicle 1 into the travel lane of Driver 2/Vehicle 2. Perhaps it is worth mentioning that once the encroachment occurs, there is not

much that roadway features can do to prevent the pre-crash event sequence, but they may still contribute to avoiding or leading to a crash (as outlined in the previous point). Implementing rumble strips could make the driver more aware, or if there is a roadside barrier it could prevent the vehicles from running off the road.

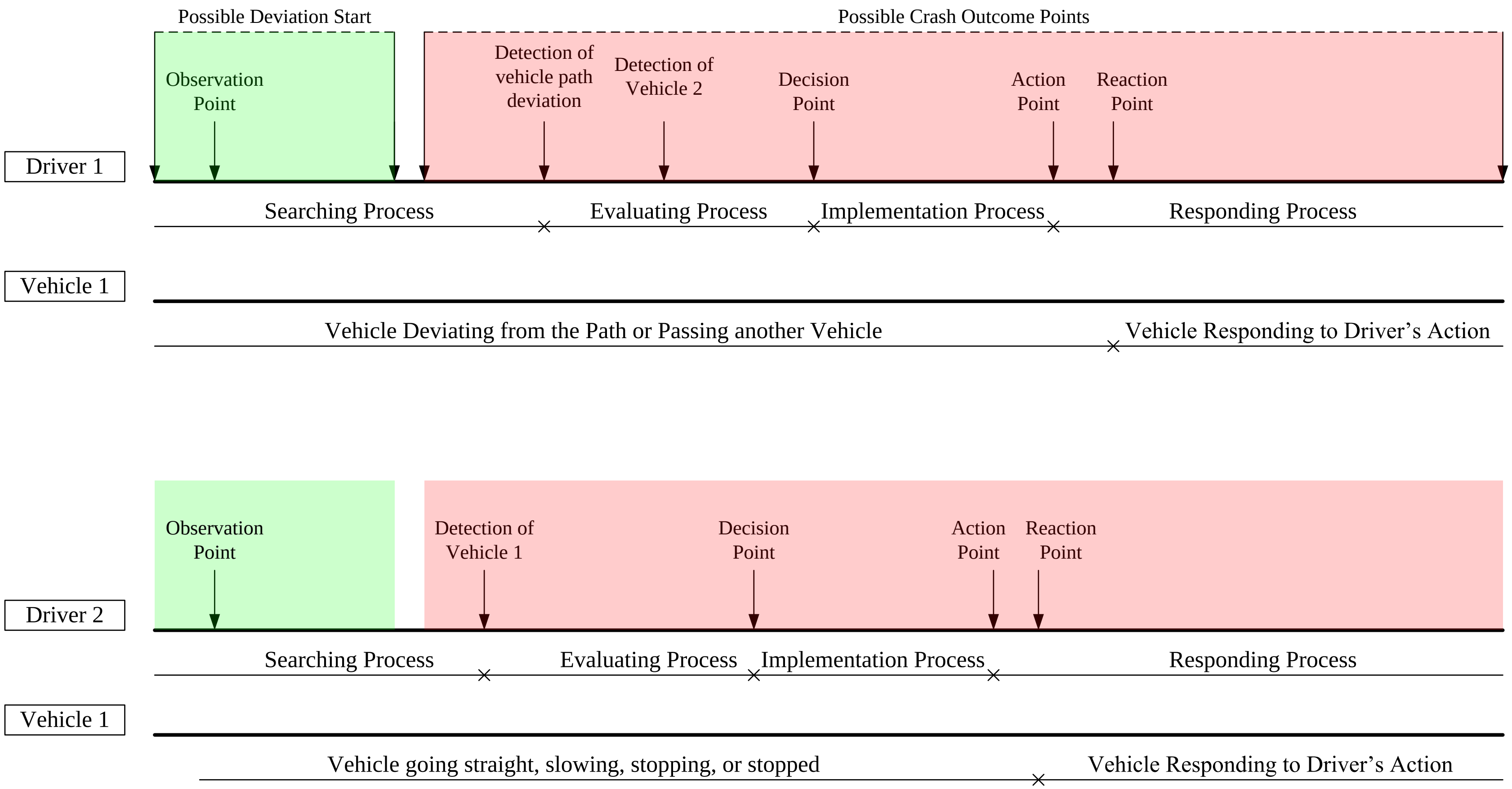
- Do traffic conditions contribute to the crash?
    - Traffic conditions might contribute to the crash in terms of volumes present on the road in general, and the time of the day when crash occurs—whether it is peak or off-peak period. Higher volumes mean higher exposure, and the probability of a vehicle being present in the opposing direction at the same time of an opposing lane encroachment increases with traffic volume.
- Does the environment contribute to the crash?
    - Environment can contribute to the crash in the case of adverse weather conditions that reduce driver visibility, increase driver workload, and impact pavement surface conditions.
- What combination of elements and their features contributed to the crash/caused two vehicles to reach the crash point at the same time?
    - The conceptual models treat crashes as emergent rather than sequential events. This means that the model outcome is the result of combined effects of driver behaviors, vehicle characteristics, roadway, traffic and environment features, and the efficiency of system constraints. Even though events related to each element (in this case both drivers and both vehicles) could be placed on a timeline, and each event does occur at a different time, this does not imply that two events occurring in sequence have a direct causal relationship.

These and other related questions are captured in the part of the conceptual crash model covering search, detection, evaluation, decision, and actions, defined in table 13.

Looking at the timeline (figure 13), this crash occurs when two vehicles meet at the same point in space and time. One of them is maintaining its intended path and, upon detection of the encroaching vehicle, attempting to avoid a collision (Driver/Vehicle 2). The other driver (Driver/Vehicle 1) drifts into the opposite direction travel lane. Driver 1 needs to detect (or not) two possible threats as a precondition to avoid the crash: deviation from their intended path and Vehicle 2 coming from the opposite direction. Driver 2 needs to detect Vehicle 1 encroaching from the opposite direction into their travel lane. This is how the timeline develops for both drivers, accounting for all the threats that drivers ought to detect, and the timeline for vehicle actions follows the timelines for their corresponding drivers.

Table 13. Conceptual Crash Models: Driver processes and events corresponding to crash scenario 22.

Process	Driver/Vehicle 1	Driver/Vehicle 2
Search—scanning, perpetual process	Driver actively searches for potential threats on the road, being aware of curvature, weather conditions, traffic	Driver actively searches for potential threats on the road, being aware of curvature, weather conditions, traffic
Detection—actual perception, awareness	Driver detects that the vehicle is drifting into the opposite direction, and a vehicle coming from the opposite direction	Driver detects vehicle coming from the opposite direction
Evaluation—recognition of threat, evaluate the need to avoid that threat	Driver evaluates the possibility of encroachment into the vehicle coming from the opposite direction	Driver evaluates the possibility of being hit by another vehicle
Decision—determination of the action	Driver determines the action: no response, slowing, stopping, warning another driver	Driver determines the action: no response, slowing, stopping
Human action—implementing decision	Driver implements previously determined action: in time, with a delay, has no time to implement decision	Driver implements previously determined action: in time, with a delay, has no time to implement decision
Vehicle action—vehicle responding to driver's action	Possible maneuvers: going straight, slowing, stopping, passing	Possible maneuvers: going straight, slowing, stopping, starting in the roadway, parking, stopping in travel lane



Source: FHWA.

Figure 13. Graphic. Example anchor points corresponding to crash scenario 22 for assessing time available and time used by drivers to perform functions.

DATA FOR PREDISPOSING FACTORS AND TARGET GROUPS

As part of the effort to marry the conceptual crash models with safety data, the project team assessed the availability and quality of data on driver, vehicle, road, traffic, and environmental characteristics that corresponded to the predisposing factors in the conceptual crash model frameworks. The project team also assessed the availability and quality of data on system constraints and precipitating events. The team executed the assessments using the following datasets: NMVCCS, GES, four HSIS States (Illinois, Minnesota, North Carolina, and Washington), Utah, and NPS. The main intent of this database assessment was to determine what data are currently available in traditional data sources to support learning how and to what level human, vehicle, roadway, and environmental elements contribute to crash occurrence. This is a logical first step to identifying opportunities to more effectively identify and understand the direct and indirect effects of these elements using enhanced data from various traditional and non-traditional data sources. The project team examined each database for the presence of predisposing factor and target group information. Target groups are human populations and/or kinds of physical locations involved in the crash type. The summary tables include descriptions of whether the data are included in the specific database, the presence of missing data, and whether it is possible to infer target groups or predisposing factors from other data elements included in the database.

Overall, a wealth of information is collected about vehicle, operator, and environmental data for police crash reports. It is typical to record causative and precipitating factors, but these rely on officers assigning responsibility, an admission of fault, or documenting blame from the vehicle operator. For these reasons, Shinar et al. and Austin found that coded vehicle and operator characteristics are quite unreliable.[34,35] Availability and reliability/accuracy of crash data for seven crash databases are presented in separate sections for vehicle, operator, environmental, roadway, and traffic characteristics. This study considers roadway characteristics for rural two-lane, highway horizontal curves and stop-controlled intersections, with stop control on the minor approaches. It then presents alternative sources for missing data and potential uses and considers areas of understanding.

Analysis of Existing Databases

Vehicle Data

Table 14 and table 15 present the data availability related to elements that characterize vehicle target groups and predisposing factors and constraints identified in the conceptual crash models. Target groups are human populations and/or kinds of physical locations involved in the crash type. Predisposing factors are specific environmental, human, or vehicle variables that influenced the function failure. For the eight databases considered in this section, most data are collected in some form. The legend at the bottom of table 15 contains a description of each of the symbols.

Table 14. Vehicle data—target groups.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Vehicle type	●	●	●	●	●	●	●	●
Vehicle year/model	●	●	●	●	●	●	●	●

●Related variable and data are available (<25 percent of cells are missing data).

Table 15. Vehicle data—predisposing factors.

Variable	NMV CCS	GES	GES	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Tire condition	●	●	D	●	●	●	●	●
Brake condition	●	●	D	●	●	●	●	●
Steering/alignment accuracy	●	●	D	●	●	●	●	--
Blind spots	●	●/I	--	--	--	--	--	--
Windshield/wiper condition	●	●	--	--	●	--	●	●
Headlight/tailing condition	●	●	D	●	●	●	●	●
Mirror presence/ condition	●	●	--	--	--	--	●	--
Side and rear window condition	●	●	--	--	--	●	●	●

D = related variable included but has been discontinued; I = variable can be inferred from other data.
●Related variable and data are available (<25 percent of cells are missing data).
--No related variable included.

It is important to note that most vehicle data are present as a subset of another categorical variable. Specifically, a range of vehicle condition variables are typically coded within a variable called “vehicle condition” or “vehicle defect.” None of the databases the project team reviewed explicitly collected these data for each crash. The vehicle condition factors are typically only coded if they are identified as a prime contributing factor in the crash. “Blind spot” is shown to be a factor that researchers can infer from the GES database. In many cases, a variable that is not directly related is collected, and it is possible to infer the factor of interest. In the case of blind spot, the GES database has a category within visual obstruction for “obstructing angles on vehicle.” It is possible to infer this to mean “blind spot.” However, this is not a typically collected variable, and it is only coded in rare cases (although a result for the overall categorical variable is coded for each crash whether there is a visual obstruction or not). Only one database specifically collected blind spot. In NMVCCS, there is a specific coding for vehicle-related blind spots. All other vehicle characteristics were collected in at least two databases. Table 16 and table 17 show the perceived reliability/accuracy of the vehicle characteristics.

The reliability categories in table 16 and table 17 are subjective at this stage, but generally refer to the following characterizations:

- Reliable data for a variable indicates correct coding nearly all (more than 90 percent) of the time.
- Unreliable data for a variable indicates correct coding less than half (i.e., less than 50 percent) of the time.
- Medium data for a variable fall somewhere in between reliable and unreliable.

Table 16. Perceived reliability/accuracy of vehicle data—target groups.

Variable	NMVCCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Vehicle type	●	●	●	●	●	●	●	●
Vehicle year/model	●	●	●	●	●	●	●	●

●Data for variable/database is reliable.

Table 17. Perceived reliability/accuracy of vehicle data—predisposing factors.

Variable	NMVCCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Tire condition	●	--	--	--	--	--	--	--
Brake condition	●	--	-	-	-	-	-	-
Steering/alignment accuracy	●	--	--	--	--	--	--	-
Blind spots	●	--	-	-	-	-	-	-
Windshield/wiper condition	●	--	-	-	--	-	--	--
Headlight/tailing condition	●	--	--	--	--	--	--	--
Mirror presence/condition	●	--	-	-	-	-	--	--
Side and rear window condition	●	--	-	-	-	--	--	--

●Data for variable/database is reliable.
--Data is unreliable.
-Data is not collected.

Table 16 and table 17 reflect the findings of Shinar et al. for police reported data, because the data collected within all databases comes from field police reports.(34) Typically the vehicle type and vehicle model/year are extremely reliable for reportable crashes. However, Shinar et al. noted a large amount of omission error for vehicle defects and contributing factors.(34) The mistakes were not typically miscoding in terms of false positives (e.g., brakes were not coded as being a cause if that was not indeed the case), but the coding was typically blank, or defects were not coded as a factor when they actually were. Mynatt et al. note that NMVCCS data for pre-crash variables are quite accurate since researchers collect and report the data instead of using police reported data.(36) The researchers note that the unique perspectives of the researchers collecting the NMVCCS data allow for a high degree of accuracy.

Driver Data

Driver characteristics are presented in table 18 and table 19. Washington had been collecting driver occupation from which it was possible to infer income but this has been discontinued. The target group variables of age and gender are the two most commonly collected operator variables. It is possible to infer driving experience from driver age from all databases, but Illinois has a category for driving experience within the crash cause categorical variable. Familiarity with location can be inferred from driver's license State as well as the distance between the driver's residence and the crash location. Although Washington had reported information about the vehicle operator's home being within 15 mi of the crash site, the variable has been discontinued. It is generally possible to infer risk-taking characteristics from other variables, such as exceeding limit or safe speed, illegal passing maneuver, reckless driving, not wearing seat belt, following too closely, or disregarding traffic control devices. Drug and alcohol impairment, fatigue, operator distraction, and operator vision are typically coded within a parent category labeled physical condition. Table 20 and table 21 summarize the perceived reliability and accuracy of the driver data. As expected, the target group data that is collected is generally reliable, but the reliability of the predisposing factors is called into question. Shinar et al. reported that police officers adequately assess direct human involvement causes and drug/alcohol causes in their investigations.[34] Utah has issues with driver age; approximately 10 percent of observations in the database include ages from 0 to 13.

Table 18. Driver data—target groups.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Age	●	●	●	●	●	●	●	●
Sex	●	●	●	●	●	●	●	●
Income	--	--	--	I/D	--	--	--	--

D = related variable included but has been discontinued; I = variable can be inferred from other data.
●Related variable and data are available (<25 percent of cells are missing data).
--No related variable included.

Table 19. Driver data—predisposing factors.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Driving experience	●	●/I	●/I	●/I	●	●	●/I	●/I
Familiarity with location	●	●/I	●/I	D	●/I	I/D	●/I	●/I
Risk taking characteristics	●/I	●/I	●/I	●/I	●/I	●/I	●/I	●/I
Drug/alcohol impairment	●	●	●	●	●	●	●	●
Fatigue	●	●	●	●	●	●	●	●
Level of self-caused distraction	●	●	--	--	●	●	●	--
Level of distraction from other occupants	●	●	●/I	●/I	●/I	●/I	●	--
Operator vision	●	--	●	--	●/I	--	--	--
Operator performance capability	●	--	●	--	●/I	--	--	--
Travel speed	●	●	●	--	--	--	○	--
Corrective action	●	●	●	●	●	●	●	●

D = related variable included but has been discontinued; I = variable can be inferred from other data.
●Related variable and data are available (<25 percent of cells are missing data).
○Related variable included but data are sometimes (>25 percent, <50 percent) missing.
--No related variable included.

Table 20. Perceived reliability/accuracy of vehicle driver data—target groups.

Variable	NMVC CS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Age	●	●	●	●	●	●	○	●
Sex	●	●	●	●	●	●	●	●
Income	--	--	--	--	--	--	--	--

●Data for variable/database is reliable.
○Data for variable/database has “medium” reliability.
--Data is not collected.

Table 21. Perceived reliability/accuracy of vehicle driver data—predisposing factors.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Driving experience	●	--	--	--	--	--	--	--
Familiarity with location	●	--	--	--	--	--	--	--
Risk taking characteristics	●	○	○	○	○	○	○	○
Drug/alcohol impairment	●	○	○	○	○	○	○	○
Fatigue	●	--	--	--	--	--	--	--
Level of self-caused distraction	●	--	-	-	--	--	--	-
Level of distraction from other occupants	●	--	--	--	--	--	--	-
Operator vision	●	-	--	-	--	-	-	-
Operator performance capability	●	-	--	-	--	-	-	-
Travel speed	●	--	--	-	-	-	--	
Corrective action	●	--	--	--	--	--	--	--

●Data for variable/database is reliable.
○Data for variable/database has "medium" reliability.
--Data is unreliable.
-Data is not collected.

Environmental Data

Table 22 and table 23 present the predisposing factor data available for environmental characteristics. Of the data categories, environmental data is the most commonly collected by each database reviewed. All agencies have available data for the environmental characteristics, except for visibility in Utah's dataset. However, Illinois experiences missing visibility data. Natural light condition, weather, and weather-related surface condition all exist as independent variables that are collected for every crash. Visibility, animal presence, and obstructions on the roadway are typically coded as an indicator within a higher-level categorical variable. The Utah database is the only one examined that does not specifically have a category for animal in roadway, but does have a category for striking an animal. This will only allow for crashes to be recorded if an animal is struck, but will not include crashes that may occur from attempting to avoid an animal.

The perceived reliability/accuracy of the environmental data is presented in table 23. The research by Shinar et al. shows that natural light condition and weather have a tendency to be somewhat reliable.(34) Weather-related surface conditions and visibility are cited as having extremely poor accuracy. It is worth noting that surface condition tends to be confused with the weather at the time of the crash. For example, the roadway could be wet, but could be coded as being dry if it is not currently raining. This reliability issue does not exist with NMVCCS data, an important point that will serves as a basis for the analysis of NMVCCS analysis presented in chapter 8. Police correctly identified sight obstructions in only three percent of crashes in the study by Shinar et al.(34) The reliability and accuracy of data about animal or obstruction presence is unknown at this time.

Table 22. Environmental data—predisposing factors.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Natural light condition	●	●	●	●	●	●	●	●
Weather	●	●	●	●	●	●	●	●
Weather-related surface condition	●	●	●	●	●	●	●	●
Visibility	●	●	●	●	●	○	--	●
Animal presence	●	●	●	●	●	●	●	●
Obstruction in roadway	●	●	●	●	●	●	●/I	●

I = variable can be inferred from other data.
●Related variable and data are available (<25 percent of cells are missing data).
○Related variable included but data are sometimes (>25 percent, <50 percent) missing.
--No related variable included.

Table 23. Perceived reliability/accuracy of environmental data—predisposing factors.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Natural light condition	●	○	○	○	○	○	○	○
Weather	●	○	○	○	○	○	○	○
Weather-related surface condition	●	--	--	--	--	--	--	--
Visibility	●	--	--	--	--	--	-	--
Animal presence	●	Δ	Δ	Δ	Δ	Δ	Δ	Δ
Obstruction in roadway	●	Δ	Δ	Δ	Δ	Δ	Δ	Δ

●Data for variable/database is reliable.
○Data for variable/database has "medium" reliability.
--Data is unreliable.
-Data is not collected.
ΔReliability unknown.

Roadway Data

Table 24 and table 25 present a summary of the availability of roadway data for horizontal curve crashes. These tables show that many roadway characteristics are not available in any of the databases. Illinois HSIS data files collect horizontal curve radius and grade, but only for those elements that are substandard, (i.e., they do not meet the minimum design criteria). Washington HSIS data collects superelevation rate data, but only for recent reconstruction efforts. Washington also collects data on vertical curve length, and entering and exiting grade, which allows for the calculation of the rate of vertical curvature. The NMVCCS database contains superelevation rate for horizontal curves, but approximately 50 percent of the superelevation rate data are missing. North Carolina HSIS has data for posted speed limit and design speed for roadway segments. From this, it is possible to infer the relationship, but data do not exist for individual horizontal curve elements. Several States have roadway lighting data in the light condition variable, if the crash occurred at night and there was roadway lighting. Lighting will also exist for daytime crashes that may go uncoded, but researchers do not typically consider the relationship between daytime crashes and roadway lighting. Databases that have a specific roadway lighting variable include Minnesota and Washington, and those are only for intersections. The NMVCCS is the only database that contains rumble strip and sign presence and condition data. NMVCCS data also lists sight distance as a contributing factor, but it is not directly measured. An important objective of one of the NMVCCS analysis was to show that researchers can use crash coordinates, traditionally not available for analysis, to determine the crash location and more specific geometrics at that location.

The perceived accuracy and reliability of roadway data is presented in table 26. Roadway data that DOTs collect are generally reliable, but not extensive. Surface type is listed as having medium reliability because if it comes from State databases, it is generally moderately reliable. The database must undergo continuous updates to reflect pavement resurfacing projects. HSIS guidebooks indicate that lane width, shoulder width, and shoulder type have a medium level of accuracy. Lane width is typically taken as the pavement surface width divided by the number of lanes, but this may be inaccurate in cases where it is measured from curb to curb. Striping determines lane width, and outside lanes may be painted as "wider" than inside lanes. There is little information to assess the reliability and accuracy of lighting presence and work zone status.

Table 24. Roadway data (for horizontal curve crashes)—target groups.

Variable	NMVCCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Roadway type	●	D	●	●	●	●	●	--
Roadway classification	--	●	●	●	●	●	M	--
Surface type	●	--	●	●	●	●	●	--
Curve radius	●	--	--	●	--	●/S	--	--
Cross-slope (superelevation)	○	--	--	M	--	--	--	--
Grade	●/I	--	--	●	--	●/S	--	--
Rate of vertical curvature	--	--	--	●/I	--	--	--	--
Access density	--	--	--	--	--	--	--	--
Shoulder slope	--	--	--	--	--	--	--	--
Lane width	●/I	--	--	●	●	●	--	--
Shoulder width	●/I	--	●	●	●	●	--	--
Shoulder type	--	--	●	○	●	●	--	--
Horizontal clearance	--	--	--	--	--	--	--	--
Barrier presence and offset	--	--	--	--	--	--	--	--
Rumble strip presence	●	--	--	--	--	--	--	--
Marking presence/type/ condition	--	--	--	--	--	--	--	--
Sign presence/type/ condition	●	--	--	--	--	--	--	--
General roadside character	--	--	--	--	--	--	--	--

Variable	NMV CCS	GE S	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Ped presence/ accommodation	●	●	●	●	●	●	●	●
Bike presence/ accommodation	●	●	●	●	●	●	●	●
Lighting presence/ character	●	--	●	○	●	●	●	--
Posted speed limit	●	●	●	●	●	●	●	●

M = related variable included but data are often (>50 percent) missing; D = related variable included but has been discontinued; I = variable can be inferred from other data; S = related variable included, but only for substandard elements.
●Related variable and data are available (<25 percent of cells are missing data).
○Related variable included but data are sometimes (>25 percent, <50 percent) missing.
--No related variable included.

Table 25. Roadway data (for horizontal curve crashes)—predisposing factors.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Available sight distance	●/I	--	--	--	--	--	--	--
Posted/advisory/design speed	●	--	●/I	--	--	--	--	--
Up and downstream character	--	--	--	--	--	--	--	--
Road condition (not weather)	--	D	●	--	●	●	●	●
Friction supply	--	--	--	--	--	--	--	--
Work zone status	●	●	●	●	●	●	●	●

D = related variable included but has been discontinued; I = variable can be inferred from other data.
●Related variable and data are available (<25 percent of cells are missing data).
--No related variable included.

Table 26. Perceived reliability/accuracy of roadway data (for horizontal curve crashes)—target groups.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Roadway type	●	●	●	●	●	●	●	--
Roadway classification	--	●	●	●	●	●	●	--
Surface type	●	--	○	○	○	○	○	--
Curve radius	●	--	--	●	--	●	--	--
Cross-slope (superelevation)	○	--	--	●	--	--	--	--
Grade	●	--	--	●	--	●	--	--
Rate of vertical curvature	--	--	--	●	--	--	--	--
Access density	--	--	--	--	--	--	--	--
Shoulder slope	--	--	--	--	--	--	--	--
Lane width	●	--	--	○	○	○	--	--
Shoulder width	●	--	○	○	○	○	--	--
Shoulder type	--	--	○	○	○	○	--	--
Horizontal clearance	--	--	--	--	--	--	--	--
Barrier presence and offset	--	--	--	--	--	--	--	--
Rumble strip presence	●	--	--	--	--	--	--	--
Marking presence/type/ condition	--	--	--	--	--	--	--	--
Sign presence/type/ condition	●	--	--	--	--	--	--	--
General roadside character	--	--	--	--	--	--	--	--
Ped presence/ accommodation	●	●	●	●	●	●	●	●
Bike presence/ accommodation	●	●	●	●	●	●	●	●
Lighting presence/ character	●		Δ	Δ	Δ	Δ	Δ	Δ
Posted speed limit	●	○	○	○	○	○	○	○

●Related variable and data are available (<25 percent of cells are missing data).
○Related variable included but data are sometimes (>25 percent, <50 percent) missing.
--Data is not collected.
ΔReliability unknown.

Table 27. Perceived reliability/accuracy of roadway data (for horizontal curve crashes)—predisposing factors.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Available sight distance	●	-	-	-	-	-	-	-
Posted/advisory/design speed	●	-	●	-	-	-	-	-
Up and downstream character	-	-	-	-	-	-	-	-
Road condition (not weather)	-	-	○	-	○	○	○	○
Friction supply	-	-	-	-	-	-	-	-
Work zone status	●	Δ	Δ	Δ	Δ	Δ	Δ	Δ

●Data for variable/database is reliable.
○Data for variable/database has “medium” reliability.
-Data is not collected.
ΔReliability unknown.

Table 28 and table 29 summarize the availability of roadway data for unsignalized intersections from the databases. Researchers can find most data (intersection type, classification, horizontal alignment, vertical alignment, shoulder and lane widths, shoulder type, superelevation rate) from the roadway data files. Specific intersection components such as turning radii, island presence, etc., are typically missing. Researchers can find auxiliary lane presence and types in three of the HSIS databases, but there is no given information for the termini and transition information. As with the horizontal curve data, sign presence and condition and a sight distance contributing factor indicator are present in the NMVCCS dataset. Table 30 and table 31 summarize the perceived accuracy and reliability of the intersection data elements, which is consistent with the roadway data reliability in table 26.

Table 28. Roadway data (for unsignalized intersections)—target groups.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Intersection type	●	D	●	●	●	●	●	--
Intersection classification	--	●	●	●	●	●	●	--
Horizontal alignment of approach	●	--	--	●	--	●	●	--
Vertical alignment of approach	--	--	--	●	--	--	●	--
Auxiliary lane presence/type	--	--	--	●	●	●	--	--
Auxiliary lane terminal/transition	--	--	--	--	--	--	--	--
Shoulder width	●/I	--	●	●	●	●	--	--
Shoulder type	--	--	●	○	●	●	--	--
Shoulder slope	--	--	--	--	--	--	--	--
Island presence/type	--	--	--	--	--	--	--	--

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Lane width	●/I	--	--	●	●	●	--	--
Turning tapers/radii	--	--	--	--	--	--	--	--
Marking presence/type/ condition	--	--	--	--	--	--	--	--
Sign presence/type/ condition	●	--	--	--	--	--	--	--
General roadside character	--	--	--	--	--	--	--	--
Ped presence/ accommodation	●	●	●	●	●	●	●	●
Bike presence/ accommodation	●	●	●	●	●	●	●	●
Lighting presence/character	●	--	●	○	●	●	●	--

D = related variable included but has been discontinued; I = variable can be inferred from other data.
●Related variable and data are available (<25 percent of cells are missing data).
○Related variable included but data are sometimes (>25 percent, <50 percent) missing.
--No related variable included.

Table 29. Roadway data (for unsignalized intersections)—predisposing factors.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Intersection skew	--	--	--	--	--	--	--	--
Superelevation/cross-slope	○	--	--	M	--	--	--	--
Intersection sight distance	●/I	--	--	--	--	--	--	--
Road condition (not weather)	--	D	●	--	●	●	●	●
Friction supply	--	--	--	--	--	--	--	--

M = related variable included but data are often (>50 percent) missing; D = related variable included but has been discontinued; I = variable can be inferred from other data.
●Related variable and data are available (<25 percent of cells are missing data).
○Related variable included but data are sometimes (>25 percent, <50 percent) missing.
--No related variable included.

Table 30. Perceived reliability/accuracy of roadway data (for unsignalized intersections)—target groups.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Intersection type	●	Δ	●	●	●	●	●	-
Intersection classification	-	●	●	●	●	●	●	-
Horizontal alignment of approach	●	-	-	●	-	●	●	-
Vertical alignment of approach	-	-	-	●	-	-	●	-
Auxiliary lane presence/type	-	-	-	●	●	●	-	-
Auxiliary lane terminal/ transition	-	-	-	-	-	-	-	-
Shoulder width	●	-	○	○	○	○	-	-
Shoulder type	-	-	○	○	○	○	-	-
Shoulder slope	-	-	-	-	-	-	-	-
Island presence/type	-	-	-	-	-	-	-	-
Lane width	●	-	-	○	○	○	-	-
Turning tapers/radii	-	-	-	-	-	-	-	-
Marking presence/type/ condition	-	-	-	-	-	-	-	-
Sign presence/type/condition	●	-	-	-	-	-	-	-
General roadside character	-	-	-	-	-	-	-	-
Ped presence/accommodation	●	●	●	●	●	●	●	●
Bike presence/accommodation	●	●	●	●	●	●	●	●
Lighting presence/character	●	-	Δ	Δ	Δ	Δ	Δ	-

●Data for variable/database is reliable.
○Data for variable/database has “medium” reliability.
-Data is not collected.
ΔReliability unknown.

Table 31. Perceived reliability/accuracy of roadway data (for unsignalized intersections)—predisposing factors.

Variable	NMVCCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Intersection skew	-	-	-	-	-	-	-	-
Superelevation/cross-slope	○	-	-	●	-	-	-	-
Intersection sight distance	●	-	-	-	-	-	-	-
Road condition (not weather)	-	○	○	-	○	○	○	○
Friction supply	-	-	-	-	-	-	-	-

●Data for variable/database is reliable.
○Data for variable/database has "medium" reliability.
-Data is not collected.

Traffic Data

Table 32 through table 35 provide a summary of the traffic related data for both horizontal curve and unsignalized intersection related crashes. No data are typically gathered related to directional traffic volumes or turning movement counts. Planned turning movements specific to vehicles involved in a crash can be found from crash data in the form of "Vehicle 1" and "Vehicle 2" movements (if more than one vehicle is involved), whether they were going straight or making a turn. It is possible to infer percent heavy trucks for Minnesota and Illinois HSIS data, since AADT and truck AADT are known factors. Obtaining the value of percent heavy trucks requires making a simple calculation. A closely following vehicle indicator variable is collected for the NMVCCS database; it can be combined with the roadway geometry variable to establish if a vehicle was following too closely on the horizontal curve. It is also possible to infer queue presence from the NMVCCS data from indicators for traffic congestion. Table 36 provides the perceived reliability and accuracy of the traffic data that are currently collected. Generally, traffic data have a medium level of reliability and accuracy. AADT data are typically not collected for every stretch of roadway or for each year, meaning that in some cases the AADT is not up to date, or has been estimated or extrapolated from other sources. It is not possible to make a concrete estimate of reliability or accuracy of planned turning movements, but based on other factors found in crash reports, it is likely that the reliability and accuracy are low. Police officers must rely on those in the crash being truthful, and must rely on crash reconstruction to identify responsibility.

Table 32. Traffic data—roadway curve target groups.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Directional daily volumes	--	--	--	--	--	--	--	--
Directional hourly volumes	--	--	--	--	--	--	--	--
Percent trucks	--	--	●	●	●/I	●/I	--	--
Bidirectional AADT	--	--	●	●	●	●	M	--

M = related variable included but data are often (>50 percent) missing; I = variable can be inferred from other data.
●Related variable and data are available (<25 percent of cells are missing data).
--No related variable included.

Table 33. Traffic data—roadway curve predisposing groups.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Closely following vehicles	●/I	--	--	--	--	--	--	--
Opposing vehicles	--	--	--	--	--	--	--	--

I = variable can be inferred from other data.
●Related variable and data are available (<25 percent of cells are missing data).
--No related variable included.

Table 34. Traffic data—unsignalized intersection target groups.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Directional daily approach volumes	--	--	--	--	--	--	--	--
Directional hourly approach volumes	--	--	--	--	--	--	--	--
Turning movement volumes	--	--	--	--	--	--	--	--
AADT of intersecting roadway	--	--	●	●	●	●	--	--

●Related variable and data are available (<25 percent of cells are missing data).
--No related variable included.

Table 35. Traffic data—unsignalized intersection predisposing factors.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Queue presence on approach	●/I	--	--	--	--	--	--	--
Planned turning movement	●	--	●	●	●	●	--	--

I = variable can be inferred from other data.
●Related variable and data are available (<25 percent of cells are missing data).
--No related variable included.

Table 36. Perceived reliability/accuracy of traffic data—roadway curve target groups.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Directional daily volumes	-	-	-	-	-	-	-	-
Directional hourly volumes	-	-	-	-	-	-	-	-
Percent trucks	-	-	○	○	○	○	-	-
Bidirectional AADT	-	-	○	○	○	○	○	-

○Data for variable/database has "medium" reliability.
-Data is not collected.

Table 37. Perceived reliability/accuracy of traffic data—roadway curve predisposing factors.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Closely following vehicles	●	-	-	-	-	-	-	-
Opposing vehicles	-	-	-	-	-	-	-	-

●Data for variable/database is reliable.
-Data is not collected.

Table 38. Perceived reliability/accuracy of traffic data—unsignalized intersection target groups.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Directional daily approach volumes	-	-	-	-	-	-	-	-
Directional hourly approach volumes	-	-	-	-	-	-	-	-
Turning movement volumes	-	-	-	-	-	-	-	-
AADT of intersecting roadway	-	-	○	○	○	○	-	-

○Data for variable/database has "medium" reliability.
-Data is not collected.

Table 39. Perceived reliability/accuracy of traffic data—unsignalized intersection predisposing factors.

Variable	NMV CCS	GES	HSIS: NC	HSIS: WA	HSIS: MN	HSIS: IL	UT DOT	NPS
Queue presence on approach	●	-	-	-	-	-	-	-
Planned turning movement	●	-	Δ	Δ	Δ	Δ	-	-

●Data for variable/database is reliable.
-Data is not collected.
ΔReliability unknown.

In summary, most variables related to direct human, indirect human, vehicular, and environmental causes have different levels of reliability. Police are generally able to identify some direct human causes, such as "failure to yield" or "failure to stop," but are unable to reliably identify "speeding" or other "improper driving."[34] For vehicle factors, police reporting is marginally successful for "inadequate braking" only. Among environmental factors, police performance does not exceed the chance level for any factors cited, and the police only correctly implicate sight obstructions in three percent of crashes. For human indirect causes, police performance was adequate only for "had been drinking."

These results are likely due to the nature of the data collection form, with these elements consisting of categories that the reporting officer must select. Overwhelmingly, the errors found are error by omission, and typically not mislabeled. However, there are factors, such as roadway curvature and surface type, that are commonly mislabeled. These errors likely occur because officers complete their reports at the police station (sometimes significantly removed from the crash in time) instead of at the scene of the crash.

Other Data Alternatives

Table 40 presents the variables that did not exist in any of the databases using police crash records considered in this research. Alternative sources of data include privately accessible data, private data collection, publicly available data, State and locally collected data, and field data collection. Six data quality characteristics (sometimes referred to as the "six-pack")—timeliness, accuracy, completeness, consistency/uniformity, integration, and accessibility—are assessed for each variable in terms of low, medium, and high.

Timeliness refers to how up to date data are for specific timeframes. Variables that are constantly changing and would require a great deal of updates receive a low (L) value for this category. Variables that typically remain unchanged over time (e.g., roadway geometry) receive a rating of high (H). Variables that may need to be collected at continuous intervals receive a medium (M) rating, such as rumble strips or barrier presence and condition.

Data accuracy is also rated as low, medium, and high, considering the range of alternative sources of collection methods. Pavement friction constantly changes along a section of roadway, and by location within the travel lane, and would be considered to have a low level of accuracy. It is possible to collect variables that result in counts or indicators extremely accurately.

Data completeness is assessed in terms of how completely the data can be collected in combination with the time it would take to collect. If it is necessary to characterize a variable through field data collection, or if a significant amount of post-processing is required, the variable receives a low rating. If it is possible to collect data in a blanket manner with automated algorithms, it receives a high rating. If field collection is necessary, achieving completeness is unlikely, and it may take a considerable effort to achieve a complete database. Integration is not really an issue for most variables, as they are easy to integrate into current databases. The exception would be real-time data such as vehicle following and opposing vehicle presence. If the vehicle in the crash is equipped with onboard sensors and cameras that are similar to those found in the SHRP2 NDS study vehicles, then the data is relatively easy to integrate. The SHRP2 NDS data could be used to establish surrogate measures for vehicle following and opposing vehicle presence. However, as noted in chapter 1, the NDS was not a practical data alternative for this study.

Uniformity and consistency is also evaluated as being low, medium, or high. This measure is assessed considering the impacts of having multiple data collectors. A degree of uniformity will exist if the variable is a count or an indicator (e.g., asset is present versus absent). If subjectivity is necessary, such as judging the distance to roadside obstacles, a lower degree of uniformity/consistency will exist.

This study assesses data accessibility in terms of ease of access and cost. Public sources, such as Google® and Bing™ products, are highly accessible, but private databases and State databases require fees or hurdles in order to obtain the data. Field data collection sources also receive a low accessibility rating due to cost of data collection equipment, time for data collection, and processing time.

Considering the six data quality characteristics outlined above, it is possible to establish a priority of missing variables by summing values of high, medium, and low. "Upstream and downstream characteristics" has an H in every category, meaning that it is possible to perform data collection quickly and reliably with complete coverage. "Rumble strip presence," "barrier presence/offset," "intersection skew," and "access density" are also at the top of the list.

Variables with several categories of L include "horizontal clearance," "marking presence/type/condition," "general roadside character," and "friction supply." "Friction supply" has the most instances of low for the six data quality characteristics. This is intuitive, since it is necessary to measure friction supply, it is constantly changing, and it varies considerably across a single section of travel lane. Models of friction supply are worth considering, but they would require pavement age, traffic level, area type (weather conditions), percent trucks, pavement type, and historical friction supply data. Roadway friction is also a function of vehicle tire type, wear, and weather-related pavement surface conditions. For these reasons, friction supply is difficult to quantify or estimate reliably, especially over a roadway's lifespan.

The following are potential uses/areas of understanding:

- Why is there a prevalence of unreliable information about driver and vehicle contributions being collected for accident reports?
- Does the focus on trying to assign responsibility impact the assessment of contributing factors?
- Is there a way to remove the human element of using driver interviews?
- Can we do better in training officers about the importance of reliable data collection?
- Shinar recommended an improved design of police accident reporting forms. He cited the extreme lack of sensitivity in recording the presence of environmental problems due to the low frequency with which environmental factors are clearly involved.

- Shinar also notes that some sections of the form may be filled out back at the station if they are perceived to be unimportant and large elements of guessing will take place (surface type is commonly misidentified between asphalt and concrete).

- Would collecting fewer data elements result in a higher reliability for data collected? For example, if the crash location is verified, roadway geometry and asset information can be integrated from State databases.

Table 40. Missing variables and potential alternative sources.

Variable	Alternative Source	Six-Pack					
		T	**A**	**C**	**I**	**X**	**U/C**
Operator income	Census	L	L	M	H	H	H
Available sight distance	GIS, E, FC, PCR, S	H	M	L	H	L	M
Up and downstream character	E, P, GIS, PCR	H	H	H	H	H	H
Access density	E, P, GIS, PCR, M	M	H	H	H	H	M
Shoulder slope	E, PCR, FC, M	H	M	L	H	L	H
Horizontal clearance	E, P, FC, M	L	M	L	M	L	L
Barrier presence and offset	E, SV, V, FC, PCR, AS, S	M	H	H	H	H	M
Rumble strip presence	E, S, V, S, AS, PCR	M	H	H	H	H	H
Marking presence/type/ condition	E, FC, PCR, AS, M	L	M	L	M	L	L
Sign presence/type/condition	E, FC, PCR, AS, M	M	M	L	M	L	L
General roadside character	E, FC, PCR	M	L	L	M	L	L
Friction supply	HPMS, FC, M	L	L	L	M	L	L
Intersection skew	GIS, E, P, PCR, FC	H	H	H	H	H	M
Auxiliary lane terminal/ Transition	GIS, E, P, FC, PCR	H	H	M	M	H	M
Island presence/type	GIS, E, P, PCR, FC, AS	M	H	M	M	M	M
Intersection sight distance	GIS, E, P, FC, PCR, S	H	M	L	H	L	M
Turning tapers/radii	GIS, E, P, FC, PCR	H	H	M	M	M	M
Directional daily volumes	TL, DN, FC, AS, MPO, M	L	M	M	H	L	M
Directional hourly volumes	TL, DN, FC, AS, MPO, M	L	M	M	H	L	M
Closely following vehicles	SHRP2 NDS	H	H	L	L	L	H
Opposing vehicles	SHRP2 NDS	H	H	L	L	L	H
Directional daily approach Volumes	TL, DN, FC, AS, MPO, M	L	M	M	H	L	M
Directional hourly approach Volumes	TL, DN, FC, AS, MPO, M	L	M	M	H	L	M
Turning movement volumes	TL, DN, FC, AS, MPO, M	L	M	M	H	L	M
Queue presence on approach	TL, DN, FC, M	L	M	M	H	L	M

Alternative source legend: E = earthmine; FC = field collection, PCR = project/contract records; M = model estimation; GIS = GIS-based tools such as Google® Earth™/Bing™ Bird's Eye, etc.; HPMS = Highway Performance Monitoring System; SV = Street View™ level tools; P = pictometry; S = safety improvement databases; AS = asset specific database; TL = TrafficLand; DN = DriveNet; MPO = metropolitan planning organization; V = video logs.
Six-pack legend: T = timeliness; A = accuracy; C = completeness; I = integration; X = accessibility; U/C = uniformity/consistency.

CHAPTER 7. EXPLORING PRECIPITATING EVENTS AND CAUSAL TYPE GROUPINGS

Precipitating events capture the specific nature of the failure in the function/event sequence that leads to the collision. Without information on precipitating events, it is difficult, if not impossible, to develop crash "causal types." As defined in the conceptual models for the high-priority, pre-crash scenarios presented in chapter 6, precipitating events are related to driver's behavior and vehicle's responsiveness to driver's action in the pre-crash period. Precipitating events for all crash scenarios are included in the conceptual models presented in figure 9, figure 11, and figure 13. Unlike the predisposing factors, data about precipitating events are not available from the traditional or most non-traditional resources, because the nature of data is different. The only likely options are the police crash report narratives or naturalistic driving studies. The purpose of these studies was to compensate for limited existing information on precipitating events of crashes. These studies, therefore, focused on exploring precipitating events and developing crash "causal types" from data, photographs, and narratives developed from detailed, on-scene crash investigations available in NMVCCS. This chapter presents the methodology used to analyze the NMVCCS data and the detailed descriptions of crash causal types obtained from this effort.

DEVELOPING CRASH CAUSAL TYPES FROM NMVCCS DATA

The study focuses on developing crash "causal types," or similar crashes grouped together based on their key precipitating events, predisposing factors, and target groups. The project team analyzed the crashes individually using data, photographs, and narratives developed from detailed, on-scene crash investigations. The purpose of these studies was to explore whether or not examining individual crash reports, photographs, and other available evidence to better understand factors leading to a crash, instead of relying solely on data summaries and modeling of electronically-coded variables, holds promise from a safety research perspective. Previous efforts have suggested that summarizing or tabulating data from more than one case results in a loss of information and that drawing causal conclusions prior to creating summaries is more beneficial.[24,25] In other words, the interactions of circumstances and causes are needed to group crashes and identify countermeasures; circumstances alone are not enough. The project team expects the studies conducted as part of this portion of the research to do the following:

- Uncover more microscopic interpretations of known empirical associations—for example, shoulder-width CMFs indicate an increase in head-on collisions with a decrease in shoulder width. Several reasons for this finding are possible, but several reasons to support an opposite finding are also possible. This is, in a sense, one of the major challenges with establishing cause–effect relationships from observational studies utilizing regression models: "When theories are weak and data are insufficient, the causal content of a regression finding cannot be convincingly questioned."[33] Detailed, crash-by-crash investigations may verify or refute alternative interpretations or uncover other interpretations. This could partially address the noted challenge with establishing cause–effect relationships from observational road safety studies by establishing evidence-based theory that supports or refutes empirical findings, with the evidence built from studying the sequence of events that led to a crash and the role the roadway features had in the crash-generating process.

- Suggest causal type groupings of similar crashes based on the combination of their key precipitating events, predisposing factors, and target groups. The hope is that by knowing these causal types, corresponding "treatments" can be identified, developed, and implemented in more cost-effective ways than if traditional groupings are used (e.g., single-vehicle, run-off-road). The project team also expects the causal groupings to contribute to current conversations on alternative ways to classify crash types for more aggregate-level safety research and analysis (e.g., estimating SPFs and CMFs).

- Determine the utility of data developed from detailed, on-scene crash investigations, not typically used by highway and traffic engineers, for understanding how roadway factors fit into the crash-generating process for different crash types.

Data and Methodology

In order to develop crash "causal types," the project team used the NMVCCS dataset. There is a total of 6,949 crashes in the publicly available dataset. The project team developed SAS codes to extract the high-priority Volpe pre-crash scenario types identified earlier, that are the focus of the detailed data analysis for this demonstration. The results showed that 674 crashes are either single-vehicle, control loss/no vehicle action; single-vehicle, road edge departure/no maneuver; or two-vehicle, opposite direction/no maneuver crashes on two-lane horizontal curves. An additional 71 crashes are SCP crashes at unsignalized intersections along two-lane roads. No area type or coordinate information is publicly available for the NMVCCS crashes, so additional data collection was necessary to verify that the identified crashes were of interest (i.e., occurring on rural two-lane horizontal curves or at unsignalized intersections on rural two-lane roads).

The project team requested the GPS coordinates for these "high priority scenarios" in order to identify the rural two-lane crashes. FHWA provided the GPS coordinates to the project team for the 745 crashes identified above after receiving them from NHTSA. Several accuracy problems were immediately evident in the GPS coordinate data. Additional quality control checks were necessary.

The accuracy of the crash GPS coordinates was evaluated crash by crash. The project team placed Google® Earth™ markers at the crash locations. They then conducted visual inspection of each crash. This step involved checking the roadway and cross-section elements and comparing the coded information in the NMVCCS data files with the Google® Earth™ and/or Google® Street View™. By the end of this process, the project team classified each crash location, as identified by its GPS coordinates, into one of three categories: crash location appears to be correct, crash location might be correct, or crash location appears to be incorrect.

Using a similar crash-by-crash inspection process, the project team then assigned three area type classifications: urban, rural, or rural-like (i.e., inside an urban boundary, but with cross-section and roadside elements that were more rural in appearance). The project team used Google® Earth™, Google® Maps™, city limit lines in Google® Maps™, and population numbers from the U.S. Census Bureau to determine the area type classifications of the crashes. The project team also considered presence of roadside curb and gutter as a key criterion in determining the area type, particularly for those crashes classified as "rural-like." By the end of this process, each crash was associated with a level of confidence in location accuracy and an area type. The results are summarized in table 41 through table 43.

Table 41. Summary of location accuracy and area type assessments—single-vehicle, horizontal curve (Volpe pre-crash scenarios 4 and 8).

Location Accuracy	Number of Crashes	Total	Rural	Rural-Like	Urban
Correct location	359	171	63	86	22
Maybe	99	47	15	18	14
Incorrect location	287	117	--	--	--

--No data.

Table 42. Summary of location accuracy and area type assessments—multi-vehicle, horizontal curve (Volpe pre-crash scenarios 22).

Location Accuracy	Number of Crashes	Total	Rural	Rural-Like	Urban
Correct location	359	31	10	13	8
Maybe	99	12	1	5	6
Incorrect location	287	28	--	--	--

--No data.

Table 43. Summary of location accuracy and area type assessments—multi-vehicle, intersection (Volpe pre-crash scenarios 22).

Location Accuracy	Total	Rural	Rural-Like	Urban
Correct location	157	10	20	127
Maybe	40	1	4	35
Incorrect location	142	--	--	--

--No data.

The project team then used the NMVCCS case viewer to identify key predisposing factors and precipitating events for the crashes of interest, which was intended to lead to a set of crash causal types. The data obtained from the case viewer included detailed crash narratives, crash diagrams, selected roadway measurements, witness interview summaries, and detailed photographs from the crash site. Four members of the project team participated in the detailed reviews of the NMVCCS cases and in categorizing key precipitating events and predisposing factors for every crash to a point. After establishing a methodology, two data analysts with general highway and traffic engineering backgrounds drove the classification process.

DEVELOPING CRASH CAUSAL TYPES USING DATA FROM DETAILED, ON-SCENE CRASH INVESTIGATIONS: NMVCCS STUDY

The contributing causal factors in the studied crashes were determined through a close examination of the detailed case reports available through the NMVCCS case viewer. The project team studied in detail all crashes identified as fitting one of the high-priority crash scenarios in rural or "rural-like" areas. These included control loss/no vehicle action and road edge departure/no maneuver for single-vehicle crashes on horizontal curves, opposite direction/no maneuver crashes on horizontal curves, and SCP at non-signal for intersections. Information pertaining to the following factors was manually extracted from the case report documents or, when not available there, from the electronically coded database: gender of driver(s), age, time of crash occurrence, information on encroaching and reacting vehicle (where applicable), presence of objects, presence of alcohol and other drugs, weather and road surface factors, vehicular factors including tire condition and tread depth, estimated pre-crash speed, rider demographics, license status, and medical concerns (if any). The project team then categorized these factors into precipitating events and predisposing factors based on the knowledge and judgment of research analysts. The project team finally grouped crashes into causal types based on common sets of precipitating events and predisposing factors. This section presents the descriptions of crash causal types obtained from this effort. The previous section described the process of identifying the relevant crashes in the NMVCCS database.

Single-vehicle crashes had precipitating events and predisposing factors associated with the one vehicle involved in the crash. The project team studied multi-vehicle crashes by looking at the precipitating events and predisposing factors associated with both the encroaching and reacting vehicles. The project team identified the encroaching vehicle in multi-vehicle crashes from the case report summary, along with items such as driver medical condition, and any information on possible "risky" driving behavior. Movements of encroaching vehicles typically were associated with descriptions such as failed to negotiate a curve, attempted passing maneuver, drifted into the other lane, and disobeyed a traffic law or failed to give right of way.

NMVCCS Analysis for Horizontal Curve Crashes

The project team identified a total of 208 single-vehicle and multi-vehicle crashes matching the high-priority scenario descriptions for horizontal curves from the NMVCCS database. Table 44 presents the causal type grouping results for the 182 single-vehicle and the 26 multi-vehicle crashes. The project team identified seven crash causal types for these critical crash scenarios. Because the sequences of initial events leading to both the single-vehicle and multi-vehicle crashes were similar (e.g., a driver leaving the travel lane), the project team developed causal type groups to be applicable to both single-vehicle and multi-vehicle/head-on crash types.

Table 44. Crash causal types for single-vehicle and multi-vehicle crashes on horizontal curves.

Crash Type	Multi-Vehicle Horizontal Curve Crashes for Scenario 22	Single-Vehicle Horizontal Curve Crashes for Scenarios 4 and 8
Fell asleep (FA)	2 (7.69)	18 (9.89)
Roadway/curve conditions (MS)	5 (19.23)	45 (24.72)
Weather conditions/black ice, hydroplane* (MS)	3 (11.53)	32 (17.58)
Inattention and/or distraction, over centerline/continued in same direction (ER)	7 (26.92)	16 (8.79)
Sun in eyes (ER)	2 (7.69)	1 (0.54)
Unknown reason (ER)	4 (15.38)	20 (10.98)
Drunk driving (AR)	1 (3.84)	15 (8.24)
Tractor trailer related (FD)	2 (7.69)	N/A
Health related problems, heart attack, passed out (FD)	N/A	7 (3.84)
Reached for something/took eyes off the road while driving (LA)	N/A	22 (12.08)
Tire blow out, braking failure (VP)	N/A	6 (3.29)
All crashes	26 (100)	182 (100)

Note: Numbers in parentheses represent the percentage of crashes.
FA = drowsy and fell asleep; MS = misjudgment of speed; ER = encroachment; AR = alcohol related; FD = failed to look/detect; LA = looked away; VP = vehicle-related problems.

Drowsy and Fell Asleep (FA)

This category consists of 2 of the 26 multi-vehicle crashes and 18 of 182 single-vehicle crashes. This causal type refers to situations where drivers completely or nearly fell asleep while driving on the road. Any crash in which the report showed the driver as drowsy, sleepy, asleep, or seriously fatigued fell into this causal type. Analysis of the 20 available case reports indicated that 7 crashes occurred during an early weekday morning, when drivers were traveling to work, of which 2 drivers suffered from sleep apnea. Of the 20 total known cases, the age of 8 drivers was above 50 years old. The average hours of sleep for the drivers the night before the crash was 6.5 hours. Medication was detected in 10 of the 20 cases in the case reports. Fourteen of the drivers involved in the drowsy-driving crash type were male. Of the 20 cases, 8 drivers had a restricted license or had no license at all. Seven crashes occurred on roadways with a posted speed limit of 55 mph. Eleven crashes occurred on a roadway where the horizontal curve radius was less than 1310 ft. In general, there were no observed patterns in adverse atmospheric conditions at the time of the crash.

Misjudgment of Speed for Roadway, Curve, and Weather Conditions (MS)

This category consists of 8 of the 26 multi-vehicle crashes and 77 of the 182 single-vehicle crashes. This causal type refers to one of the most frequent crash types where the driver, due to weather conditions (i.e., rain, snow, wet, and/or icy roadway conditions), or lack of experience and familiarity with the roadway misjudges an appropriate operating speed that enables the vehicle to negotiate the curve within the travel lane. Analysis of the 85 available case reports indicated that there were 50 cases where drivers had been travelling too fast for the curve and/or roadway conditions. Thirty-nine of the 85 drivers were female drivers. Of the 85 total cases, 29 crashes occurred in the evening or at night, and 20 crashes occurred in the morning (6 a.m.–9 a.m.). There were seven drivers whose license had been previously suspended or revoked, or who never had a license at all.

Of the 85 known cases, reports indicated that 7 drivers had consumed alcohol prior to the crash. Sixty-seven of the 85 total crashes occurred on a roadway where the horizontal curve radius was less than 1310 ft, of which 31 crashes occurred on a roadway with a horizontal curve radius less than 330 ft. There were 35 drivers for whom excessive speed was identified and reported in the case report. Forty-four drivers were between 16 and 23 years old.

Encroachment: Inattention, Distraction, and Sun Glare (ER)

This category consists of 13 of the 26 multi-vehicle crashes and 37 of the 182 single-vehicle crashes. This crash type describes the cases where the driver fails to negotiate the curve due to internal and external distraction, inattention, sun glare, objects falling into their eye, and other unknown reasons. The driver continues on a straight path or a path that is drifting in one direction or the other and either encroaches into the opposite lane of travel or encroaches into the roadside. Stated reasons for distraction and/or inattention include the driver talking to the passengers in the vehicle, looking for another vehicle in the rear-view mirror, sun glare, and the driver blacking out, among others. Analysis of 50 case reports indicated that the ages of 13 drivers were less than 25 years. There were 21 female drivers in this category. Of the 50 total cases, 12 crashes occurred in the evening and at nights (after 6 p.m.). Twenty-three out of the 50 total crashes occurred due to distracted or inattentive driving. Of these, 4 crashes involved drivers focusing on or retrieving other objects in the vehicle. Thirty-one out of the 50 total crashes occurred on a roadway where the horizontal curve radius was less than 1,310 ft.

In the single-vehicle category, 8 out of 37 crashes occurred on curves to the right, with the first encroachment of the cases inside the curve at 50 percent (4 out of 8). Twenty-nine out of the 37 single-vehicle crashes occurred on curves to the left. In 25 of these 29 cases, the first encroachment was to the outside of the curve. For crashes on curves to the left, 19 out of 26 crashes occurred when drivers continued on a straight path or a path that was drifting in one direction or the other, and 27 percent occurred when the driver encroached into the roadside, then overcorrected and encroached into or through the opposite lane.

Alcohol Related: Drunk Driving (AR)

This category consists of 1 of the 26 multi-vehicle and 15 of the 182 single-vehicle crashes. Crashes were grouped into this category if the crash report indicated that the drivers involved had blood alcohol concentrations (BAC) levels greater than 0.08. Of the 16 total crashes, 7 occurred on a Friday night or weekend. In this category, the age of 5 drivers was below 24 years.

Nine of these crashes occurred during the mid- and late-evening on weekdays and weekends. Of the 16 cases, there were 12 male drivers in this category. Five of the 16 drivers in this category had BAC levels that exceeded 0.2. Ten of the 16 drivers had also used other drugs or medications prior to the crash.

Five of the 16 drivers had a revoked or suspended driver license; 2 had been emotionally upset prior to the crash. Three had rarely driven on the road where the crash occurred. Eleven of the 16 crashes occurred on a roadway where the horizontal curve radius was less than 1,310 ft; three of these crashes occurred on a roadway where the horizontal curve radius was less than 330 ft.

Failed to Look/Detect and Other (FD)

This category consists of 2 of the 26 multi-vehicle and 7 of the 182 single-vehicle crashes. This causal type captures several remaining crash types that were not prevalent enough to create separate categories for. One type refers to drivers who fail to detect opposing vehicles travelling in the opposite direction while passing a tractor trailer or box truck. The combination of a large leading vehicle and horizontal curve restricts the encroaching vehicle driver from detecting the oncoming vehicle. The driver then attempts a passing maneuver, which leads to a crash.

The crash type also includes drivers who have health problems resulting in passing out suddenly, heart attack, etc., ultimately leading to a crash. Of the nine total crashes, six involved male drivers. In 7 of the 9 crashes, the driver's age exceeded 50 years; all drivers involved in this crash causal type used medication. The drivers involved in six of the nine crashes used more than one type of drug. Seven out of 9 crashes occurred on a roadway where the horizontal curve radius was less than 1,310 ft.

Looked Away (LA)

This category consists of 22 of the 182 total single-vehicle related crashes. This causal type refers to drivers who while driving, reached for something (e.g., water bottle, pack of cigarettes, food), which resulted in a run-off-road crash. The driver took his/her eyes off the road for a short amount of time, resulting in the vehicle departing the roadway. Analysis of 22 crash case reports indicated that the ages of 12 drivers were 16 to 25 years old. There were 13 male drivers in this category. Four of the 22 total crashes occurred in the evening and at night (after 6 p.m.). Another four occurred in the morning (6 a.m. to 9 a.m.).

Three of the 22 crashes occurred when it was raining and the roadway was wet. Thirteen crashes occurred on a roadway where the horizontal curve radius was less than 1,310 ft of which one crash occurred on a roadway where the horizontal curve radius was less than 330 ft.

Vehicle Related Problems (VP)

This category consists of 6 of the 182 single-vehicle related crashes. This causal type refers to crashes that involved vehicle-related problems, such as tire blowouts or braking failure. Analysis of the 6 crash case reports indicated that the vehicle model year for all the vehicles was at least six years old at the time of the crash. Two of the six crashes occurred in the evening and at night (after 6 p.m.). The average tire tread depth for all vehicles involved in the crashes was 0.18 inches. Of the six total crashes, one occurred during rainy conditions, and the roadway was wet at the time of the crash. Three of the 6 crashes occurred on a roadway where the horizontal curve radius is less than 1,310 ft.

NMVCCS Analysis for Unsignalized Intersection Crashes

The project team identified a total of 33 SCPs at non-signal crashes from the NMVCCS database and developed 3 crash causal types to characterize these crashes. Table 45 presents preliminary results for the 33 identified SCP crashes at unsignalized intersections. Detailed definitions of these contributing causal factors and types are explained after the table.

Table 45. Crash causal types for multi-vehicle crashes at unsignalized intersections—multi-vehicle unsignalized intersection crashes (scenario 31).

Crash Type	Causal Factor	Number	Percent
Misjudgment of speed (MS)	Right of way vehicle’s speed	2	6.06
Misjudgment of speed (MS)	Roadway/weather conditions	2	6.06
Failed to give right of way (FR)	Inattention and/or distraction	7	21.21
Failed to give right of way (FR)	False assumptions	2	6.06
Failed to give right of way (FR)	Unknown reason	4	12.12
Inadequate surveillance (IS)	Recognition error	15	45.45
Inadequate surveillance (IS)	Obstructed view	1	3.03
All crashes	—	**33**	**100**

—Not applicable.

Misjudgment of Speed (MS)

This category consists of 4 of the 33 SCP crashes at unsignalized intersections. This causal type refers to drivers who misjudged the crossing vehicle’s speed (i.e., the speed of the vehicle with the right of way). This causal type also refers to the crash type where the driver entering from the minor road, due to weather conditions (i.e., rain, snow, wet, and/or icy roadway conditions) misjudges the speed for the roadway and is unable to perform the required maneuver (e.g., acceleration and turning) at the intersection. Similarly, the weather conditions have some impact on the ability of the driver on the major road to avoid the collision. The driver ages for the stopped vehicles was 15 to 18 years old. Two of the four total crashes occurred at intersections with significant skew.

Failed to Give Right of Way (FR)

This category consists of 13 of the 33 SCP crashes at unsignalized intersections. This causal type refers to drivers who were inattentive, distracted, made false assumptions of other’s actions due to lack of experience and other unknown reasons. Of the 13 total crash cases, 7 occurred because of inattention or distracted driving. The drivers involved in these crashes were talking to other passengers in the car, on the cell phone, adjusting controls in the car, or were in a state of dilemma. In four of these seven crashes, the vehicle on the major road approached from the left of the vehicle that should have given the right of way. Two of the 13 total crashes that occurred because the driver made false assumptions about the other vehicle’s requirement to stop.

Inadequate Surveillance (IS)

This category consists of 16 of the 33 SCP crashes at unsignalized intersections. This causal type refers to drivers proceeding into the intersection with inadequate surveillance, recognition error, or with an obstructed view (i.e., limited intersection sight distance). This crash type accounted for the majority of crashes at unsignalized intersections. Of the 16 total crash cases, 15 occurred because of recognition error, of which 4 crashes involved drivers younger than 20 years old. Seven of the 16 total crashes occurred at skewed intersections.

CHAPTER 8. DATA COLLECTION FOR CRITICAL CRASH SCENARIOS

This chapter summarizes data collection for three studies the project team executed using enhanced datasets, each assessing the potential of a specific enhanced and robust dataset for increasing the project team's understanding of the influence of various predisposing factors on safety. Specifically, the studies estimated and examined parameters that quantify relationships between expected number of crashes (by type) at a location during a defined time period and the predisposing roadway, traffic, and weather factors at that location. Based on the previously discussed conceptual crash models, the presence of one or more specific predisposing factors over time (e.g., rainy weather) does not in itself "cause" a crash. It does, however, by its presence, have some level of influence on whether the driving task will be carried out successfully by the driver. Therefore, one would expect that influential predisposing factors are associated with the expected number of crashes at an aggregate level. The focus of the methodology and interpretation was not on the specific parameter estimates. Rather, the methodology focused on demonstrating the construction and/or use of an enhanced dataset of predisposing factors and exploring whether the variables not typically collected appear to influence the expected number of crashes and act as a confounder for variables that are typically collected (and therefore result in possible over- or under-statements of such a traditional variable's influence on safety).

The three enhanced datasets used for analysis were as follows:

- **Unsignalized intersections along rural two-lane highways in North Carolina and Ohio**—built using a combination of State and local crash, traffic, and roadway inventory files, Google® Earth™, Google® Street View™, field measurements, and NOAA data. Predisposing factors in this dataset that are not traditionally available include intersection sight distance; vertical grade; intersection angle; pavement quality; weather patterns; and, the presence, type, and condition of various types of traffic-control devices.

- **Horizontal curves on rural two-lane highways in Washington**—built by combining State crash, traffic, and roadway inventory files with a detailed roadside-features inventory and NOAA data. Predisposing factors in this dataset that are not traditionally available include horizontal curve characteristics; vertical grade; and, the presence, type, and location of various roadside features.

- **Horizontal curves on rural two-lane highways in Utah**—built using data from an extensive effort by the UDOT to gather, identify, and process detailed information on all above-ground assets and road characteristics along State routes using LiDAR. The data collection effort appeared to be the first of its kind executed by a State DOT, and the dataset holds significant potential for safety analysis.

The following sections describe data collection procedures corresponding to the unsignalized intersection and horizontal curve scenarios. Chapter 9 covers analysis methodologies, results, and findings.

DATA COLLECTION PROCEDURES FOR INTERSECTIONS

The project team gathered most data elements discussed in this section initially for use in NCHRP Project 17-59, considering the safety impacts of intersection sight distance (ISD). The intersections used in this study consisted of the subset of four-leg stop-controlled intersections on rural two-lane highways. For NCHRP 17-59, data were collected at the intersection approach level, but for the current analysis, the project team aggregated the data to the intersection level for consistency with current methods employed in the *Highway Safety Manual*.[39]

The project team collected most elements as part of NCHRP 17-59, and the data collection procedures for those elements are summarized in the NCHRP project interim report. The project team collected additional data from the desktop specifically for this research for unsignalized intersections in North Carolina and Ohio. The project team collected data at additional sites in North Carolina, to increase the sample size, since SCP crashes are not common. For these sites, the project team collected field data, following the NCHRP 17-59 protocol.

Table 46 through table 49 provide the specific data variables for crash, roadway, and traffic/operations categories that the project team identified, and whether crash-based cross-sectional analyses typically consider the data to be required or desired. Required elements represent the minimum information necessary to conduct the analysis, and their incorporation into the model is essential to the essence of the study. Meanwhile, desired elements are more ancillary in nature and, if available, serve to enhance the analysis by allowing for a more detailed investigation. Note that while several of the variables in the tables are defined as desired, all variables were collected (both required (Req.) and desired (Des.)) for each study location. The summary tables also display the data collection method of each data element. The project team collected them either in the office or the field. It is necessary to field verify the data elements that are collected in the office because the available data is not always up to date, and it ensures that no physical significant changes have been done to the intersections of interest.

Table 46. Intersection crash data elements of interest.

Data Element	Need	Description/Comment	Collection Location	Data Sources
Location	Req.	Typically consists of county, route, and milepost	Office	HSIS database
Type	Req.	Often recorded as first harmful event	Office	HSIS database
Initial direction of vehicles	Req.	Direction of travel immediately before crash	Office	HSIS database
Sequence of events	Des.	Events of crash listed in chronological order	Office	HSIS database
Vehicle type	Des.	Identifies the type of vehicle based on police report	Office	HSIS database

Table 47. Intersection roadway data elements of interest.

Data Element	Need	Description/Comment	Collection Location	Data Sources
Major road functional class	Des.	E.g., principal/minor arterial, major/minor collector, local	Office	HSIS database
Minor road functional class	Des.	E.g., principal/minor arterial, major/minor collector, local	Office	HSIS database
Area type	Req.	E.g., rural, urban	Office	HSIS database
Number of through lanes	Req.	Both directions of the major road	Office (verified in field)	Google® Earth™
Divided/undivided major road	Req.	Undivided, divided, or two-way left-turn lane	Office (verified in field)	Google® Earth™
Standard available ISD	Des.	Measured looking left and right for each minor route approach	Field	Field
Best available ISD	Des.	Measured looking left and right for each minor route approach	Field	Field
Confirm visibility of traffic control	Des.	Simple yes/no confirmation that minor road driver has clear vantage of the traffic control device(s) at major road	Field	Field

Data Element	Need	Description/Comment	Collection Location	Data Sources
Horizontal curvature	Des.	Indicating presence and direction looking left and right at each minor route approach	Office (verified in field)	Google® Earth™
Lane width	Des.	Width of travel lanes in feet	Field	Field (NC), Google® Earth™ (OH)
Shoulder width	Des.	Width of paved shoulder in feet	Field	Field (NC), Google® Earth™ (OH)
Access density	Des.	Number of unsignalized intersections or driveways along major road within 0.25 mi of subject intersection (count a 4-leg intersection as 2)	Office	Google® Earth™
Major driveways	Des.	Number of driveways within 250 ft of intersection on major route	Office	Google® Earth™
Minor driveways	Des.	Number of driveways within 250 ft of intersection on minor route	Office	Google® Earth™
Lighting presence	Des.	Field check for luminaires at the intersection	Field	Field
Vertical grade	Des.	Measured with smart level at 3 locations (100, 250, and 500 ft) left and right along major road in both directions along the major route	Field	Field
Intersection angle	Des.	Approximate angle between major road and minor road	Office	Google® Earth™
Presence of left-turn lanes	Des.	Along major road in both directions	Office (verified in field)	Field
Presence of right-turn lanes	Des.	Along major road in both directions	Office (verified in field)	Field
Presence of left-turn lane	Des.	Along minor road for both approaches	Office (verified in field)	Field
Presence of right-turn lane	Des.	Along minor road for both approaches	Office (verified in field)	Field

Data Element	Need	Description/Comment	Collection Location	Data Sources
Terrain	Des.	Level, rolling, mountainous	Office (verified in field)	Field
Quality of sight distance	Des.	Appraise ISD quality in a subjective way (scores from 1–3 with 1 being least objects and 3 being many objects) in both directions along the major route	Office	Webinar

Table 48. Intersection traffic/operations data elements of interest.

Data Element	Need	Description/Comment	Collection Location	Data Sources
Traffic volume of major road	Req.	Including year that volume was determined	Office	State DOT, HSIS database
Traffic volume of minor road	Req.	Including year that volume was determined	Office/Field	State DOT, local agencies, field counts
Traffic control	Req.	Stop, yield, etc.	Office (verified in field)	Field
Posted speed limit	Des.	Nearest upstream posted speed limit on major route	Field	Field
Presence of on-street parking	Des.	Along major road	Office (verified in field)	Field
Pavement deficiency	Des.	Presence of observable cracking, pot holes, rutting, etc.	Office	Google® Street View™
Stop bar	Des.	Presence of minor route stop bar	Office	Google® Street View™
Number of stop signs	Des.	Presence of multiple stop signs on minor route approaches	Office	Google® Street View™
Reflective post	Des.	Presence of red reflective panel on stop sign support	Office	Google® Street View™
Large sign	Des.	Presence of oversized stop signs	Office	Google® Street View™
Number of intersection warning signs	Des.	The presence and number of advance intersection warning signs on major route	Office	Google® Street View™

Data Element	Need	Description/Comment	Collection Location	Data Sources
Distance to intersection warning sign	Des.	The distance from the intersection to the advance intersection warning sign	Office	Google® Earth™
Number of stop ahead warning signs	Des.	The presence and number of advance stop ahead warning signs on minor route	Office	Google® Street View™
Distance to stop ahead warning sign	Des.	The distance from the intersection to the advance stop ahead warning sign	Office	Google® Earth™
Through-edge line extension	Des.	The presence of an extension of the through edge line using a short skip pattern	Office	Google® Street View™
Advisory speed	Des.	The presence of an advisory speed plaque on major route	Office	Google® Street View™
Speed reduction	Des.	Speed reduction from posted speed to advisory speed on major route in mph	Office	Google® Street View™
Rumble	Des.	Presence of rumble strips on major route	Office	Google® Street View™
Major RRPMs	Des.	Presence of raised reflective pavement markings on the major route	Office	Google® Street View™
Minor RRPMs	Des.	Presence of raised reflective pavement markings on the minor route	Office	Google® Street View™

Table 49. Intersection weather data elements of interest.

Data Element	Need	Description/Comment	Collection Location	Data Sources
Mean temperature	Des.	Mean annual temperature of intersection in degrees	Office	NOAA
Number of 90 plus days	Des.	Number of days where high temperature reaches 90 °F	Office	NOAA
Number of 32 less days	Des.	Number of days where low temperature reaches 32 °F	Office	NOAA
Total snow	Des.	Total annual snowfall recorded at intersection in inches	Office	NOAA
Days with precipitation	Des.	Number of days annually receiving one tenth, one half, and 1 inch of rain	Office	NOAA

As previously stated, the project team gathered most data elements initially for use in NCHRP Project 17-59, conducted from May 2012 through November 2014, and the interim report provides information on the data collection strategies. Information on additional data elements collected for this study are provided below.

Pavement Deficiency

Using the latest imagery from Google® Earth™ and Google® Street View™, the project team developed an indicator for deficient pavement conditions based on observable cracks, pot holes, rutting, etc. If the project team noted any of these factors for a site, the indicator received a value of one; otherwise the indicator received a zero. It is important to note that using images of this type is only useful for a qualitative review of pavement conditions.

Stop Bars

The project team used Google® Earth and Google® Street View to determine whether or not the minor road had a stop bar. If only one approach had a stop bar, the project team coded the intersection as having stop bars present. This circumstance occurred infrequently and usually resulted from one approach having a faded or worn out stop bar. The project team used Google® Street View™ historical imagery to confirm those instances.

Dual Stop Signs

The project team used Google® Street View™ to determine whether one or two stop signs were used at the minor approaches. The project team gave a dual stop sign indicator a value of one if one or both approaches had two stop signs. In some cases, the stop signs were off the pavement on both the left and right sides of the minor route approach. In other cases, the left stop sign was placed on a channelized median in the center of the roadway. If only one stop sign was present on both minor route approaches, the value was given as zero.

Retroreflective Panel on Sign Post

The project team used Google® Street View™ to determine whether or not a red retroreflective panel was mounted to the support of the stop sign(s). They coded it as present if the panel appeared on one or both approaches; otherwise the indicator received a zero.

Large Sign

The project team used Google® Street View™ to determine whether or not oversized stop signs were used at the minor approaches. They coded it as present if oversized signs appeared on one or both approaches. The Manual on Uniform Traffic Control Devices (MUTCD) standard stop sign is 30 by 30 inches; therefore, the indicator received a value of one if the stop sign was visibly larger than a standard sign through a comparative evaluation. The large signs frequently were identified by having two posts to accommodate for the added size of the sign; otherwise the indicator received a zero.

Advance Intersection Warning Sign

The project team used Google® Street View™ to determine if an advance intersection warning sign (MUTCD signs W2-1 through W2-8) was present on the major road before the intersection. These signs were sometimes difficult to find because the distance from the sign to the intersection varied. This resulted in using Google® Street View™ to incrementally move down the road until the sign was located. If the sign was not present within 1,500 ft of the intersection, it was coded as not having an advance intersection warning sign present. The project team used a separate indicator variable to record the presence of double advance intersection warning signs (one warning sign on each side of the roadway). Additionally, the project team recorded the distance from the warning signs to the center of the intersection for intersections with advance warning signs on the major route.

Advance Stop Ahead Warning Sign

An indicator for the presence of an advance stop ahead warning sign (either stop sign symbol and arrow (MUTCD sign W3-1) or stop ahead) on the minor road approaching the intersection was determined using Google® Street View™. Similar to advance intersection warning signs, these signs were sometimes difficult to find because the distances from the sign to the intersection varied. If the sign was not present with 1,500 ft of the intersection, the project team coded it as not having an advance stop ahead warning sign present. The project team used a separate indicator variable to record the presence of double advance stop ahead warning signs (one warning sign on each side of the roadway). Additionally, they recorded the distance from the warning signs to the center of the intersection for intersections with advance warnings signs on the minor route.

Through-Lane Edge Line Extension

The project team used Google® Earth™ to identify the presence of a through-lane edge line extension on the major route of the four-leg intersection. The through-lane edge line consisted of a skip line painted at intersections with either a high-crash history or where vehicles are known to stop back a far distance from the edge line of the major route. The project team coded the indicator as one if the skip line was present; otherwise they coded it as zero.

Advisory Speed Plaque

The project team used Google® Street View™ to develop an indicator for the presence of an advisory speed plaque (MUTCD sign W13-1P). If an advisory speed plaque was attached to an advance intersection warning sign on the major route, a value of one was coded, otherwise the value was zero.

Speed Reduction

The project team developed the speed reduction by taking the difference between the posted speed limit and the posted advisory speed, in mph. If they found no speed reduction between the posted speed limit and the advisory speed, they recorded a value of zero.

Rumble Strips

An indicator for the presence of rumble strips on the major route was determined by using the latest imagery from Google® Street View™. The project team recorded a value of one if rumble strips were present; otherwise they recorded a value of zero.

Weather Data Collection

The project team obtained weather data from the NOAA National Climatic Data Center. The project team collected land-based station data from the nearest stations to the study intersections. In total, there were eight weather stations in North Carolina and seven weather stations in Ohio. The nearest weather stations were generally located approximately 15 to 20 mi away from most intersections. Data from land-based stations included temperature, dew point, relative humidity, precipitation, wind speed and direction, visibility, atmospheric pressure, and types of weather such as hail, fog, and thunder.[(40)] Data reporting occurred hourly, daily, monthly, and annually. Data quality and biases were captured and reported for land-based stations. This level of detail allowed for data capture at the time of individual crashes (for most stations). Aggregate yearly data were collected for this project to match the time scale used for crash and AADT data at the intersection level. The project team recorded the following data elements for each intersection included in the study:

- MeanTemp—mean temperature for the intersection over the course of the year, in degrees Fahrenheit.
- NDays90Plus—number of days within the year the high temperature at the intersection reached 90 °F or greater, in days.
- NDays32Less—number of days within the year the low temperature at the intersection reached 32 °F or less, in days.
- TotalSnow—total snowfall recorded at the intersection for the year, in inches of snowfall.
- NDaysPreciptenthinch—number of days with at least 0.1 of an inch of precipitation at the intersection, in days.
- NDaysPreciphalf—number of days with at least 0.5 inches of precipitation at the intersection, in days.
- NDaysPreciponeinch—number of days with at least 1 inch of precipitation at the intersection, in days.
- WeatherStation—name of the weather station used near the intersection.

Weather station data was not always available or complete at the closest station for every year. If one month of data was missing, the averages, and totals, for the year were incomplete. According to the NOAA station summaries, the weather data are approximately 90 to 95 percent accurate for most sites included in the study. The project team used data imputation to determine average temperature for years with missing data. Imputation was based on months in which data were included, nearby land-based stations, and previous years' data. Mean temperatures for the year do not fluctuate greatly and considering these other factors allowed for the project team to achieve a high degree of accuracy.

For sites with missing precipitation data, the project team considered data from nearby land-based stations, as well as previous years' data and monthly totals to impute a yearly total. The yearly precipitation numbers fluctuated more than temperatures, but the perceived degree of accuracy is relatively high based on the factors used to impute the data. The project team imputed roughly 10 percent of the yearly temperature and precipitation data due to missing data.

DATA COLLECTION PROCEDURES FOR HORIZONTAL CURVES IN WASHINGTON

This section discusses strategies for building an enhanced dataset to explore predisposing factors associated with horizontal curve crashes on rural two-lane roads in Washington State. The majority of data elements the project team used for analysis were obtained from the HSIS database and WSDOT Roadside Features Inventory program. Additionally, the project team collected weather data from NOAA for land-based stations near rural horizontal curve locations in Washington State. Table 50 through table 53 provide the specific data variables for roadway, roadside, horizontal curve, and crash categories the project team identified, and whether the data are typically required or desired for a crash-based cross-sectional analysis. It is similar to the tables provided in the previous section on predisposing factors at unsignalized intersections. The summary tables also include the information on whether the variable is required or desired, as well as the data collection method associated with each of the variables.

Table 50. Washington roadway crash data elements of interest.

Data Element	Need	Description/Comment	Collection Location	Data Sources
Location type	Req.	Crash location type	Office	HSIS database
Accident type	Req.	Type of crash that occurred	Office	HSIS database
Vehicle direction	Req.	Direction of the vehicle in the crash related to roadway component	Office	HSIS database
Sequence of events	Des.	Events of crash listed in chronological order	Office	HSIS database
Vehicle type	Des.	Type of vehicle involved in the crash	Office	HSIS database
Roadway surface	Des.	Condition of the road surface where the crash occurred	Office	HSIS database

Table 51. Washington roadside crash data elements of interest.

Data Element	Need	Description/Comment	Collection Location	Data Sources
Route number	Req.	Location/linkage element	Office	HSIS database
Functional class	Des.	Segment roadway classification	Office	HSIS database
Roadway width	Des.	Total roadway width for the roadway segment	Office	HSIS database
Lane width	Des.	Lane descriptor	Office	HSIS database
Number of lanes	Req.	Lane descriptor	Office	HSIS database
Shoulder width	Des.	Shoulder descriptor	Office	HSIS database
Left shoulder type	Des.	Shoulder descriptor	Office	HSIS database
Right shoulder type	Des.	Shoulder descriptor	Office	HSIS database
Left-turn lane presence	Des.	Surface descriptor	Office	HSIS database
Left acceleration length and width	Des.	Surface descriptor	Office	HSIS database
Right acceleration length and width	Des.	Surface descriptor	Office	HSIS database
Speed limit	Des.	Legal speed limit on the curve	Office	HSIS database
Average annual daily traffic	Req.	Yearly volume that was determined	Office	HSIS database
Terrain type	Des.	Surface descriptor	Office	HSIS database
Right-turn lane presence	Des.	Surface descriptor	Office	HSIS database

Data Element	Need	Description/Comment	Collection Location	Data Sources
Horizontal curvature	Des.	Radius of the curve measured along rural two-lane roads	Office	HSIS database
Horizontal curve length	Des.	Length of the horizontal curve	Office	HSIS database
Maximum superelevation	Des.	Maximum superelevation for the horizontal curve	Office	HSIS database
Degree of the curve	Des.	Horizontal curve data	Office	HSIS database
Curve deflection angle	Des.	Horizontal curve deflection angle	Office	HSIS database
Vertical curve length and type	Des.	Length and type of the vertical curve	Office	HSIS database
Area type	Req.	E.g., rural or urban	Office	HSIS database
Vertical grade	Des.	Measured for the grade of the vertical curve along rural two-lane roads	Office	HSIS database

Table 52. Washington road side crash data elements of interest.

Data Element	Need	Description/Comment	Collection Location	Data Sources
Concrete barrier length	Des.	Length of the concrete barrier	Office	Roadside Inventory Program (GPS equipment, field)
Concrete barrier offset	Des.	Distance indicator	Office	ArcGIS™
Percent of concrete barrier presence	Des.	Percentage of concrete barrier presence in the horizontal curve	Office	ArcGIS, Microsoft® Excel
Guardrail length	Des.	Length of Guardrail	Office	Roadside Inventory Program (GPS equipment, field)
Guardrail offset	Des.	Distance indicator	Office	ArcGIS™
Percent of guardrail presence	Des.	Percentage of guardrail presence in the horizontal curve	Office	ArcGIS™, Microsoft® Excel
Tree count	Des.	Number of trees associated with the curve	Office	Roadside Inventory Program (GPS equipment, field)
Tree offset	Des.	Average distance indicator	Office	ArcGIS™
Tree diameter	Des.	Average diameter of the trees in the curve	Office	Roadside Inventory Program (GPS equipment, field)

Data Element	Need	Description/Comment	Collection Location	Data Sources
Fixed object count	Des.	Number of fixed objects associated with the curve	Office	Roadside Inventory Program (GPS equipment, Field)
Fixed object offset	Des.	Average distance indicator	Office	ArcGIS™

Table 53. Washington weather crash data elements of interest.

Data Element	Need	Description/Comment	Collection Location	Data Sources
Mean temperature	Des.	Mean annual temperature of intersection in degrees	Office	NOAA
Number of 90 plus days	Des.	Number of days where high temperature reaches 90 °F	Office	NOAA
Number of 32 less days	Des.	Number of days where low temperature reaches 32 °F	Office	NOAA
Total snow	Des.	Total annual snowfall recorded at intersection in inches	Office	NOAA
Days with precipitation	Des.	Number of days annually receiving one tenth, one half, and 1 inch of rain	Office	NOAA

All crash data, roadway, and roadside data elements collection occurred in an office environment using a number of different data sources. These data sources included the HSIS database, WSDOT's roadside inventory database, ArcGIS^{TM} maps, and NOAA, as shown in table 50. The following sections outline the details on key parts of the data collection process.

Crash Data

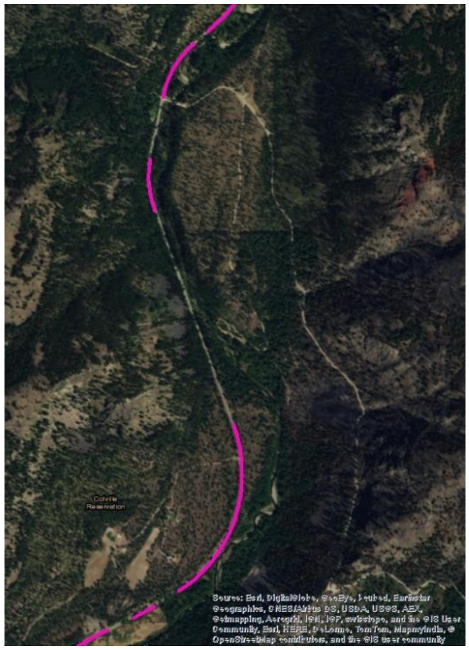
The project team collected crash data for crashes occurring on rural two-lane horizontal curves in Washington State for the years 2008–2012. They obtained the crash data from the HSIS database accident and vehicle file. The project team developed a C++ code, using the route and milepost number to link crash data to roadway data, and using the crash case numbers to link accident files and vehicle files. The project team did not include crashes involving pedestrians and bicyclists. The project team considered single-vehicle and multi-vehicle crashes for analysis, which they defined by the following criteria:

- Single-vehicle crashes—a subset of crashes in which only one vehicle was involved in a roadway or roadside crash.
- Multi-vehicle crashes—a subset of crashes in which two vehicles were involved. The team did not consider any crashes that involve more than two vehicles for this analysis.

Roadway Data

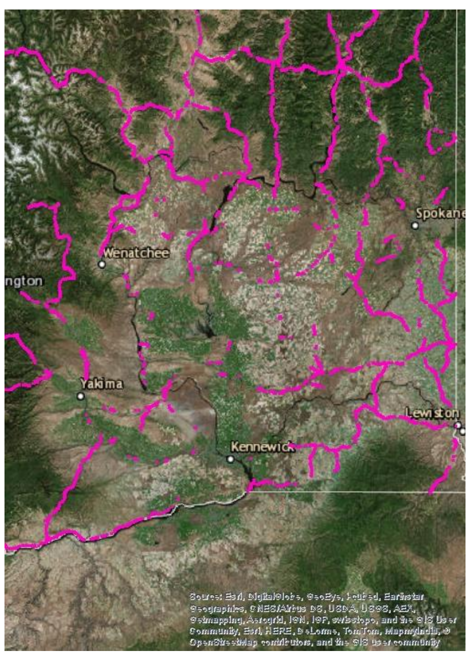
The project team obtained the horizontal and vertical curve data from the curve file in the HSIS database, located by route and milepost. GPS latitude and longitude information were not available for the curve data, but they would need them to link curve characteristics to the roadside features. The project team therefore acquired a linear referencing system route feature class/GIS layer from WSDOT, which was a roadway network, and identified the route field. Figure 14 shows curve information imported into ArcGIS^{TM}. The project team used ArcGIS^{TM} and Google® Earth™ to verify the accuracy of curve data and to filter rural two-lane horizontal curves from the entire dataset of all facility types.

The project team used route and milepost information to link the curve data with the roadway inventory file from the HSIS database. The project team built queries in C++ and SAS to link roadway data elements such as AADT, functional classification, lane width, and shoulder width to the horizontal and vertical curve data. Some curves had multiple AADTs associated with them; the project team assigned higher values of AADT to the curves in those cases.



World Imagery Map source credits: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community. Images with Labels map source credits: Esri, HERE, DeLorme, TomTom, MapmyIndia, ©OpenStreetMap contributors, and the GIS user community.

Figure 14. Screenshot. Illustration of curves imported into ArcGIS^{TM}.



World Imagery Map source credits: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community. Images with Labels map source credits: Esri, HERE, DeLorme, TomTom, MapmyIndia, ©OpenStreetMap contributors, and the GIS user community.

Figure 15. Screenshot. Illustration of curves imported into ArcGIS^{TM}.

Roadside Features Inventory Program

RFIP was a systemwide program of limited scope that WSDOT undertook to provide information on the number, types, and locations of roadside features for the main purpose of safety analysis. Figure 16 demonstrates how WSDOT has handled the roadside data from collection to reporting phase. WSDOT had been scheduled to complete the roadside inventory data collection by late 2015 or early 2016, but the program was stopped. Data are still available for road segments where the roadside was inventoried prior to the program ending.

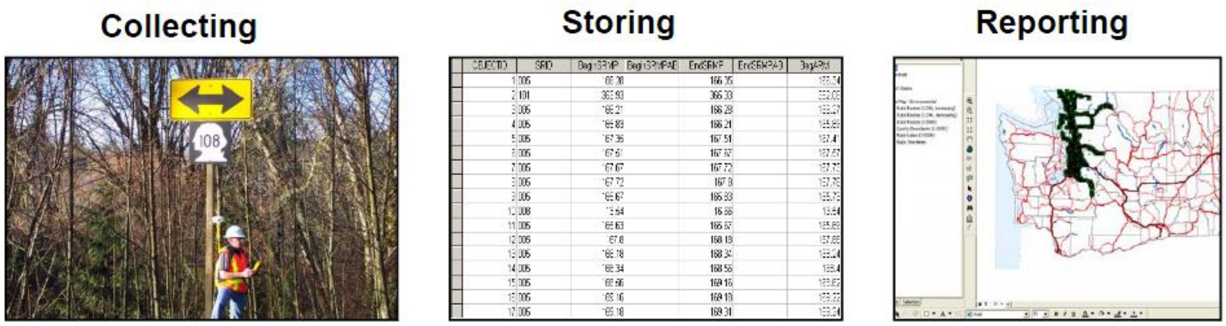
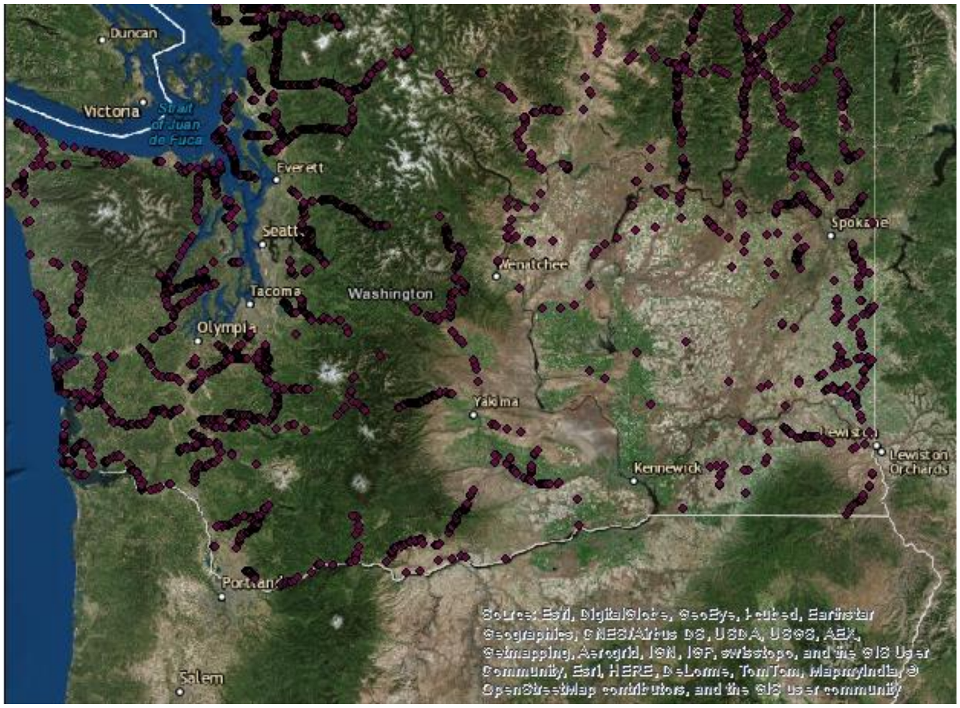


Figure 16. Timeline. Roadside inventory data collection, storage, and reporting procedures.[(41)]

WSDOT used GPS equipment to identify and collect information on the roadside features and attributes. The information on roadside inventory data was provided to the project team in the form of an Excel file. The project team imported the information from the Excel file to ArcGIS™ using the GPS coordinate information provided for the roadside features. Figure 17 shows a screenshot of the roadside features data imported into ArcGIS™. The roadside features this project considered for analysis were the presence and locations of the following features: concrete barrier, guardrail, special-use barrier, tree, and fixed object.



World Imagery Map source credits: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community. Images with Labels map source credits: Esri, HERE, DeLorme, TomTom, MapmyIndia, ©OpenStreetMap contributors, and the GIS user community.

Figure 17. Graphic. Example of roadside features data imported from Microsoft® Excel to ArcGIS™.

Figure 18 shows a "non-traditional" barrier offset calculation performed by the project team using ArcGIS™. The project team overlaid the barrier information and route centerlines from ArcGIS™ onto Google® Earth™'s images to check the accuracy of the alignment. The project team used the "near" distance calculation tool in the ArcGIS™ proximity toolbox to calculate the nearest distance from the barrier (thick solid line, shown in orange) to centerline (dotted line, shown in white). The project team used the calculated nearest distance and lane width to compute the barrier offset value. This tool calculated the nearest distance between the barrier and centerline, and found it to be acceptable for majority of cases. The "near" function could be inaccurate if the barrier is not generally parallel with the centerline.

World Imagery Map source credits: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community. Images with Labels map source credits: Esri, HERE, DeLorme, TomTom, MapmyIndia, ©OpenStreetMap contributors, and the GIS user community.

Figure 18. Graphic. Barrier offset measurements in ArcGIS™.

Weather Data

The project team obtained weather data from the NOAA National Climatic Data Center. They used the climate data online search tool to obtain the past weather and climate data for Washington State. Data from land-based stations included temperature, precipitation, wind, snowfall, relative humidity, and atmospheric pressure. The data were available on a daily, monthly, and yearly basis. The project team collected aggregate yearly data for this project to match the time scale used for crash, roadway, and traffic data at the curve analysis level. Figure 19 shows all weather stations in Washington. There were some inconsistencies in the weather station location information. Sometimes weather stations over time had differing coordinate/elevation information. The project team determined that the majority of the differences were small (within 1 to 3 mi). Hence, this did not affect the inclusion of weather data into the analysis. The following data elements were recorded from each weather station:

- Mean Temperature—mean temperature for the curve over the course of the year, in degrees Fahrenheit.

- Ndays90plus—number of days within the year in which the high temperature at the curve reached 90 °F or greater.

- Ndays32less—number of days within the year in which the low temperature at the curve reached 32 °F or less.

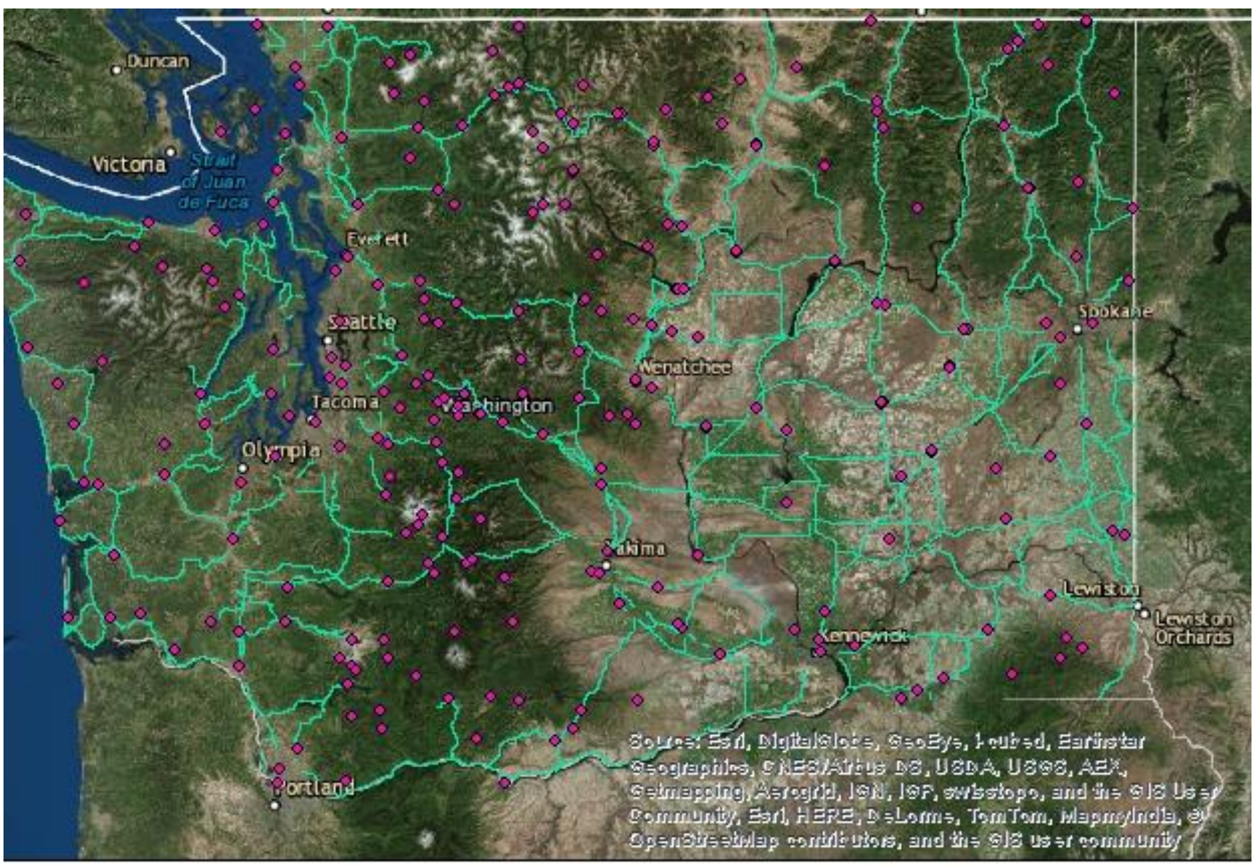
- TotalSnow—total snowfall recorded at the curve for the year, in inches of snowfall.

- Ndayspreciptenthinch—number of days with at least 0.1 of an inch of precipitation at the curve, in days.

- Ndayspreciphalfinch—number of days with at least 0.5 inches of precipitation at the curve, in days.

- Ndayspreciponeinch—number of days with at least 1 inch of precipitation at the curve, in days.

- WeatherStation—name of the weather station used near the curve.



World Imagery Map source credits: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community. Images with Labels map source credits: Esri, HERE, DeLorme, TomTom, MapmyIndia, ©OpenStreetMap contributors, and the GIS user community.

Figure 19. Graphic. Weather stations in Washington.

The project team obtained weather station data for all weather stations in Washington for 5 years (2008–2012). Weather station data were not available or complete for the closest station to each curve every year. The project team considered only those weather stations that had 90 percent or more available weather station data (i.e., less than 10 percent of data missing). The project team used data imputation to determine average temperature, snowfall, and precipitation for included years or stations with some missing data. Imputation was based on the previous years' data and nearby weather stations. In total, the project team had to impute roughly 10 percent or less of the yearly temperature and precipitation data due to missing data.

DATA COLLECTION PROCEDURES FOR HORIZONTAL CURVES IN UTAH

This section discusses strategies for building an enhanced dataset to explore predisposing factors associated with horizontal curve crashes on rural two-lane roads in Utah. As previously noted, UDOT undertook an extensive data collection effort to gather, identify, and process detailed information on all above-ground assets and road characteristics along State routes using LiDAR. The effort appeared to be the first of its kind executed by a State DOT. The database was expected to provide access to detailed road information that is traditionally not available for safety analysis, including grade, superelevation, roadside objects and offsets, roadside slopes, sign and pole presence, and pavement condition, among others. Data were available through UDOT's online data portal, a central clearinghouse of all public UDOT data. The project team relied on roadway inventory developed from LiDAR data and processed and calibrated by one or more data collection contractors, resulting in direct and easy access to a significant number of predisposing roadway variables not typically available in traditional datasets, including cross slope and vertical grade. However, the project team was processing the data in a way to support asset management and the accuracy of certain data elements (e.g., horizontal curvature) was at a level consistent with that need and inconsistent with safety analysis. Additional data processing was necessary.

Curve Estimation Method

The horizontal curvature that UDOT provided via its data portal came in short, broken segments. In its original form, the project team could not use the processed LiDAR data to identify complete curves and their key features (i.e., radii, defection angles, lengths). However, the project team developed a procedure and computer codes to manipulate the data and estimate the horizontal curve properties. The following sections provide a brief description of the horizontal curve data from Utah and details of the algorithm and estimation process. Figure 20 shows an example screenshot of a horizontal curve that is split into many short segments. The cyan arc represents the entire horizontal curve, while the yellow arc is one of the "broken" short segments.

ROUTE	DIRECTION	REGION	MANDLI_STA	MANDLI_END	UDOT_START	UDOT_END_M	CURVE_CLAS	CURVE_DEGR	CURVE_RADI	CURVE_LENG
0006P	P	0	12.364	12.647	12.351	12.633	A	-1.213	-4723.48	1493.491
0006P	N	0	12.822	12.601	12.81	12.589	A	1.14	5025.95	1166.211
0006P	N	0	12.601	12.584	12.589	12.571	A	1.2	4774.65	93.367
0006P	N	0	12.584	12.458	12.571	12.445	A	1.067	5369.8	665.317
0006P	N	0	12.458	12.433	12.445	12.421	A	1.655	3461.98	129.568
0006P	N	0	12.433	12.323	12.421	12.311	A	1.158	4947.82	582.669

Source: FHWA.

Figure 20. Screenshot. Example of "broken" segments of a horizontal curve from the attribute table in ArcGIS™.

The curve property estimation process implemented the following key steps:

- Step 1—Import the curve shape file into GIS software (ArcGIS™) and compute the two-dimensional Cartesian UTM coordinates from longitudes and latitudes in WGS84.

  Add X and Y fields to the Attribute table and use the Calculate Geometry tool in ArcGIS™ to convert the longitude and latitude in degrees into X and Y in meters. The project team would do all the calculations later in using these two-dimensional X–Y coordinates in meters. Figure 21 is a screenshot showing the Calculate Geometry tool for converting GPS coordinates from degrees to meters in ArcGIS™. The entire data file underwent this step of data conversion.

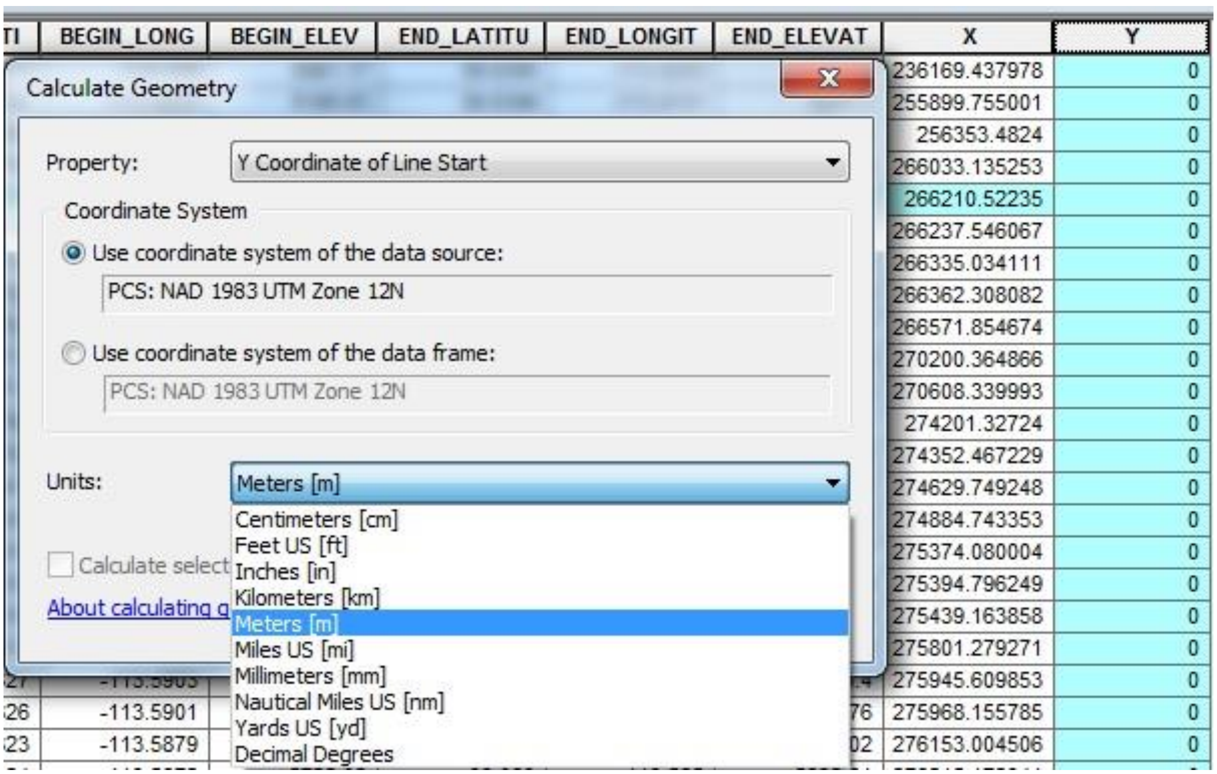
- Step 2—Export the attribute table to a CSV data file and import it into Excel.

  Using the data export tool in ArcGIS™, the project team exported the attribute table into a CSV data file. They then imported the data into an Excel spreadsheet for further processing and calculation.

- Step 3—Combine the short segments and estimate point of curvature (PC) and point of tangent (PT) locations, curve radii, deflection angles, and curve length.

  The direction of all segments on the decreasing milepost direction was reversed. In the data file, the decreasing milepost direction was coded as "N" (the increasing milepost direction was coded as "P"). The project team developed some Visual Basic for Applications (VBA) codes to search all records in the data file and systematically re-code all segments on the decreasing milepost direction. The new direction changed from "N" to "P" while the new beginning milepost received the value of ending milepost and the value of beginning milepost became the new ending milepost.

The project team then sorted the data by route number and the newly reversed mileposts in proper order. The project team developed a computer program in VBA to scan through each route and detect PC and PT locations based on various characteristics of each short segment (e.g., radius is very large for tangent and within a reasonable range for a short segment on curve). They used the mileposts of PC and PT to estimate curve length. The project team calculated the deflection angle from the estimated curve length and the estimated curve radius. After this step, the project team merged the short "broken" curve segments into complete horizontal curves with the estimated values for their key geometric features.



Source: FHWA.

Figure 21. Screenshot. The Calculate Geometry tool in ArcGIS™ for converting coordinates from degrees into meters.

- Step 4—Examine and clean up the data. This step helped screen out abnormal and missing values. The project team examined the data and cleaned up the data using the following criteria:

    - Curves with missing GPS coordinates.
    - Curves with missing traffic volumes.
    - Abnormally long (a few miles) and abnormally short (less than 0.05 mi) curves.
    - Curves with at least one crash coded as intersection-related.

After this process, the dataset went from about 6,500 curves down to 4,416 horizontal curves. It is obvious that, at least in some cases, data for curves were not missing or unreliable completely at random. The developed algorithms had the most challenges with curves having small deflection angles. The algorithm also had problems accurately detecting and estimating curves for winding stretches of roadway where no or very short tangents exist between curves. Finally, data often seemed inaccurate for roads in mountainous terrain.

Given the fact that the Utah curve dataset was new, the project team spent a considerable amount of time and effort to examine, explore and process the data for a safety study. The core element of this process was developing a method of combining "broken," short curve segments into complete horizontal curves and estimating their key geometric characteristics. The project team spent at least 500 person-hours on this effort. This included the time spent on examining the data, testing various ideas developing the algorithms, writing and debugging computer codes, and visually verifying all results in Google® Earth™. The project team adopted VBA programing language in Microsoft® Excel for the job. It is certainly possible to implement the concept in other programing languages as well. While this process took a considerable amount of time and resources, it is also worth noting that this was the first attempt to explore this type of data. With the knowledge and experience gained through this process, similar future efforts will be unlikely to require as much time.

Traffic Volume Layer

Traffic volume information for the Utah roadway network stored in a GIS shapefile was available through UDOT's data portal. The files had AADT for each roadway segment from as far back as 1981 through 2012. After downloading the data files, the project team imported them into ArcGIS™ software for data conversion into a CSV data format.

The data in .CSV format was brought into Excel and merged to each horizontal curve based on route number and milepost. Road segments in the AADT data file (no change in AADT) on rural two-lane roads are often very long and the horizontal curves often completely fall within one of these long segments. In some instances, horizontal curves belong to two different roadway segments with different AADTs. This often takes place where there is an intersection with the horizontal curve. In this case, the project team calculated the AADT for the horizontal curve as the weighted average of the two different AADTs. They eventually dropped horizontal curves with intersections from the final dataset and did not include them in the analysis due to the influence of intersections. Four years of AADT data was merged to horizontal curves (2009–2012) and the 4-year average was computed.

Number of Through Lanes (Through Lane Layer)

Similar to AADT data, the through lane layer and lanes layer came in GIS shape file format. The project team imported the layer into ArcGIS™ and converted the data into an Excel spreadsheet format for merging. Each data record in both data layers had basic route identifier and mileposts. The project team used route number, start milepost and end milepost of each roadway segment to merge number of lanes to each horizontal curve. The project team decided to use number of lanes from both through lane layer and lanes layer. This provided two sources for the same variable that were useful for data verification purposes. The merging process for the entire dataset was performed in Excel using VBA codes that were specifically developed for this purpose. In the end, only two-lane road segments stayed in the dataset.

Passing Lanes and Two-Way Left-Turn Lane (Lanes Layer)

The presence of passing lanes and two-way left-turn lanes was available in the lanes data layer. Each record in this data file had route number, start milepost and end milepost. The project team used these pieces of identification information to merge passing lanes and two-way left-turn lanes to the horizontal curves. The end results showed that a passing lane was present at only one curve and five curves had two-way left-turn lanes. These numbers were too small for any meaningful analysis so the project team dropped these curves from the dataset.

Shoulder Width and Shoulder Material

Shoulder information was available in the shoulder data layer. The project team used route number, start milepost, end milepost and side of shoulder for identifying and merging both left and right shoulders' horizontal curve segments.

Area Type (Urban Code Layer)

Area type information (urban or rural), as determined by existing urban boundaries, was available in urban code layer. The project team matched route number, and mileposts in this data file with corresponding variables in the curve file. In the end, only segments that were coded as "rural" remained in the dataset.

Roadway Surface Width (Pavement Condition Layer)

Lane width was not available in the database. An alternative option was to estimate the road surface width from surface area available in the pavement condition data layer. If it was possible to achieve a reasonably accurate estimation of surface width, it would have been easy to calculate lane width from surface width, number of lanes, and shoulder widths on both sides of the road. However, further examination of the data showed that surface area available in the pavement condition file was aggregated for a very long section of roadway (several miles). Even with both surface area and number of lanes known for the entire section, the surface width estimate was only an average number for a very long stretch of roadway in which both lane width and shoulder width changed. Therefore, the estimated lane width was only the average value over a section of roadway that is several miles long. The project team determined that the average lane width was not meaningful at this level of aggregation. The average surface width was calculated and merged to each horizontal curve segment based on the route number, start and end mileposts. However, lane width could be reliably calculated and used for this study.

Roadside Barrier Length and Offset (Barrier Layer)

Statewide roadside barrier information was available in the barrier data layer. Because the database was built for asset management purposes, a wide variety of information related to barriers was provided. In the safety research context, barrier length, barrier offset, and barrier type could be considered some of the key elements. For this study, the focus was on barrier presence, length, and offset.

The barrier location and length were easily calculated from start and end mileposts of each barrier that were readily available in the data file. Barrier offset, however, was not available and the project team could not easily calculate it. The project team explored a number of options; among them, was using ArcGIS™ to calculate the nearest distance from the barrier to the centerline of each route. Although, this method produced reasonably accurate results for a large number of locations, it was imprecise for many others due to an inaccurate centerline alignment in GIS. Because of the relatively small distance between the barrier and the centerline, both longitudinal positions of the centerline and barrier needed to be accurate, at least relative to each other. However, in many instances, this was not the case. Figure 22 shows an example where this barrier offset estimation method did not give an accurate and reliable result. From figure 22, it is obvious that the barrier locations on both sides of the roads (the blue lines) are quite accurate and that they align reasonably well with the actual barriers in Google® Earth™. However, the roadway centerline from GIS is not consistent with the centerline from Google® Earth™. In such situations, the calculated distance from the centerline to the barrier is not accurate and the overall results of this method are not reliable and should not be used. This may change in the future if the data quality is improved and the longitudinal position of roadway centerline in GIS is found to be consistently accurate.

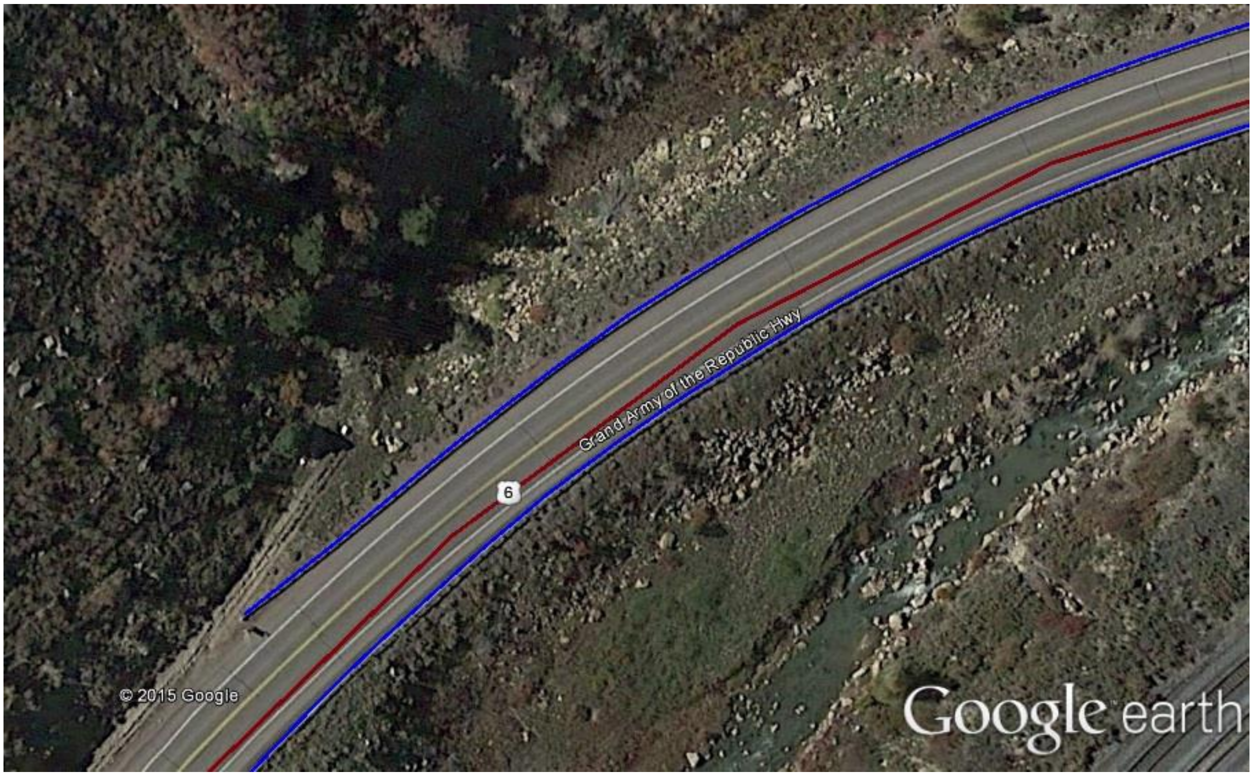


Figure 22. Graphic. An example of inaccurate longitudinal positions of centerline and barriers from Google® Earth™.[42]

The project team randomly checked numerous roadside barrier locations and found that in all those locations that were checked, the barriers were located in close proximity to the paved shoulders. Therefore, shoulder width could be considered as an alternative for barrier offset. Figure 23 shows an example of a curve with barriers located at the right and left edges of a paved shoulder. Barrier length was calculated and merged to the curve segments based on route number, start milepost, and end mileposts.

©VHB.

Figure 23. Photograph. An example of roadside barriers at the edges of paved shoulders.

Crash Data

The project team obtained crash data files separately because they were not available through UDOT's data portal. Typically, crashes are located and merged to road segments using location identifiers such as route number and milepost. Although both the curve file and the crash data file had these identifiers, UDOT tagged all crashes in recent years with GPS coordinates. The project team decided to develop a new method for merging crashes based on GPS information.

The underlying concept of this new method is that a crash is located within a given road segment if the distance from that crash location to either end is smaller than the distance between two ends of such road segment. Information on the number of vehicles involved was used to separate and categorize single-vehicle and multi-vehicle crashes. Two specific types of crashes were also identified and merged: multi-vehicle, opposite direction, no maneuver crashes; and single-vehicle, control loss/road edge departure crashes (the high-priority crash types identified earlier in this report). In the end, the project team identified and counted 5 years of crashes (2009–2013) in five crash categories, and merged them to each horizontal curve segment:

- All crashes—crashes of all types and severity levels.
- Single-vehicle crashes—all crashes with only one vehicle involved.
- Multi-vehicle crashes—all crashes with at least two vehicles involved.
- Scen_22 crashes—multi-vehicle, opposite direction, no maneuver crashes.
- Scen_48 crashes—single-vehicle, control loss/road edge departure crashes.

Weather Data

The project team collected weather data for the State of Utah from the NOAA land-based stations. Monthly data for all weather stations in Utah from 2009 to 2014 were downloaded.

However, 2014 data was incomplete and the project team decided to drop it from the dataset. The remaining 5 full years of weather data included the following data elements:

- DP01—Number of days with at least 0.1 inch of precipitation.
- DP05—Number of days with at least 0.5 inch of precipitation.
- DT32—Number of days with minimum temperature below 32 °F.
- DT90—Number of days with minimum temperature above 90 °F.
- Mntm—Monthly mean temperature (0.1 degree).
- Tsnw—Total snow fall (in 0.1 inch).

The dataset had weather data for more than 150 weather stations located throughout the State of Utah. However, a significant number of weather stations had missing data points. The project team decided that a weather station must have at least 54 months of data (out of 60 months for 5 years) to remain in the dataset. The project team adopted this 54-month cutting point based on the idea that no more than 10 percent of data points were missing (consistent with the other studies in this report). This cut the number of available weather stations from more than 150 down to 86 stations. The project team adopted a data imputation procedure for filling in the missing values. The data imputation was based on data from a nearby weather station for the same month. If data from the nearest weather station was also missing for that month, the project team estimated the missing value and imputed it using data from the same month of the previous year. The project team performed the data imputation in Excel using an application written in VBA they developed specifically for this task.

The project team then merged data from each weather station to each horizontal curve segment based on the distance between them. This was based on the assumption that weather conditions at each horizontal curve were the same as its nearest weather station. It is worth noting that this method did not take the effects of terrain on weather patterns into account. Five years of data (2009–2013) were merged and aggregated into 5-year average values.

Visual Screening of Data in Google® Earth™

The project team took two extra steps to visually check and verify data associated with all 4,416 curves in Google® Earth™.

- Step 1—Mark PC and PT locations in Google® Earth™. In this step, the project team used GPS coordinates of PC and PT to create place markers for Google® Earth™ using Keyhole Markup Language (KML). PC and PT markers were coded with different colors for easy identification (red for PC and green for PT). Both PC and PT markers were attached with curve identifiers and key curve features. This made these key pieces of information readily available for verification when the curve was checked in Google® Earth™ (discussed in Step 2). Figure 24 is an example of PC and PT markers in Google® Earth™. The PC's GPS-based location is marked with the red place marker and the green one represents PT. Both PC and PT markers have almost identical information (except milepost) attached to them.

- Step 2—Check and verify all curves in Google® Earth™. The KML files with place markers for PC and PT locations of all curves were imported into Google® Earth™. The project team then scanned through each curve and visually checked to verify the consistency between the key curve features (which was based on mileposts and other measurements, attached to both PC and PT markers) and what the GPS-based locations of PC and PT appear to be. The project team occasionally used a distance measuring tool for verifying the curve length. Figure 24 is an example of a curve where its basic features appear to be accurate. The information attached to both PC and PT shows that this curve has a deflection angle of 31.01 degrees and is 0.499 mi long. Judging the curve in Google® Earth™, these numbers appear to be reasonable. A distance of 0.5 mi resulted from a quick measurement in Google® Earth™ (along the curve, from PC to PT). Therefore, all pieces of information associated with this curve were consistent and the curve was tagged in the data file for analysis.

Figure 24. Graphic. An example of curve with accurate information from Google® Earth™.(43)

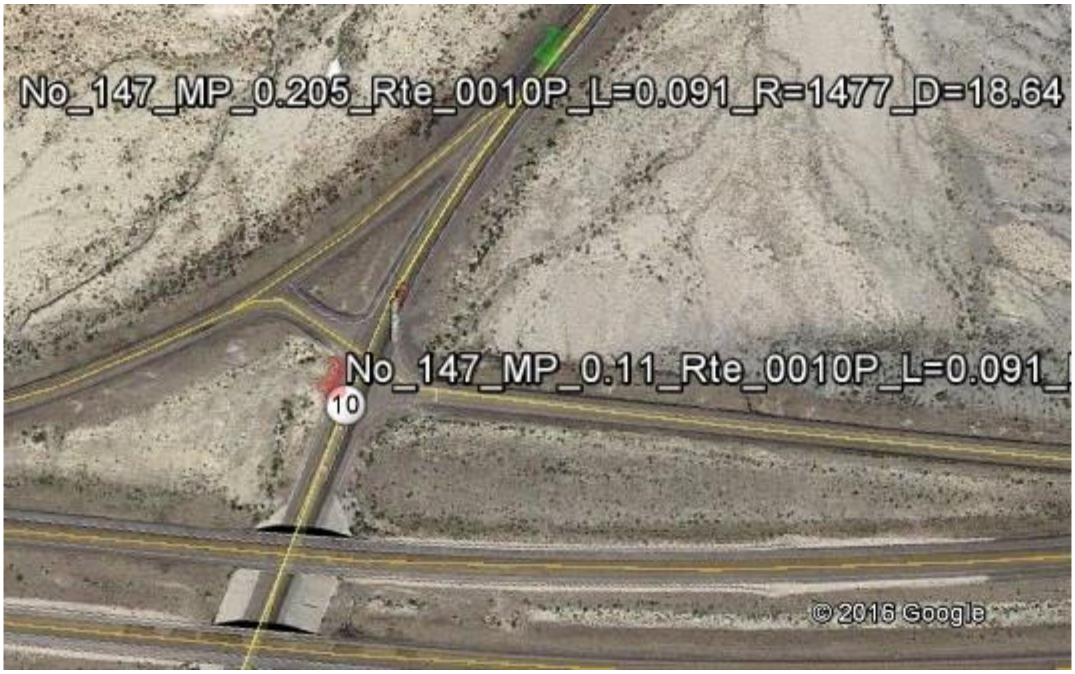
Figure 25 is an example of a curve with inconsistent information. The marker labels indicate that the curve is 0.059 mi long. However, the PC and PT are located at almost the same location. The project team eliminated this curve, together with all other curves with similarly inconsistent information, from the dataset.

It is worth noting that in this case, the GPS locations of PC and PT appear to be inaccurate. During the data screening process, the inaccurate GPS-based locations appear to be more frequent in mountainous areas. Given the terrain of Utah's mountains, poor GPS signal reception could have been the reasons for these inaccuracies in recorded GPS coordinates. The project team recommends caution when using this GPS part of the dataset for any purpose that requires good location accuracy.

©2016 Google® Earth™.

Figure 25. Graphic. An example of discrepancy between curve length and GPS coordinates from Google® Earth™.(44)

Through this process, the project team also screened the data to identify horizontal curves that were located at or near one or more intersections. If they found a curve to be at or near an intersection with another paved road, they tagged it for removal from the dataset. This ensured crashes within those curves were not affected by the intersections in any way. Figure 26 is an example of a horizontal curve located at an intersection. The traffic within this curve was certainly affected by the intersection and interchange. However, there was no indication of the intersection from the data itself. The intersection was only identified visually in Google® Earth™. This curve was eventually removed from the final dataset.



©2015 Google® Earth™.

Figure 26. Graphic. An example of curve at or near intersection from Google® Earth™.[(45)]

This data verification process in Google® Earth™ also helped collect information on winter closure. With the "Roads" layer activated, Google® Earth™ provided sections of roadway that are closed in the winter months. Figure 27 is an example of winter closure information in Google® Earth™. During this data screening process, if a curve was found to be within a section of roadway with the "closed winters" label, it was tagged with an indicator variable.

©2015 Google® Earth™.

Figure 27. Graphic. Example of winter closure information in Google® Earth™.[46]

This Google® Earth™–based data screening process resulted in a much smaller but cleaner dataset. The final dataset had 1,755 horizontal curve segments remaining.

CHAPTER 9. ASSESSING BENEFITS AND TRADE-OFFS OF ENHANCED DATA IN SRSM

As noted in earlier chapters of this report, three studies of enhanced datasets were executed, each assessing the potential of a specific enhanced and robust dataset for increasing what can be learned about the influence of various predisposing factors on safety. These studies continued to build on the marriage of the conceptual models and data concepts presented in chapter 6, with a specific focus on datasets built to maximize available information on predisposing factors. The project team executed this within the constraints of the project's schedule and budget.

BENEFITS AND TRADE-OFFS OF ENHANCED DATA IN STATISTICAL ROAD SAFETY MODELING

The three studies focus on using enhanced and robust data sources to try to improve estimates that quantify the relationships between the expected number of crashes at a location during some defined time period and the predisposing roadway, traffic, and weather factors at that location. Based on the conceptual crash models, the presence of one or more specific predisposing factors (e.g., rainy weather) does not in itself cause a crash. It does however, by its presence, have some level of influence whether or not the driving task will be carried out successfully (or unsuccessfully) driver by driver. Therefore, the project team expects that predisposing factors are associated with the expected number of crashes at an aggregate level. Count regression models are commonly used to search for these associations (see Lord and Mannering for a comprehensive review of methodologies and challenges).(37) The use of regression models to estimate if, how, and by how much a change in one or more predisposing factors affects the expected number of crashes remains a safety research area of great interest and some controversy (e.g., Hauer, Elvik).(33,38) This group of studies is based on the following set of ideas:

- Refined crash type definitions combined with more complete datasets that lower the probability of omitted variable biasing effects on regression model estimates will increase the repeatability of safety findings developed using cross-sectional regression models.

- Increasing the repeatability of safety findings developed using cross-sectional regression models will lead to uncovering "cause–effect" relationships between the expected number of crashes at a location during some defined time period and the predisposing roadway, traffic, and weather factors at that location.

Analysis Methodology

The project team used multivariate regression to estimate statistical relationships between specific crash frequencies and a set of predictor variables. In this case, expected crash frequency, as defined by typically established crash types (i.e., all crashes, single-vehicle crashes, and multi-vehicle crashes) as well as the critical Volpe pre-crash scenarios, was the dependent variable of interest. Predictor variables included predisposing factors available from typical State agency databases (i.e., traffic volume and facility-level identifying characteristics), more advanced State agency databases (i.e., horizontal and vertical curvature), and from supplemental (i.e., "non-traditional") databases. The project team estimated coefficients during the modeling process for each of the predictor variables. The coefficients represented the expected change in the dependent variable (expected crash frequency) due to changes in the predictor variable, all else being equal.

When developing crash prediction models, the current state-of-the-practice is to assume a log–linear relationship between crash frequency and site characteristics. The project team applied generalized linear modeling techniques to develop the models, and specified a log–linear relationship using a negative binomial error structure. The negative binomial error structure has gained recognition as the appropriate method for crash counts than the normal distribution that is assumed in conventional regression modeling. The negative binomial error structure also has advantages over the Poisson distribution in that it allows for dispersion (commonly overdispersion) that is often present in crash data. The appropriate model specification was determined after exploratory data analysis for intersections and horizontal curves on rural two-lane highways.

The project team employed the following protocol to develop the multivariate models:

- Step 1—Identify the base models with traffic volume only. This was important for determining the appropriate functional form for traffic volume.

- Step 2—Explore predictor variables typically available for State agencies in terms of crash, roadway, and traffic data. These data did not include horizontal or vertical alignment data, as most agencies do not have these data for their roadway network.

- Step 3—Explore predictor variables available for State agencies with a more advanced inventory of their roadway network. This included data available from roadway inventory files, from the Utah LiDAR data collection, and from GIS Layers and HSIS data available for States included in the analyses.

- Step 4—Explore predictor variables available from State agencies and from all supplemental data sources for intersections and horizontal curves.

- Step 5—Identify models that include all available predictors, regardless of statistical significance and correlations among predictor variables (with the exception of perfect multicollinearity).

- Step 6—Identify models that include statistically significant and marginally significant variables, while considering correlations among predictor variables. The project team relaxed specific criteria for statistical significance at various levels throughout the analysis due to sample size concerns. Multiple confidence levels (e.g., 80, 90, 95 percent) are identified in each set of model estimation results. These levels vary by specific study, driven primarily by different sample sizes and observations of initial estimation runs. The assumptions the project team used for the specific critical crash scenarios are characterized in the presentation of the findings.

- Step 7—Compare general findings, parameter estimates, and other model properties resulting from estimation using a traditional dataset, an advanced dataset, and an enhanced dataset built by combining advanced and supplemental sources.

The first step of the modeling process was to identify the proper functional form for the exposure term in the crash prediction models. For SCP and multi-vehicle crashes at unsignalized intersections, models were developed using major and minor AADT, or total entering volume (the sum of major and minor AADT), as the only predictor variable(s). Figure 28 and figure 29 give the general form of this model, but the project team investigated various model forms to determine the most appropriate functional form. The general forms of the models for horizontal curves are presented in figure 30 and figure 31. The decision on which form to use was based on an evaluation of parameter estimates and cumulative residual (CURE) plots.

$$\frac{Crashes}{year} = AADT_{maj}^{\beta_1} * AADT_{min}^{\beta_2} * e^{\alpha}$$

Figure 28. Equation. Base model for intersections with major and minor route AADT separate.

$$\frac{Crashes}{year} = TotalEntering^{\beta_3} * e^{\alpha}$$

Figure 29. Equation. Base model for intersections with total entering volume.

$$\frac{Crashes}{year} = L * AADT^{\beta_4} * e^{\alpha}$$

Figure 30. Equation. Base model for curves with segment length as offset variable.

$$\frac{Crashes}{year} = AADT^{\beta_5} * L^{\beta_6} * e^{\alpha}$$

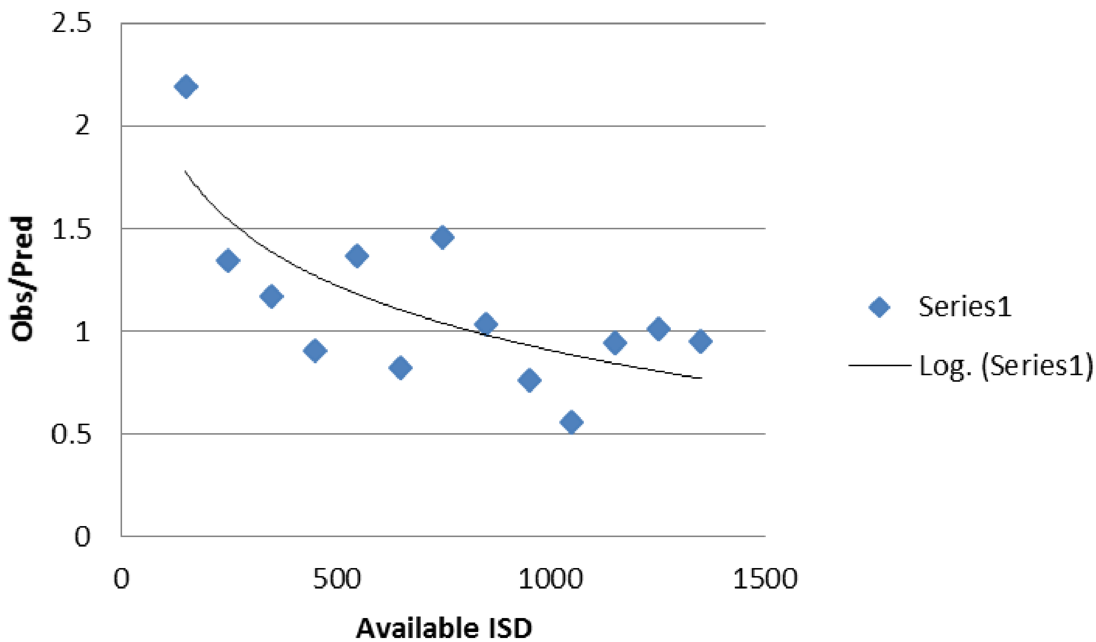
Figure 31. Equation. Base model for curves with segment length as predicted.

Where:

α, β_1–β_6 = parameters estimated in the model calibration process.
$AADT_{maj}$ = annual average daily traffic on the major approach.
$AADT_{min}$ = annual average daily traffic on the minor approach.
$TotalEntering$ = total entering volume per day.
$AADT$ = annual average daily traffic for the horizontal curve.
L = horizontal curve length.

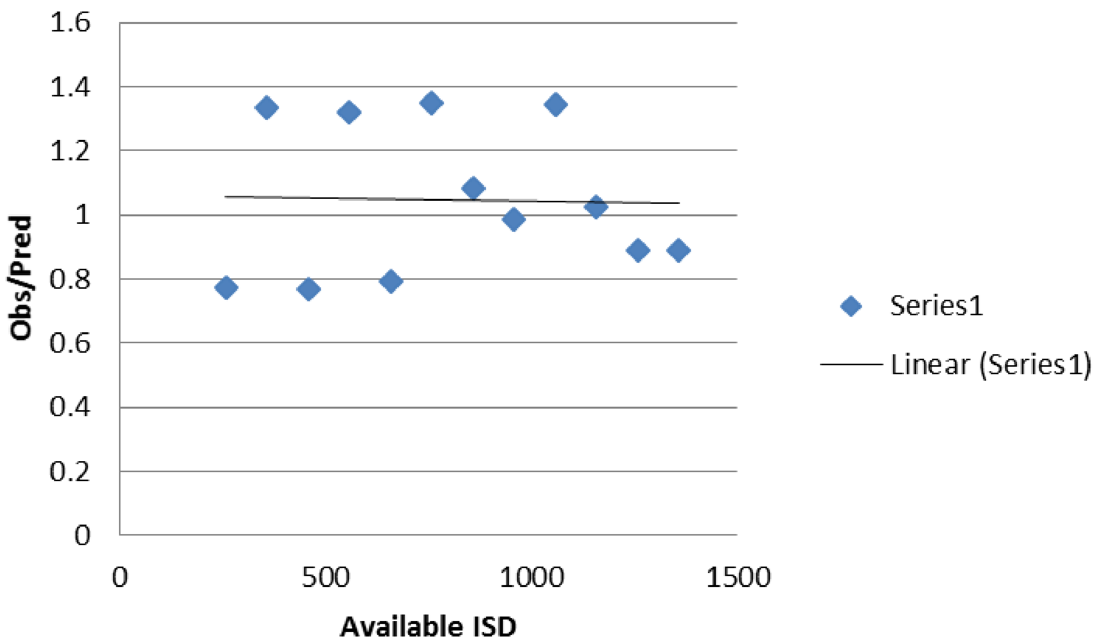
The CURE plot shows the cumulative residuals (difference between observed and predicted values for each observation) plotted in increasing order for the predictor variable of interest. The CURE plot shows how well the model fits the data with respect to the predictor variable of interest. In this study, the project team plotted the cumulative residuals along with the upper and lower 95-percent confidence limits. The CURE plot should oscillate around zero and remain within the confidence limits. Horizontal curve length is included in figure 30 as an offset variable. In this case, the coefficient for curve length is forced to be 1.00. This implies that for two horizontal curves with the same traffic volume, the curve that is twice as long will predict twice as many crashes. In figure 31, horizontal curve length is not constrained to have a parameter of 1.00. This constraint was tested in the analyses' results presented in chapter 8.

For models only considering statistically significant or marginally significant predisposing factors, the project team used variable introduction exploratory data analysis to identify variables that have a relationship with the outcome, and what functional form that relationship takes. In this case, the base model (which includes only AADT, total entering volume, and segment length) served to predict crashes and the predicted crashes were compared to observed crashes across predictor variables to determine if a relationship exists. Figure 32 provides a conceptual example of the ratio of observed to predicted crashes versus ISD. It is clear that crashes are underpredicted for intersections with short available ISD and overpredicted for intersections with high values of available ISD. The relationship is clear, and appears to be nonlinear. This leads to the conclusion that available ISD has a relationship with crash frequency and the predisposing factor should be considered in the model in a nonlinear form. Figure 33 presents a conceptual example for which no identifiable relationship exists between the observed and predicted value ratio and available ISD. In this case, the predisposing factor has no relationship with crash frequency. The project team used this methodology to identify the presence and functional form of factors in crash prediction models.



Source: FHWA.

Figure 32. Chart. Conceptual example of a relationship between ratio of observed/predicted crashes to available intersection sight distance with a clear relationship.



Source: FHWA.

Figure 33. Chart. Conceptual example of a relationship between ratio of observed/predicted crashes to available intersection sight distance with no relationship.

Methodological Considerations of Datasets

While beginning to explore the sensitivity of parameter estimation and other model properties under the three data scenarios—traditional, traditional and advanced data, and all data including the enhanced data collected—there are several estimation considerations. Crash prediction models are regression models, and therefore the results are subject to the data used to develop them. Drawing on concepts from applied econometric modeling, the subject of using alternative data sources and capturing more data is an issue of irrelevant variables, omitted variable bias, multicollinearity, and endogeneity. As demonstrated by Himes et al. for speed modeling, four perspective modeling scenarios and the likely estimation outcomes should be considered.[(47)] Table 54 briefly summarizes the modeling scenarios and likely estimation outcomes in terms of crash prediction modeling.

Table 54. Crash modeling scenarios and possible outcomes.

Modeling Scenario	Likely Outcome	Implication
A predisposing factor is included in the model specification, but is irrelevant to predicting safety.	The parameter estimates are unbiased, but the standard errors increase. There is a higher level of uncertainty in model predictions.	There is no additional benefit to collecting information about the predisposing factor. It does not add any value.
A predisposing factor is excluded from the model, but is relevant to predicting safety.	The model parameters are biased. The level of bias is proportional to the magnitude of correlation between the predisposing factor and the variables included in the model. The influence of other factors on safety is likely under- or over-estimated. Additionally, this systematic error may lead to the false conclusion that an NB model will provide unbiased results.	Excluding relevant predictors may lead to biased interpretations in multiple ways. Every effort should be made to ensure that systematic errors are not present. This includes collecting as many as possible relevant predictors.
A predisposing factor is included in the model, but is correlated with other predisposing factors in the model.	In the extreme case of perfect multicollinearity, the parameters cannot be estimated. In more typical cases, the individual parameter variances are inflated (inefficient) but unbiased. This is not a problem, particularly in cases where the modeler is interested in prediction.	Including relevant predictors may lead to insignificant variables in the model.
A predisposing factor is included in the model, but is not independent of the model disturbance.	The parameter estimates are biased because the variable violates the exogeneity assumption of the estimator. The influence of all variables in the model on crash frequency is likely under- or over-estimated.	The exogeneity of predictors must be considered, especially for safety related treatments (e.g., intersection turn lanes).
A predisposing factor is included in the model, but the sample cannot support additional variables.	The model will suffer from data sparseness. Indicator variables will have bins with no observations and large gaps will exist for continuous variables for which parameter estimates will have no meaning. For these values, no conclusions can be drawn.	For each indicator variable added to the model, the number of observations in each bin decreases. For continuous variables, the range of data must be fully supported for each dimension that is created by the additional variable.

The scenarios and outcomes in table 54 show that omitted predisposing factors may result in estimates that are biased in multiple ways. Model parameters can be biased due to the exclusion of a relevant predictor, and that bias is proportional to the magnitude of correlation between the omitted variable and the included predictors. This is shown in a similar study by Mitra and Washington, who examined the omission of geometric, traffic regulatory information, weather, sun glare, proximity to drinking establishments, and proximity to schools in intersection crash models.[(48)] The authors found that exclusion of relevant variables overstates the effect of minor road AADT by as much as 40 percent and major road AADT by 14 percent.

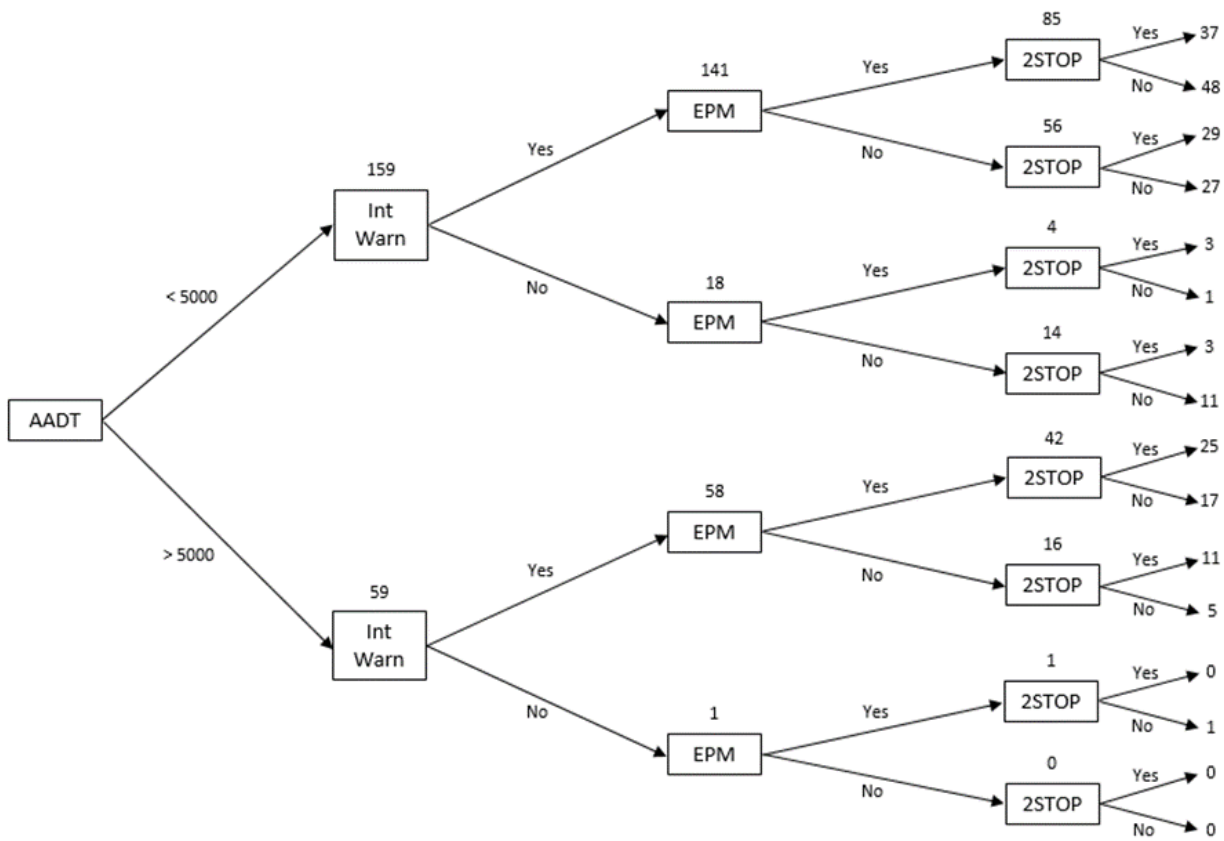
Hauer notes that, in every model, there are some omitted variables, or variables that serve as proxies for other variables.[(49)] For example, data on affluence-related predisposing factors (i.e., income, vehicle traits, education, etc.) are typically not collected, but other predisposing factors collected may serve to capture some of these traits (i.e., spatial factors, driver age). As more relevant data are collected, the surrogate effect of these other traits may diminish. Therefore, the user of the crash prediction model (or SPF) should understand the relationships between correlated predisposing factors and what level of detail is "good enough." It may be more labor intensive and more expensive to collect very detailed data; however, in certain cases easier-to-collect data may suffice.

Including correlated predictor variables may contribute to increased standard errors, but reduce omitted variable bias. When predisposing factors are highly correlated, the model coefficient for any one predictor variable depends on which other predictor variables are included and excluded. Thus, as noted by Kutner et al., the model estimate does reflect the effect of the predictor variable on the response variable but only a marginal or partial effect, given whatever other correlated predictor variables are in the model.[(50)] The authors show that a model that includes highly correlated predictor variables will result in unbiased estimates of the mean outcome; however, the parameter estimates for predictor variables will have high standard errors because the parameter estimates will vary greatly from one sample to the next. Therefore, simple interpretation of the coefficients may be unwarranted, particularly because isolated effects are impractical. Simple interpretation requires that all other factors remain constant; this may be unreasonable if other factors are highly correlated with the factor of interest and will therefore change as the factor of interest changes. The rule of thumb for what defines a highly correlated variable varies, but researchers should be aware of potential issues with collinearity.

Berk et al. show that systematic errors, such as omitted variables or incorrect functional forms, will create excess variation around the fitted values, which may appear to be overdispersion.[(51)] The authors note that it is unlikely one can distinguish whether excess variation around the fitted values is due to misspecification or true overdispersion. They further note that in addition to the negative binomial, there are other distributions that can be used, but they can lead to different empirical results that may be just as arbitrary. However, if the estimated Poisson regression coefficients are biased and inconsistent, the regression coefficients from a negative binomial model with systematic error will continue to be biased and inconsistent.

The violation of exogeneity is presented in table 54. While this can be of concern for estimating the safety effectiveness of predisposing factors, this study will not discuss it in great detail. Other research has shown that cross-sectional models, including models for which the left-hand-side variable (e.g., crash frequency) influences the value for some right-hand-side variables (e.g., presence of lighting), should be subject to increased scrutiny. If crash frequency or severity is an influencing factor for a treatment to be used, then a cross-sectional model will not result in an unbiased estimate of the treatment's effectiveness. For example, if an agency uses a treatment at sites with twice the crash frequency of sites without the treatment (all other factors aside), a resulting 20-percent reduction in crash frequency will still appear to be an increase in crash frequency from the cross-sectional model.

When developing crash prediction models, it is important to consider how many variables can reasonably be included in the model given the number of crashes included in the sample. Figure 34 presents the sample size for all SCP crashes in the dataset, distributed by factor in the model. The counts at each division represent the number of crashes in the dataset for that particular bin. For example, intersections with greater than 5,000 vehicles per day, an advance intersection warning sign, no edge line pavement markings, and the presence of single stop sign have a total of 5 SCP crashes. Other bins have no sample size. For these bins with few observations, it is nearly impossible to establish an odds ratio. At this point, the results are meaningless for several scenarios.



Source: FHWA.

Figure 34. Graphic. Disaggregation of crashes by additional variable.

Additionally, sparse data are a consideration for continuous variables, as shown in figure 35. For example, when major route AADT and minor route AADT are included in the crash prediction model for SCP crashes, the bounds of applicability are typically presented. In this case, the major route AADT is appropriate for 267 to 9,700 vehicles per day, and the minor route AADT is appropriate for 60 to 3,700 vehicles per day. However, when the sample size of crashes is presented together, it becomes readily apparent for where the model is appropriate. It is difficult to make interpretations for any volume of minor route AADT when the major route AADT is fewer than 1,000 vehicles per day. Additionally, there is very little data to support interpretations when the major route AADT is above 5,000 vehicles per day and the minor route AADT is between 1,250 and 3,250 vehicles per day. Researchers should use caution when making interpretations outside these bounds.

Sum of rightangle	Column Labels													
Row Labels	0-249	250-499	500-749	750-999	1000-1249	1250-1499	1500-1749	1750-1999	2000-2249	2250-2499	3000-3249	3250-3499	3500-3749	Grand Total
0-499	0	0												0
500-999	0	0	1											1
1000-1499	2	2	1	3	7									15
1500-1999	0	0	6	4	7	3	16							36
2000-2499		1	1	7	7	2	4	3						25
2500-2999	0	0	0	1	5	0	6	1						13
3000-3499	5	0	0	3		4	0	0		4				16
3500-3999	0	0	4	1		2	7	3	1	2				20
4000-4499			0	0	4	0	0	0	0	0			3	7
4500-4999		3	16	0	2	4	0	0	1					26
5000-5499		0	1	2	3	2	0	4	0					12
5500-5999		0	0	2	2	1		0	1					6
6000-6499		4	1	5	0			4						14
6500-6999								0	0			11		11
7000-7499			1	3										4
7500-7999				0	0									0
8000-8499				7	2									9
9000-9499											2			2
9500-9999											1			1
Grand Total	7	10	32	38	39	18	33	15	3	6	3	11	3	218

Source: FHWA.

Figure 35. Screenshot. Disaggregation of crashes by continuous variables in Microsoft® Excel.

In summary, researchers developing crash prediction models should be cognizant of sample size and sample applicability. They should consider a minimum number of crashes (by crash type) for developing models with a specific number of variables in mind. Other considerations include:

- Prevalence of treatment or variable in model. If the treatment is rare or is only implemented at rural sites, a larger sample may be required to meet a minimum number of crashes.

- Variable form. When possible, the number of categorical variables should be kept to a minimum. If a categorical form makes more sense from a safety effectiveness standpoint, it is worth using. However, including a categorical variable further subdivides the data every time.

- Minimum number of outcomes per bin. It is important to have observations with crashes and with no crashes in each bin.

With all this in mind, analysis under this first group of studies would seek to compare general findings, parameter estimates, and other model properties resulting from estimation using a traditional dataset, advanced dataset, and an enhanced dataset built by combining advanced and supplemental sources. The intent is to attempt to characterize the impacts and benefits of using enhanced and robust data sources to try to improve estimates that quantify the relationships between the expected number of crashes at a location during some defined time period and the predisposing roadway, traffic, and weather factors at that location.

The following section summarizes detailed statistical analyses for the three high-priority crash scenarios identified in chapter 5 and presented in more detail in the appendix. Details on data collection and methodologies were described in chapter 6 and chapter 8, and chapter 9 under the general study descriptions of "Benefits and Trade-Offs of Enhanced Data and Refined Crash Type Definitions." The SCP analysis is presented in the unsignalized intersection study section, using data from North Carolina and Ohio. The combination "control loss/no vehicle action" and "road edge departure/ no maneuver" scenario, and the "opposite direction/no maneuver" scenario are presented in the horizontal curve study section, using data from Washington State and Utah. For the intersection study, the project team estimated the separate models for "multi-vehicle" all crashes and SCP crashes. For the horizontal curve study, separate models were estimated for "single-vehicle" and "multi-vehicle" all crashes and critical crash scenarios; a total of four models.

Benefits and Trade-Offs of Enhanced Data and Refined Crash Type Definitions: Unsignalized Intersection Study

The project team collected the data elements for predisposing factors in table 14 at 88 rural four-leg intersections with stop control on the minor approaches. The dataset includes 42 intersections in North Carolina and 46 intersections in Ohio, with 2008 to 2012 crash data. This provided 434 site-years of data since 6 sites in Ohio did not have 2012 crash and AADT data. The final sample consisted of 365 multi-vehicle crashes and 218 SCP crashes in the dataset.

Table 55 presents summary statistics for North Carolina and table 56 presents summary statistics for Ohio. Crash statistics are presented separately for multi-vehicle crashes and SCP crashes. The data show a sharp contrast between Ohio and North Carolina in terms of several predisposing factors. The North Carolina sites had nearly twice as many annual multi-vehicle and SCP crashes as Ohio sites. The data in table 55 show that North Carolina sites had slightly higher major and minor route AADTs, greater intersection skew, narrower shoulders on the mainline, and lower average intersection sight distance. Conversely, Ohio sites had a lower average annual temperature, greater average annual snowfall, and a slightly higher average speed limit on the major route.

Table 55. Summary of North Carolina data collection site characteristics.

Variable	MV Crash	SCP Crash
Maximum crashes	7	4
Total crashes	239	137
Average annual crash frequency	1.14	0.65
Average annual temperature	59.5 °F	59.5 °F
Average days with half inch of precipitation	79.0 days	79.0 days
Average annual snow	3.7 inches	3.7 inches
Average AADT major	3,770 vehicles	3,770 vehicles
Average AADT minor	1,110 vehicles	1,110 vehicles
Average approach angle	71.5 degrees	71.5 degrees
Average speed limit	50.8 mph	50.8 mph
Average lane width	10.5 ft	10.5 ft
Average shoulder width	1.5 ft	1.5 ft
Average ISD	943 ft	943 ft

Table 56. Summary of Ohio data collection site characteristics.

Variable	MV Crash	SCP Crash
Maximum crashes	6	6
Total crashes	127	81
Average annual crash frequency	0.57	0.36
Average annual temperature	50.2 °F	50.2 °F
Average days with half inch of precipitation	81.8 days	81.8 days
Average annual snow	31.2 inches	31.2 inches
Average AADT major	2,570 vehicles	2,570 vehicles
Average AADT minor	970 vehicles	970 vehicles
Average approach angle	79.0 degrees	79.0 degrees
Average speed limit	53.6 mph	53.6 mph
Average lane width	10.7 ft	10.7 ft
Average shoulder width	2.4 ft	2.4 ft
Average ISD	1,188 ft	1,188 ft

The dataset consists of predisposing factors in continuous form and categorical form. Continuous factors are measured on a spectrum and the summary statistics are presented in table 57. Categorical factors consist of indicator variables for the presence or absence of specific characteristics, or as different categories of a single characteristic (e.g., functional classification). In this dataset, categorical factors are unordered, and do not represent a scale.

Table 57 provides descriptive statistics for categorical factors. Since all intersections were on rural two-lane roads, there were no sites that had a median. However, the descriptive statistics are shown in table 58 to present information on all variables that were collected.

Table 57. Descriptive statistics for continuous factors.

Variable	Minimum	Maximum	Mean	Standard Deviation
Multi-vehicle crash count (crash/year)	0	7	0.84	1.26
Straight crossing path crash count (crash/year)	0	6	0.50	0.91
Major route average annual daily traffic (vehicle/day)	267	9,700	3,150	1,886
Minor route average annual daily traffic (vehicle/day)	60	3,704	1,037	728
Posted speed limit (mph)	35	55	52.22	4.94
Access density within quarter mile (number of pts)	0	35	9.16	7.09
Shoulder width (ft)	0	6	1.98	1.24
Median width (ft)	0	0	0	0
Lane width (ft)	9	12	10.61	0.73
Number of driveways within 250 ft on major	0	7	1.31	1.49
Number of driveways within 250 ft on minor	0	5	1.76	1.32
Distance to advance intersection warning sign (ft)	0	1,266	561	381
Distance to advance stop ahead sign (ft)	0	1,300	599	301
Speed reduction from posted to advisory (mph)	0	35	3.57	7.52
Mean temperature (°F)	48.0	61.8	54.67	4.90
Number of days where high reached 90 °F	2	91	33.96	23.47
Number of days where low reached 32 °F	43	137	105.94	29.55
Number of days with at least 0.1 inch of precipitation	48	120	80.48	11.56
Number of days with at least 0.5 inch of precipitation	13	49	29.02	6.46
Number of days with at least 1 inch of precipitation	3	23	10.01	3.26
Total annual snowfall (inches)	0	83	17.93	17.37
Intersection angle (degrees)	27.5	90	75.88	13.90
Intersection sight distance quality	1	3	1.91	0.36
Proportion of entering volume from minor route	0.03	0.50	0.25	0.12
Value of minimum absolute grade at 500 ft (%)	0	3	0.61	0.79
Value of average absolute grade at 500 ft (%)	0	4.9	0.98	1.03

Variable	Minimum	Maximum	Mean	Standard Deviation
Value of maximum absolute grade at 500 ft (%)	0	7.1	1.34	1.40
Minimum left intersection sight distance (ft)	239	1,321	932	376
Average left intersection sight distance (ft)	321	1,321	1,067	273
Maximum left intersection sight distance (ft)	395	1,321	1,201	230
Minimum right intersection sight distance (ft)	200	1,321	949	358
Average right intersection sight distance (ft)	362	1,321	1,068	279
Maximum right intersection sight distance (ft)	370	1,321	1,188	262
Minimum intersection sight distance (ft)	200	1,321	886	385
Average intersection sight distance (ft)	356	1,321	1,068	264
Maximum intersection sight distance (ft)	443	1,321	1,250	185

Table 58. Descriptive statistics for categorical factors.

Variable	Categories	Frequency	Percent
Deficient pavement	Yes	90	20.5
Deficient pavement	No	350	79.5
Stop line presence on minor road approach	Yes	125	28.4
Stop line presence on minor road approach	No	315	71.6
Presence of one stop sign on approach	Yes	320	72.7
Presence of one stop sign on approach	No	120	27.3
Presence of two stop signs on approach	Yes	165	37.5
Presence of two stop signs on approach	No	275	62.5
Presence of reflective post on stop sign	Yes	15	3.4
Presence of reflective post on stop sign	No	425	96.6
Presence of oversized stop sign	Yes	100	22.7
Presence of oversized stop sign	No	340	77.3
Presence of advance int. warning sign	Yes	320	72.7
Presence of advance int. warning sign	No	120	27.3
Presence of double advance int. warning	Yes	10	2.3
Presence of double advance int. warning	No	430	97.7
Presence of stop ahead warning sign	Yes	370	84.1
Presence of stop ahead warning sign	No	70	15.9
Presence of double advance stop ahead	Yes	95	21.6
Presence of double advance stop ahead	No	345	78.4
Presence of edge line extension	Yes	200	45.5
Presence of edge line extension	No	240	54.5
Presence of speed advisory on major	Yes	100	22.7
Presence of speed advisory on major	No	340	77.3

Variable	Categories	Frequency	Percent
Presence of RRPM on major road	Yes	330	75.0
Presence of RRPM on major road	No	110	25.0
Presence of RRPM on minor road	Yes	115	26.1
Presence of RRPM on minor road	No	325	73.9
Presence of intersection lighting	Yes	20	4.5
Presence of intersection lighting	No	420	95.5
Presence of curve on both major approaches	Yes	115	26.1
Presence of curve on both major approaches	No	325	73.9
Presence of right-turn lane on major approach	Yes	5	1.1
Presence of right-turn lane on major approach	No	435	98.9
Presence of left-turn lane on major approach	Yes	3	0.7
Presence of left-turn lane on major approach	No	437	99.3
Presence of right-turn lane on minor approach	Yes	20	4.5
Presence of right-turn lane on minor approach	No	420	95.5
Presence of left-turn lane on minor approach	Yes	5	1.1
Presence of left-turn lane on minor approach	No	435	98.9
Functional classification	Principal arterial	10	2.3
Functional classification	Major arterial	25	5.7
Functional classification	Minor arterial	95	21.6
Functional classification	Major collector	235	53.4
Functional classification	Minor collector	30	6.8
Functional classification	Local	45	10.2
Terrain	Level	215	48.9
Terrain	Rolling	220	50
Terrain	Mountainous	5	1.1

As table 57 and table 58 show, there was a wide variety of predisposing factors considered for four-leg intersections with stop-control on the minor approaches. However, the data collection sites contained very few observations for several factors considered. For these factors, the outcome (i.e., crashes occurred vs no crashes occurred) does not vary for the data element being collected. When an independent variable contains responses of only one type, it provides no information about the odds ratio. Two options exist for models with an independent variable (in this case predisposing factor) in question:

- Remove the independent variable.
- Combine the independent variable with other independent variables, if possible. One example would be whether the presence of a left-turn lane on one minor approach observes no crashes. The variable could be combined with right-turn lane to form a combined variable for the presence of any auxiliary turn lane. This would allow the response to use information for both outcomes.

Including variables with zero sample counts can also provide estimates for independent variables that are quite large, are counterintuitive, or have standard errors that are quite large and are numerically unstable. Agresti notes that a danger with sparse data is that one might not realize that a true estimated effect is infinite and, as a consequence, will report the effects and associated statistical inferences that are invalid and are highly unstable.[52] With sparse data, any small changes in the data or model specification may result in large changes in parameter estimates.

The descriptive statistics show that, while collected, it is inadvisable to consider several predisposing factors in models due to small samples. Frequency presents the number of site-years of available data for each category and, if divided by five, the number of sites that had that factor. Three sites had stop signs with reflective posts, resulting in 15 site-years of data. This results in too few observations to provide meaningful results. Other factors that the project team did not consider in the final models include the following:

- Presence of double advance intersection warning signs.
- Presence of intersection lighting.
- Presence of right-turn lane on the major approach.
- Presence of left-turn lane on the major approach.
- Presence of right-turn lane on the minor approach.
- Presence of left-turn lane on the minor approach.

Additionally, the project team considered neither principal arterial and major arterial functional classifications nor mountainous terrain in isolation. The project team considered these categories in combination with other categories. For example, they combined principal arterial with major arterial and minor arterial to compare all arterials to other roadway functional classes. Mountainous terrain was combined with rolling terrain to create a flat terrain versus non-flat terrain indicator.

Finally, correlation matrices are presented to further show the existing two-way associations between independent predisposing factors. The correlation value provides the degree of linear relationship between two factors, ignoring all other factors, and does not provide causality. A value near +1 or –1 means that a nearly perfect linear relationship exists between the factors and a value near 0 indicates no relationship between the predictor variables.

Model Estimation Results and Discussions

Table 59 and table 60 include preliminary model estimation results for multi-vehicle crashes and SCP crashes, respectively. Six models were developed for each crash type classification:

- Model 1—considering an extensive number of predisposing factors collected from alternative and traditional sources.

- Model 2—considering only statistically significant predisposing factors from alternative and traditional sources.

- Model 3—considering predisposing factors expected to be available in advanced State agency databases.

- Model 4—considering only statistically significant predisposing factors expected to be available in advanced State agency databases.

- Model 5—considering predisposing factors collected from traditional State agency databases.

- Model 6—considering only statistically significant predisposing factors from traditional State agency databases.

Due to the rare nature of multi-vehicle and SCP crashes at rural four-leg intersections, relaxed values of statistical significance were identified for presentation of results. Predisposing factors significant with at least 80-percent confidence were included in Models 2, 4, and 6; however, asterisks are used to differentiate between variables that were significant with 80-percent confidence and variables that were significant with 90-percent confidence.

Variables that were significant in Model 1 were not necessarily the same variables that were significant in Model 2. In most cases, more variables were significant when fewer variables were included in the model. Removing correlated, insignificant predictors provided results that were more efficient (i.e., the standard errors were reduced) and led to more and/or different variables that were significant in the models.

The preliminary results show that many predisposing factors collected from alternative data sources are statistically significant predictors of multi-vehicle crashes and SCP crashes. The project team found several traffic control related variables to be statistically significant, especially safety-related traffic control devices. However, some safety related treatments may suffer from selection bias; e.g., the presence of two stop signs on a minor approach. A significant increase in multi-vehicle crashes and SCP crashes is associated with these signs; however, it is possible that two stop signs are used because of crash history. The model will reflect the estimation association, but will be biased.

Table 59. Negative binomial regression models for multi-vehicle crashes.

Variable	**Factors From Alternative and Traditional Sources**		**Factors In Advance State Agency Databases**		**Factors From Traditional State Agency Databases**	
Model Number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Log-major AADT	2.753**	1.821**	1.064	1.031	3.207**	3.170**
Log-minor AADT	–1.861**	–1.864**	–1.019**	–0.986**	–1.722**	–1.689**
Proportion minor AADT	14.173**	13.893**	9.457**	9.273**	13.864**	13.614**
Access density within quarter mile	–0.008	--	--	--	--	--
Lane width	–0.419**	–0.376**	–0.319**	–0.315**	–0.366**	–0.351**
Shoulder width	–0.034	--	–0.002	--	–0.028	--
Major route driveways (250 ft)	0.011	--	--	--	--	--

Variable	Factors From Alternative and Traditional Sources		Factors In Advance State Agency Databases		Factors From Traditional State Agency Databases	
Model Number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Minor route driveways (250 ft)	0.159**	0.151**	--	--	--	--
Deficient pavement	–0.256	--	--	--	--	--
Stop line presence	0.283**	0.303**	--	--	--	--
Two stop signs present	0.548**	0.572**	--	--	--	--
Large stop sign	–0.088	--	--	--	--	--
Intersection warning on major	0.662	0.966**	--	--	--	--
Distance to intersection warning	–0.0005	–0.0008*	--	--	--	--
Stop ahead on minor	0.736**	0.800**	--	--	--	--
Double stop ahead on minor	–0.430**	–0.311*	--	--	--	--
Distance to stop ahead	–0.0006	–0.0009**	--	--	--	--
Edge line extension	–0.067	--	--	--	--	--
Speed advisory plaque present	0.478	0.661**	--	--	--	--
Speed reduction advised	–0.041*	–0.049**	--	--	--	--
RRPMs on major	–0.153	--	--	--	--	--
RRPMs on minor	0.138	--	--	--	--	--
Mean temperature	0.019	--	--	--	--	--
Number of 90 or greater days	–0.002	--	--	--	--	--
Number of 32 or less days	0.007	--	--	--	--	--
Total snowfall	–0.009	–0.009*	--	--	--	--
Number of tenth-inch precipitation days	0.004	--	--	--	--	--
Number of half-inch precipitation days	–0.024	--	--	--	--	--
Number of 1-inch precipitation days	0.029	--	--	--	--	--

Variable	Factors From Alternative and Traditional Sources		Factors In Advance State Agency Databases		Factors From Traditional State Agency Databases	
Model Number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
North Carolina indicator	0.439	--	0.575**	0.612**	0.412**	0.448**
Intersection angle	–0.010*	–0.013**	–0.009**	–0.008**	--	--
Intersection sight distance quality	–0.456*	–0.473**	--	--	--	--
Presence of curve on both major approaches	–0.156*	--	–0.209*	–0.214*	--	--
Speed limit	0.052**	0.037**	0.025**	0.026**	0.024**	0.024**
Maximum grade at 500 ft (absolute value)	0.186**	0.160**	0.172**	0.184**	--	--
Maximum intersection sight distance left	–0.0001	--	–0.0005*	–0.0006**	--	--
Minimum intersection sight distance left	–0.001*	--	–0.0001	--	--	--
Maximum intersection sight distance right	0.0004	–0.008**	–0.009**	–0.009**	--	--
Minimum intersection sight distance right	–0.004	--	0.001**	0.001**	--	--
Flat terrain indicator	–0.098	--	–0.098	--	--	--
2009 indicator	0.154	0.159	0.143	0.142	0.115	0.115
2010 indicator	0.246	0.315**	0.216	0.216	0.189	0.188
2011 indicator	–0.061	–0.069	–0.101	–0.101	–0.139	–0.140
2012 indicator	0.227	0.123	0.153	0.150	0.101	0.098
Major route arterial indicator	–0.179	--	0.017	--	0.078	--
Interaction maximum ISD right and major route AADT	0.001*	0.001**	0.001**	0.001**	--	--
Constant	–13.345	–3.590	–2.579	–2.603	–15.396	–15.470

Variable	Factors From Alternative and Traditional Sources		Factors In Advance State Agency Databases		Factors From Traditional State Agency Databases	
Model Number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Log-likelihood	–454.32	–459.82	–472.34	–472.58	–484.61	–484.82
Dispersion parameter	0.045	0.109**	0.209**	0.208**	0.351**	0.353**

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
--No data.

Table 60. Negative binomial regression models for straight crossing path crashes.

Variable	Factors From Alternative and Traditional Sources		Factors In Advance State Agency Databases		Factors From Traditional State Agency Databases	
Model Number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Log-major AADT	3.497**	3.120**	1.831**	1.732**	3.593**	3.576**
Log-minor AADT	–2.805**	–2.712**	–1.498**	–1.426**	–2.192**	–2.189**
Proportion minor AADT	20.857**	19.543**	13.087**	12.435**	16.928**	16.832**
Access density within quarter mile	–0.039*	--	--	--	--	--
Lane width	–0.475**	–0.315**	–0.331**	–0.312**	–0.398**	–0.416**
Shoulder width	–0.178**	–0.175**	–0.093	--	–0.085	--
Major route driveways (250 ft)	0.121	--	--	--	--	--
Minor route driveways (250 ft)	0.225**	0.204**	--	--	--	--
Deficient pavement	–0.607*	–0.809**	--	--	--	--
Stop line presence	–0.433*	--	--	--	--	--
Two stop signs present	1.354**	1.145**	--	--	--	--
Large stop sign	–0.382*	--	--	--	--	--
Intersection warning on major	0.352	--	--	--	--	--
Distance to intersection warning	0.0001	--	--	--	--	--
Stop ahead on minor	0.815*	0.448**	--	--	--	--
Double stop ahead on minor	–.765**	–0.721**	--	--	--	--
Distance to stop ahead	–0.0002	--	--	--	--	--
Edge line extension	–0.461	--	--	--	--	--

Variable	Factors From Alternative and Traditional Sources		Factors In Advance State Agency Databases		Factors From Traditional State Agency Databases	
Model Number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Speed advisory plaque present	0.778*	0.576**	--	--	--	--
Speed reduction advised	–0.030	--	--	--	--	--
RRPMs on major	0.075	--	--	--	--	--
RRPMs on minor	0.320	--	--	--	--	--
Mean temperature	0.001	--	--	--	--	--
Number of 90 or greater days	–0.004	--	--	--	--	--
Number of 32 or less days	0.006	0.010**	--	--	--	--
Total snowfall	–0.019*	–0.021**	--	--	--	--
Number of tenth-inch precipitation days	0.003	--	--	--	--	--
Number of half-inch precipitation days	–0.010	--	--	--	--	--
Number of 1-inch precipitation days	0.003	--	--	--	--	--
North Carolina indicator	0.275	--	0.390*	0.614**	0.492**	0.553**
Intersection angle 85 or greater	–0.619**	–0.895**	–0.759**	–0.651**	--	--
Intersection sight distance quality	–1.086**	–1.174**	--	--	--	--
Presence of curve on both major approaches	–0.085	--	–0.213	--	--	--
Speed limit	0.052**	0.042**	0.045**	0.045**	0.046**	0.045**
Maximum grade at 500 ft (absolute value)	0.109	0.116**	0.104*	0.148**	--	--
Maximum intersection sight distance left	0.00002	--	–0.0001	--	--	--
Minimum intersection sight distance left	–.001	--	–0.0004	--	--	--
Maximum intersection sight distance right	–0.0005	–0.0005*	–0.00002	--	--	--

Variable	Factors From Alternative and Traditional Sources		Factors In Advance State Agency Databases		Factors From Traditional State Agency Databases	
Model Number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Minimum intersection sight distance right	–0.008*	–0.008**	–0.008**	–0.009**	--	--
Flat terrain indicator	–0.228	–0.556**	–0.245	--	--	--
2009 indicator	0.064	0.188	0.058	0.048	–0.008	–0.005
2010 indicator	0.232	0.102	–0.023	–0.021	–0.048	–0.047
2011 indicator	–0.141	–0.176	–0.244	–0.267	–0.311	–0.317
2012 indicator	–0.115	–0.010	–0.157	–0.174	–0.208	–0.216
Major route arterial indicator	–0.994**	–0.922**	–0.326*	–0.445**	–0.345*	–0.379**
Interaction minimum ISD right and major route AADT	0.001**	0.001**	0.001**	0.001**	--	--
Constant	–11.237	–9.940	–7.524	–7.896	–17.006	–16.793
Log-likelihood	–324.83	–331.15	–358.64	–361.33	–372.28	–372.82
Dispersion parameter	2.7e-7	3.4e-15	0.388**	0.432**	0.652**	0.653**

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
--No data.

The results also show that parameter estimates for predisposing factors included in the analyses along with only variables that are available in the limited datasets are likely biased. For example, for multi-vehicle crashes, the estimated parameter corresponding to posted speed limit is 0.028 for the model considering only variables from the limited State database category, 0.031 for the model considering variables from the advanced State database category, and 0.057 for the model considering variables from the combination of traditional and non-traditional sources. This indicates that the model considering only variables from the limited State data underestimates the effect of posted speed limit by 51 percent. The presence of lighting has an estimated coefficient of –0.562 in the model estimated using the advanced State database classification and –0.350 in the model estimated with data from the combination of traditional and non-traditional sources, indicating that the effect of intersection lighting is overestimated by 38 percent when certain key variables are not available for the model specification. Most predisposing factors appear to be underestimated when using the more limited datasets, indicating that they are correlated with omitted variables that have an “opposite” safety effect; however, the results are not consistent across all factors.

Comparing the model with only statistically significant predisposing factors drawn from the combined traditional and non-traditional data sources versus the model estimated from the same database with all predisposing factors (regardless of statistical significance), it becomes clear that the estimated effects of some factors are highly dependent on the inclusion of other factors, for both magnitude and direction. For example, there is a correlation between lighting presence and the presence of left-turn lanes on the minor route. When considering the model with all predisposing factors for multi-vehicle crashes (regardless of statistical significance), the effect of lighting differs substantially from the model with only statistically significant predisposing factors. This is because the project team included the presence of a left-turn lane in the model estimate using the combined traditional and non-traditional data sources, and the presence of left-turn lanes has a large effect in this model. However, the difference between the lighting parameter in the fully specified model estimated using the advanced State database category and the model with only statistically significant predictors from the advanced State database category is very small. This is because the effect of minor route left-turn lanes is also estimated to be very small when this dataset is used. Similar issues are present for SCP crashes as well. This shows that differences in model estimation results likely stem from issues and impacts of omitted variable bias, irrelevant variables, and correlation between independent variables.

Changes in the estimated dispersion parameter are consistent with the research by Berk et al.[(51)] The dispersion parameter estimate is largest for the model estimated with traditional State data and is statistically significant, indicating the need for a negative binomial model. For the model estimated using variables available in the advanced State database category, the magnitude of dispersion parameter estimate is reduced by nearly 50 percent, but it is still statistically significant. When estimating the model using the combined traditional and alternative data sources, the dispersion parameter reduces to nearly zero, and is no longer statistically significant. These findings suggest that true dispersion is not present in the data, but could instead be the result of a misspecification in the form of omitted variables. In this case, the resulting parameter estimates remain biased for the models estimated with advanced State data or reduced State data, even when correcting for the apparent dispersion using the negative binomial model. The findings are similar for SCP crashes.

The parameter estimates for the predisposing factors in the model are consistent in direction of effect and magnitude between multi-vehicle crashes and SCP crashes. This is not surprising given that the SCP crashes constitute 60 percent of all multi-vehicle crashes at the selected study sites. However, it was somewhat surprising that a similar number of statistically significant variables were found for comparative models of SCP crashes and multi-vehicle crashes given the smaller sample size. It appears as if a removal of “noise” that comes from looking at one specific crash pattern offsets the reduced sample size. The estimated effects for SCP crashes are generally greater in magnitude than those for multi-vehicle crashes, indicating that the significant factors have a greater impact on that specific crash type than they do for a general combination of crash types.

There are a few factors that are statistically significant for SCP crashes that are not for multi-vehicle crashes and vice versa. The additional factors that are significant for multi-vehicle crashes are generally related to the mainline (e.g., presence of an advance intersection warning sign), which may be related to crash types other than SCP crashes (e.g., rear-end crashes). Additional terrain and intersection sight distance factors are significant for SCP crashes only, providing further insights into what factors are related to SCP crashes, but may be lost in models for all multi-vehicle crashes.

When weighing the results of the models for varying levels of data, and for differing crash types, a trade-off occurs between the cost of additional data and the benefit of having the additional data. In some cases, data exist in other sources, but it is not typical to integrate them with roadway and crash data (e.g., traffic sign data). For existing data from other sources, the additional cost is very small, since these data were collected for another purpose (e.g., sign condition inventory). However, it is necessary to collect some data at the physical site (e.g., standard available intersection sight distance for each approach direction). These data can be very expensive, and very time consuming to collect. In this case, the intended purpose of the crash prediction model must be clear and the impact on safety of the data also must be clear. If the data element has a large impact on safety and is correlated with other variables, it will be highly beneficial to collect. If the variable has little impact on safety and is not related to other variables, there will be very little benefit. Additionally, if the goal is to understand the influence of variables on safety, it is important to consider all potential confounders. If the goal is to use the crash prediction model to predict crash frequency, the inclusion of additional relevant predictors is not as important. In both cases, it is imperative that researchers have a sound theoretical understanding of what factors are associated with the crash type of interest, and how the predisposing factors of interest interact with one another.

To provide an example of the importance of relationships among explanatory variables, table 61 presents models with and without the proportion of AADT on the minor route side by side. The project team removed the interaction term between major route AADT and intersection sight distance for convenience.

Table 61. Illustration of important relationships in safety prediction models.

Variable	Model With PropAADT	Model Without PropAADT	Correlation With PropAADT	Percent Difference
Log-major AADT	4.071**	0.505**	–0.33	88
Log-minor AADT	–2.906**	0.391**	0.60	113
Proportion minor AADT	20.257**	--	1.00	--
Lane width	–0.280**	–0.212	–0.15	24
Shoulder width	–0.201**	–0.175**	0.10	13
Minor route driveways (250 ft)	0.217**	0.189**	–0.03	13
Deficient pavement	–0.927**	–0.575**	–0.25	38
Two stop signs present	1.159**	0.933**	0.13	19
Stop ahead on minor	0.417**	0.381*	0.17	9
Double stop ahead on minor	–0.589**	–0.434**	0.21	26
Speed advisory plaque present	0.526**	0.420**	0.19	20
Number of 32 or less days	0.009**	0.007*	0.06	22
Total snowfall	–0.019**	–0.017**	0.07	11
Intersection angle 85 or greater	–1.028**	–1.084**	0.16	5
Intersection sight distance quality	–0.951**	–0.940**	0.06	1
Speed limit	0.044**	0.049**	–0.03	11
Maximum grade at 500 ft (absolute value)	0.096*	0.210**	0.05	119

Variable	Model With PropAADT	Model Without PropAADT	Correlation With PropAADT	Percent Difference
Maximum intersection sight distance right	–0.001**	–0.001*	0.01	0
Minimum intersection sight distance right	0.001**	0.001**	–0.09	0
Flat terrain indicator	–0.649**	–0.386**	–0.05	41
2009 indicator	0.182	0.164	0.01	10
2010 indicator	0.092	0.126	0.01	37
2011 indicator	–0.170	–0.132	0.01	22
2012 indicator	–0.004	–0.019	–0.02	79
Major route arterial indicator	–0.886**	–0.992**	–0.14	12
Constant	–17.136**	–6.945**	--	--
Log-likelihood	–333.87	–350.47	--	--
Dispersion parameter	0.017	0.169	--	--

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
--No data.

The model that includes PropAADT (the proportion of entering traffic on the minor route) is capturing the important relationship between major and minor route traffic volumes. The model that does not include PropAADT assumes that there is no relationship between the traffic volumes.

Table 61 shows that this relationship is significant, and should be accounted for in crash prediction models for rural four-leg intersections with stop control on the minor approaches. When PropAADT is removed from the model, the parameter estimates for several factors, including both traffic volumes, change substantially. In fact, the 95-percent confidence intervals for major route traffic are significantly different between models. The same is true for minor route traffic volumes. These results imply that the effect of any traffic volume should not be considered in isolation (i.e., the parameter estimates should not be used to determine a potential effect) because there is not an isolated effect of an approach's AADT. Holistically, this finding leads to the recommendation that researchers consider plausible relationships before data collection and modeling ensue. If researchers ignore these potential interrelationships, the estimates for predisposing factors in crash prediction models may lead to inappropriate conclusions.

Additionally, researchers should give careful consideration to the degree of correlation among predisposing factors included in crash prediction models. For example, a strong, positive correlation was found between major route AADT and minor route AADT (approximately 60-percent linear correlation).

Table 62 presents the models considering major and minor route volumes separately and together. In the model considering both major and minor route AADT, the standard error of parameter estimates is higher, meaning that the estimates are less efficient. However, the predictors are contributing separate information, and the parameter estimates do not change substantially, or significantly.

Table 62. Parameter estimates for major and minor route AADT for SCP crashes.

Variable	Model With Major AADT	Model With Minor AADT	Model With Major and Minor AADT
Log-major AADT	0.650 (0.148)	--	0.428 (0.161)
Log-minor AADT	--	0.614 (0.126)	0.471 (0.136)
Log likelihood	–404.04	–401.28	–397.67

--Variable not included in final model specification.

When considering the indicator for North Carolina and the presence of edge line extensions for inclusion in the analysis, the linear correlation between the two factors is approximately 95 percent. Essentially, nearly all intersections in North Carolina had edge line extensions, while no sites in Ohio had edge line extensions. Table 63 presents models considering these factors individually and together, while accounting for major and minor route volumes.

Table 63. Parameter estimates for North Carolina indicator and edge line extension indicator for SCP crashes.

Variable	Model With Major AADT	Model With Minor AADT	Model With Major and Minor AADT
Log-major AADT	0.352 (0.163)	0.360 (0.162)	0.377 (0.164)
Log-minor AADT	0.469 (0.137)	0.482 (0.138)	0.490 (0.138)
North Carolina indicator	0.413 (0.180)	--	–0.384 (0.584)
Edge line extension indicator	--	0.468 (0.177)	0.826 (0.575)
Log likelihood	–395.03	–394.17	–393.95

--Variable not included in final model specification.

Individually, the predisposing factors provide similar estimates and are both statistically significant with 95-percent confidence. The magnitude of effect is essentially the same, as are the estimates for major and minor route AADT. When considered together, the parameter estimates no longer make sense, and neither factor is statistically significant with even 85-percent confidence. As Kutner et al. stated, considering both factors in the model does not inhibit the ability to obtain a good fit of the data; however, the individual effects of the parameters are in question and the magnitudes of effect indicated by the model do not indicate which factor is the key predictor variable over the other.(50) With a different set of data, the magnitude and direction of effect will be inconsistent with the current estimates for both factors.

It is up to the researcher to determine which predisposing factor should be included in the model, based on sound theory. In this case, the edge line extension indicator is capturing the unobserved differences between intersections in North Carolina and intersections in Ohio. There is only one intersection in North Carolina (in this dataset) that does not have edge line extensions and there are no sites in Ohio that have edge line extensions. If the edge line extension indicator is chosen for the model, the estimated effect will be incorrectly applied, leading to a biased estimate for edge line extensions. The project team retained the indicator for North Carolina in this model and dropped the indicator for edge line extensions.

BENEFITS AND TRADE-OFFS OF ENHANCED DATA AND REFINED CRASH TYPE DEFINITIONS: WASHINGTON HORIZONTAL CURVE STUDY

The project team collected data from the files maintained by HSIS and supplemented by additional information on weather from NOAA and roadside features from RFIP. The database included 9,363 curve segments on rural two-lane roads, observed over 5 years from 2008 to 2012. During that period, there were 5,607 single-vehicle crashes; 2,468 multi-vehicle crashes; 2,492 crashes classified as the high-priority scenario of combination control loss/no vehicle action and road edge departure/no maneuver; and 601 crashes classified as the high-priority scenario of opposite direction/no maneuver. Table 64 presents summary statistics for all the variables in the data collection study for Washington State.

Traffic data consists of two elements: the AADT and the percent of trucks. The AADT was available for each of the years from 2008 to 2012. AADT varied between 87 and 25,844 vehicles per day with a mean of 2,545. The percent truck information was not available for all the curve segments. Of the 9,363 curve segments, percent truck information was available only for 9,164 segments, which were retained for the analysis. Percent trucks ranged between 0 and 66 percent with a mean of 17 percent.

Data for geometric variables, such as lane width, shoulder width, horizontal alignment, and vertical alignment were available in separate HSIS data files, which the project team merged for the analysis. Lane width values ranged from 9 to 20 ft with a mean around 11 ft. Shoulder widths ranged from 0 to 15 ft with a mean around 4 ft. The speed limit variable ranged from 25 and 65 mph, with the mean value very close to 52 mph. For each horizontal curve, data were available for length, degree of the curve, radius, curve angle, and maximum superelevation. Vertical alignment data in the HSIS files included the incoming grades and outgoing grades of vertical curves. The project team created a variable indicating the presence or absence of vertical curve for analysis. Of the total segments, approximately 70 percent were associated with a vertical curve presence. The segment length in the dataset varied between 0.01 and 1.29 mi, with the mean value of 0.11 mi. Degrees of curve vary from slightly more than 0 to 57 degrees, with most of the curves falling between 1 and 10 degrees. Naturally, curve radius and degree of curve are one in the same, as one is calculated from the other. Having information on horizontal and vertical alignment was considered to be an advanced dataset.

Weather data, which includes temperature, snowfall, and precipitation data, comprised one part of the alternative (i.e., enhanced) variable dataset used in the analysis. The project team collected the data through the NOAA database, as mentioned in chapter 6. The mean temperature values ranged between 14 and 56 °F with a mean around 47 °F. Besides mean temperature, the project team also used the number of days with temperature greater than 90 °F and the number of days with temperature below 32 °F as temperature variables. The project team considered total snowfall and the total number of days when precipitation was between 0.1 and 1 inch as wet-condition variables included in the dataset. The average snowfall was around 3.44 inches with a minimum value of 0.02 inches and a maximum value of 63.38 inches of snowfall during the winter months.

Table 64. Descriptive statistics for all variables in Washington roadway segment dataset.

Variable	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
Average annual daily traffic (vehicle/day)	9,363	2,545	2,658	87	25,844
Truck (%)	9,164	17.02	8.72	0	66
Lane width (ft)	9,363	11.42	0.90	9	20
Shoulder width (ft)	9,363	3.93	2.24	0	15
Horizontal curve angle (degrees)	9,363	27.73	22.65	0.26	177.95
Maximum superelevation (ft/ft)	9,363	0.008	0.022	0	0.14
Speed limit (mph)	9,363	51.76	8.37	25	65
Segment length (mi)	9,363	0.11	0.10	0.01	1.29
Curve radius (ft)	9,363	1,925.51	1,946.20	100	11,871
Degree of the curve	9,363	6.84	7.66	0.48	57.3
Grade (%)	9,363	1.76	2.06	0	9.87
Concrete barrier length (mi)	101	0.20	0.35	0.008	1.39
Concrete barrier presence in curve (%)	101	41.81	33.09	2.38	100
Guardrail length (mi)	2,228	0.26	0.275	0.01	1.68
Guardrail presence in curve (%)	2,228	59.99	36.41	1.02	100
Special-use barrier length (mi)	20	0.21	0.25	0.02	1.12
Special-use barrier presence in curve (%)	20	49.38	32.61	5.55	100
Tree count	911	2.18	1.96	1	25
Tree average diameter (ft)	911	5.59	2.69	0.82	12.42
Fixed object count	1,028	2.28	2.68	1	32
Mean temperature (°F)	3,728	47.04	6.13	13.9	55.24
Number of days with temperature greater than 90 °F (day)	3,728	1.87	0.92	1	5
Number of days with temperature less than 32 °F (day)	3,728	9.95	3.96	3	17
Total snowfall (inches)	3,728	3.44	8.00	0.02	63.38
Number of days with tenth inch precipitation (day)	3,728	9.34	3.31	3	14
Number of days with half inch precipitation (day)	3,728	3.80	2.21	1	7

Variable	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
Number of days with 1-inch precipitation (day)	3,728	1.93	1.07	0	4
Total SV crashes (crash)	9,363	0.60	1.22	0	22
Total MV crashes (crash)	9,363	0.26	1.21	0	37
All crashes (crash)	9,363	0.86	1.95	0	44
Total SV scenario 48 (crash)	9,363	0.26	0.73	0	12
Total MV scenario 22 (crash)	9,363	0.064	0.30	0	6

The alternative, enhanced dataset also included variables and information from the RFIP dataset that WSDOT developed. The project team obtained information on the length of roadside features for concrete barrier, guardrail, and special-use barrier. The project team also obtained the count of trees, and fixed objects and the diameter for trees using the roadside inventory data. The count of trees and fixed objects along the horizontal curves ranged from 1 to 25 and 1 to 32, respectively, with mean values for trees and fixed object counts of 2.18 and 2.28, respectively. The sample size of guardrail was significant (2,228 observations) when compared to concrete barrier (101) and special-use barrier (20), as expected. As a result, the project team used guardrail in the estimation of crash frequency models using alternative dataset variables for single-vehicle and multi-vehicle crashes, as well as for the high-priority crash types.

Variables and sample sizes for each crash frequency model was a function of the type of dataset for each of the crash type classifications. Traffic and cross sectional, including AADT, truck percentages, lane and shoulder widths, and speed limit are variables that tend to be available in most traditional datasets used for crash modeling. The descriptive statistics for traditional variables, along with the single-vehicle and multi-vehicle “all” and “critical” crash numbers, are provided in table 65.

Table 65. Descriptive statistics for traditional dataset variables.

Variable	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
Average annual daily traffic (vehicle/day)	9,164	2,593.64	2,666.51	139	25,844
Truck (%)	9,164	17.02	8.72	0	66
Lane width (ft)	9,164	11.44	0.89	9	20
Shoulder width (ft)	9,164	3.97	2.23	0	15
Speed limit (mph)	9,164	51.94	8.30	25	65
Total SV crashes (crash)	9,164	0.61	1.23	0	22
Total MV crashes (crash)	9,164	0.26	1.16	0	34
All crashes (crash)	9,164	0.88	1.92	0	36
Total SV scenario 48	9,164	0.27	0.74	0	12
Total MV scenario 22	9,164	0.06	0.30	0	6

Alignment variables, including horizontal alignment and vertical alignment data, were considered available only in advanced State agency databases. The HSIS database for Washington contained roadway inventory data on horizontal curves that included curve length, radius, superelevation, and other curve identifiers. Similarly, it contained information on vertical curves and grades. The descriptive statistics for the traffic and roadway inventory data that classifies itself under the advanced State agency database variables are provided in table 66. The values for horizontal curve angles that were smaller than 1 degree and greater than 140 degrees (40 observations) seemed abnormal and the project team deleted them from the dataset. As a result, the sample size decreased from 9,164 in traditional dataset to 9,124 in advanced State agency variable dataset.

Table 66. Descriptive statistics for advanced State agency database variables.

Variable	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
Average annual daily traffic (vehicle/day)	9,124	2,597.06	2,669.15	139	25,844
Truck (%)	9,124	17.03	8.72	0	66
Lane width (ft)	9,124	11.44	0.89	9	20
Shoulder width (ft)	9,124	3.97	2.23	0	15
Speed limit (mph)	9,124	51.94	8.30	25	65
Horizontal curve angle	9,124	27.30	21.51	1.00	137.50
Maximum superelevation	9,124	0.008	0.02	0	0.14
Segment length	9,124	0.11	0.10	0.01	1.29
Horizontal curve radius	9,124	1,933.38	1,938.66	100	11,871
Degree of the curve	9,124	6.73	7.51	0.48	57.3
Vertical curve presence ("1 = present", "0 = not present")	9,124	0.74	0.44	0	1
Grade (%)	9,124	1.77	2.06	0	9.87
Total SV crashes (crash)	9,124	0.61	1.23	0	22
Total MV crashes (crash)	9,124	0.26	1.16	0	34
All crashes (crash)	9,124	0.88	1.92	0	36
Total SV scenario 48	9,124	0.27	0.74	0	12
Total MV scenario 22	9,124	0.06	0.30	0	6

The weather and roadside features information that constitutes the alternative, enhanced dataset, along with the traffic data and geometric data is shown in table 67. The sample size of the dataset—3,686 observations—is much smaller than the traditional variable dataset and advanced State agency variable dataset. The amount of available information for each of the variables determined the sample size for the dataset. There was a total of 3,686 segments in the dataset with the weather information. Guardrail was present on 26 percent (973 out of 3,686 segments) of the total segments in the enhanced dataset.

Table 67. Descriptive statistics for alternative and traditional database variables.

Variable	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
Average annual daily traffic (vehicle/day)	3,686	2,602.95	2,553.81	182	21,600
Truck (%)	3,686	16.63	7.72	0	59
Lane width (ft)	3,686	11.53	0.93	10	20
Shoulder width (ft)	3,686	4.05	2.16	0	15
Speed limit (mph)	3,686	52.98	7.76	25	65
Horizontal curve angle	3,686	26.01	20.31	1.05	136.93
Segment length	3,686	0.12	0.11	0.01	1.19
Horizontal curve radius	3,686	1,993.39	1,863.12	103	11,675
Degree of the curve	3,686	5.61	5.44	0.49	55.63
Vertical curve presence (“1 = present”, “0 = not present”)	3,686	0.78	0.42	0	1
Grade (%)	3,686	1.84	2.03	0	8
Guardrail length	3,686	0.07	0.17	0	1.35
Guardrail presence (%)	3,686	15.77	32.37	0	100
Tree count	3,686	0.22	0.88	0	13
Tree average diameter (ft)	3,686	0.57	1.91	0	12.42
Fixed object count	3,686	0.26	1.26	0	32
Mean temperature	3,686	47.16	5.87	13.9	55.24
Number of 90 or greater days	3,686	1.88	0.92	1	5
Number of 32 or less days	3,686	9.93	3.94	3	17
Total snowfall	3,686	3.20	7.08	0.02	63.38
Number of 0.1-inch precipitation days	3,686	9.34	3.32	3	14
Number of 0.5-inch precipitation days	3,686	3.80	2.21	1	7
Number of 1-inch precipitation days	3,686	1.93	1.07	0	4
Total SV crashes (crash)	3,686	0.55	1.15	0	11
Total MV crashes (crash)	3,686	0.20	0.82	0	17
All crashes (crash)	3,686	0.76	1.62	0	23
Total SV scenario 48	3,686	0.23	0.67	0	9
Total MV scenario 22	3,686	0.05	0.30	0	6

Model Estimation Results and Discussion

Results of single-vehicle and multi-vehicle crash frequency models are displayed in table 68 through table 71. All tables consisted of six models that the project team developed for each crash type classification (similar to the intersection crash models):

- Model 1—considering an extensive number of predisposing factors collected from alternative and traditional sources.
- Model 2—considering only statistically significant predisposing factors from alternative and traditional sources.
- Model 3—considering predisposing factors expected to be available in advanced State agency databases.
- Model 4—considering only statistically significant predisposing factors expected to be available in advanced State agency databases.
- Model 5—considering predisposing factors collected from traditional State agency databases.
- Model 6—considering only statistically significant predisposing factors from traditional State agency databases.

Similar to the model estimation at rural four-leg intersections, relaxed values of statistical significance were identified for presentation of results. Predisposing factors significant with at least 80-percent confidence were included in Models 2, 4, and 6; asterisks serve to differentiate between variables that were significant with 80-percent confidence and variables that were significant with 90-percent confidence.

Variables were introduced into the models as per the model classification. The consideration for either including a variable into the model (for Models 2, 4, and 6) or excluding it from the model was the examination of statistical significance value. If the *p*-value was less than 0.20, the variable was included in the model.

The preliminary results show that many predisposing factors were found to be statistically significant for all the models. Specifically, most of the traffic and geometry variables were found to be significant for all the models for all crash types. As expected, as the value of AADT increased, the expected numbers of crashes increased. However, the effect of AADT on multi-vehicle crashes was slightly different than in the single-vehicle crashes. For example, the coefficient for Log AADT for the model including alternative dataset for single-vehicle and multi-vehicle crashes were 0.842 and 0.966 respectively, both significant with 90-percent confidence.

Table 68. Negative binomial regression models for single-vehicle crashes on rural two-lane horizontal curves.

Variable List	Factors From Alternative and Traditional Sources		Factors In Advance State Agency Databases		Factors From Traditional State Agency Databases	
Model Number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Log AADT	0.907**	0.871**	0.822**	0.816**	0.814**	0.813**
Truck (%)	0.005	--	0.0007	--	0.0003	--
Lane width	–0.035	--	–0.044**	–0.045**	–0.055**	–0.055**
Shoulder width	–0.017	--	–0.004	--	–0.016**	–0.016**
Speed limit	0.041**	0.040**	0.031**	0.031**	0.027**	0.027**
Horizontal curve radius	–0.00004**	–0.00004**	–0.00006**	–0.00006**	--	--
Vertical curve presence	–0.104	–0.099*	–0.039	--	--	--
Grade (%)	0.009	--	–0.001	--	--	--
Mean temperature	–0.011*	--	--	--	--	--
Number of 90 or greater days	0.132**	0.159**	--	--	--	--
Number of 32 or less days	–0.026**	–0.029**	--	--	--	--
Total snowfall	–0.015*	--	--	--	--	--
Number of 0.1-inch precipitation days	–0.026	--	--	--	--	--
Number of 0.5-inch precipitation days	0.074	--	--	--	--	--
Number of 1-inch precipitation days	–0.197**	–0.112**	--	--	--	--
Guardrail length	0.150	--	--	--	--	--
Guardrail presence (%)	–0.001*	--	--	--	--	--
Tree count	0.034	--	--	--	--	--
Tree average diameter	0.014	0.024**	--	--	--	--
Fixed object count	0.020	--	--	--	--	--
Segment length	1 (offset)	1 (offset)	1 (offset)	1 (offset)	1 (offset)	1 (offset)
Constant	–8.632	–9.409	–8.030	–7.992	–7.714	–7.713
Log-likelihood	–3,273.155	–3,279.262	–8,707.394	–8,708.195	–8,779.263	–8,779.273
Dispersion parameter	0.933	0.945	0.974	0.976	1.004	1.004

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
--No data.

Table 69. Negative binomial regression models for multi-vehicle crashes on rural two-lane horizontal curves.

Variable List	Factors From Alternative and Traditional Sources		Factors in Advance State Agency Databases		Factors From Traditional State Agency Databases	
Model Number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Log AADT	1.234**	1.19**	1.388**	1.427**	1.393**	1.425**
Truck (%)	–0.009	–0.016**	–0.006*	--	–0.005	--
Lane width	0.139**	0.131**	0.038	--	0.033	--
Shoulder width	0.088**	0.094**	0.015	--	0.010	--
Speed limit	–0.007	–0.009*	–0.009**	–0.010**	–0.011**	–0.012**
Horizontal curve radius	–0.00002	--	–0.00001	--	--	--
Vertical curve presence	0.034	--	–0.100	–0.133**	--	--
Grade (%)	0.004	--	–0.019	--	--	--
Mean temperature	0.034	0.0377*	--	--	--	--
Number of 90 or greater days	0.006	--	--	--	--	--
Number of 32 or less days	–0.052**	–0.049**	--	--	--	--
Total snowfall	0.018*	--	--	--	--	--
Number of 0.1-inch precipitation days	–0.074**	–0.092**	--	--	--	--
Number of 0.5-inch precipitation days	0.111	--	--	--	--	--
Number of 1-inch precipitation days	–0.313**	--	--	--	--	--
Guardrail length	0.497**	0.344**	--	--	--	--
Guardrail presence (%)	–0.002	--	--	--	--	--
Tree count	0.057	0.089**	--	--	--	--
Tree average diameter	0.014	--	--	--	--	--
Fixed object count	0.0393*	--	--	--	--	--
Segment length	1 (offset)	1 (offset)	1 (offset)	1 (offset)	1 (offset)	1 (offset)
Constant	–13.515	–13.111	–12.425	–12.269	–12.432	–12.287
Log-likelihood	–1,573.731	–1,579.171	–4,441.958	–4,444.502	–4,470.716	–4,472.085
Dispersion parameter	1.949	2.01	2.757	2.752	2.750	2.747

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
--No data.

Table 70. Negative binomial regression models for pre-crash scenario 48 (control loss and road edge departure) for single-vehicle crashes on rural two-lane horizontal curves.

Variable List	Factors From Alternative and Traditional Sources		Factors in Advance State Agency Databases		Factors From Traditional State Agency Databases	
Model Number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Log AADT	0.854**	0.838**	0.736**	0.748**	0.704**	0.704**
Truck (%)	0.012**	0.009*	–0.003	--	–0.004*	–0.004*
Lane width	–0.166**	–0.155**	–0.165**	–0.172**	–0.190**	–0.190**
Shoulder width	–0.056**	–0.056**	–0.032**	–0.033**	–0.065**	–0.065**
Speed limit	0.068**	0.068**	0.044**	0.042**	0.029**	0.029**
Horizontal curve radius	–0.0002**	–0.0002**	–0.0002**	–0.0002**	--	--
Vertical curve presence	–0.096	--	–0.032	--	--	--
Grade (%)	0.008	--	–0.006	--	--	--
Mean temperature	–0.018**	–0.016**	--	--	--	--
Number of 90 or greater days	0.050	--	--	--	--	--
Number of 32 or less days	–0.054**	–0.051**	--	--	--	--
Total snowfall	–0.030*	–0.036*	--	--	--	--
Number of 0.1-inch precipitation days	–0.047*	–0.073**	--	--	--	--
Number of 0.5-inch precipitation days	0.025	--	--	--	--	--
Number of 1-inch precipitation days	–0.116	--	--	--	--	--
Guardrail length	0.107	--	--	--	--	--
Guardrail presence (%)	–0.003**	–0.003**	--	--	--	--
Tree count	–0.035	--	--	--	--	--
Tree average diameter	0.042**	0.032**	--	--	--	--
Fixed object count	0.063**	0.059**	--	--	--	--
Segment length	1 (offset)	1 (offset)	1 (offset)	1 (offset)	1 (offset)	1 (offset)
Constant	–7.390	–7.347	–6.827	–6.839	–5.854	–5.854
Log-likelihood	–1,979.157	–1,981.221	–5,457.375	–5,458.181	–5,613.679	–5,613.679
Dispersion parameter	1.401	1.418	1.612	1.612	1.991	1.991

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
--No data.

Table 71. Negative binomial regression models for pre-crash scenario 22 (opposite direction/no maneuver) for multi-vehicle crashes on rural two-lane horizontal curves.

Variable List	Factors From Alternative and Traditional Sources		Factors in Advance State Agency Databases		Factors From Traditional State Agency Databases	
Model Number	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Log AADT	1.126**	1.160**	1.021**	1.033**	0.954**	0.999**
Truck (%)	–0.016	–0.017*	–0.004	--	–0.006	--
Lane width	0.172**	0.162**	0.077*	0.074*	0.037	--
Shoulder width	0.106**	0.104**	0.061**	0.056**	0.020	--
Speed limit	0.032**	0.036**	0.017**	0.016**	0.001	--
Horizontal curve radius	–0.0004**	–0.0004**	–0.0003**	–0.0003**	--	--
Vertical curve presence	–0.041	--	–0.128	--	--	--
Grade (%)	–0.010	--	0.034	--	--	--
Mean temperature	0.031	--	--	--	--	--
Number of 90 or greater days	0.053	--	--	--	--	--
Number of 32 or less days	–0.059*	--	--	--	--	--
Total snowfall	0.021	--	--	--	--	--
Number of tenth-inch precipitation days	–0.012	--	--	--	--	--
Number of half-inch precipitation days	–0.186	–0.090**	--	--	--	--
Number of 1-inch precipitation days	0.118	--	--	--	--	--
Guardrail length	0.474*	0.636**	--	--	--	--
Guardrail presence (%)	0.003	--	--	--	--	--
Tree count	0.111	0.114*	--	--	--	--
Tree average diameter	–0.001	--	--	--	--	--
Fixed object count	0.025	--	--	--	--	--
Segment length	1 (offset)	1 (offset)	1 (offset)	1 (offset)	1 (offset)	1 (offset)
Constant	–15.724	–15.217	–12.334	–12.396	–10.874	–10.727
Log-likelihood	–665.140	–668.498	–1,938.754	–1,939.866	–2,005.333	–2,006.364
Dispersion parameter	2.353	2.394	1.818	1.823	2.255	2.249

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
-- No data.

The analysis results showed that many predisposing factors collected from alternative data sources were statistically significant predictors of single-vehicle crashes, multi-vehicle crashes, combination control/loss and road edge departure single-vehicle crashes, and opposite direction/no maneuver multi-vehicle crashes. Several weather-related variables were found to be statistically significant, especially temperature-related and snowfall variables. The project team found some of the roadside variables not to be statistically significant in the fully specified model estimated using alternative and traditional variables database category. However, these variables were highly significant in the model with only statistically significant predictors estimated with the enhanced dataset (i.e., alternative plus traditional). These seem to indicate that, similar to the unsignalized intersection analysis, parameters estimated with data containing only traditional variables are likely biased, as several variables available in only the enhanced dataset were relevant. For example, for all single-vehicle crashes, the estimated parameter corresponding to the posted speed limit was 0.027 for the model considering only variables from the traditional State database category, 0.031 for the model considering variables from the advanced State database category, and 0.041 for the model considering variables from the enhanced dataset. This indicates that the model considering only variables from the traditional State data underestimated the effect of posted speed limit by 34 percent. For multi-vehicle, “high-priority” crashes, lane width had an estimated coefficient of 0.077 in the model estimated using the advanced State database category and 0.172 in the model estimated with data from the enhanced dataset, indicating that the effect of lane width may be underestimated by 55 percent when certain key variables are not available for model specification. Most predisposing factors appear to be underestimated when using the more limited datasets, indicating that they are correlated with omitted variables that have an opposite safety effect; however, the results were not consistent across all factors.

When comparing the model with only statistically significant predisposing factors drawn from the combined alternative and traditional sources versus the model estimated from the same database with all predisposing factors (regardless of statistical significance), it becomes clear that the estimated effects of some factors were highly dependent on the inclusion of other factors, both for magnitude and direction. For example, number of days with at least 1 inch of precipitation was correlated with the other precipitation-related variables. When considering the model with all predisposing factors for all single-vehicle crashes (regardless of statistical significance), the effect of the number of days with 1 inch of precipitation differs in magnitude from the model with only statistically significant predisposing factors. This is because the precipitation variables included in the model were estimated using the combined alternative and traditional data sources, and the presence of the wet-condition variables has a large effect in this model.

However, the difference in the parameter estimates for this variable is very small. Similar issues exist for other crash types as well. It is necessary to note that these differences in model estimation results are due to the impacts of apparent overdispersion, which stems because of omitted variable bias, irrelevant variables, and greater correlation between independent variables.

Similar to the unsignalized intersection analysis, the dispersion parameter estimate was largest for the model estimated with traditional State data. It was also statistically significant, indicating the need for a negative binomial regression model for all crash types with an exception to the multi-vehicle more refined crash type (opposite direction/no maneuver crashes). For the multi-vehicle “high-priority” crash prediction model, the model estimated using the combination of alternative and traditional sources results in a dispersion parameter that was 4 percent higher, but it was still statistically significant. For the other crash types, the dispersion parameter was smaller when estimating the models using variables from advanced State agency database and/or combination of alternative and traditional sources. For single-vehicle “high-priority” crash types, the model estimated using variables available in the advanced State database category resulted in a dispersion parameter estimate that was 20 percent smaller. When estimating the model using the combined alternative and traditional data sources, the dispersion parameter was 30 percent smaller. These findings suggest that including as many relevant variables as possible in the model specification reduces the magnitude of the dispersion parameter. The findings are similar for other crash types.

The parameter estimates for the predisposing factors in the model were consistent in direction of effect and magnitude for the single-vehicle “all” crashes and single-vehicle “high-priority” crashes. However, for the multi-vehicle crashes, the model parameters were less stable and more difficult to explain. The parameter estimates for lane width and shoulder width were positive, indicating that the expected number of multi-vehicle crashes increases as these widths increase in magnitude. Overall, it appears that the model estimated with a reduced sample size, more relevant variables, and more refined crash type definitions provided better estimates for the crash prediction models.

BENEFITS AND TRADE-OFFS OF ENHANCED DATA AND REFINED CRASH TYPE DEFINITIONS: UTAH HORIZONTAL CURVE STUDY

Similar to the unsignalized intersection study, the Washington State horizontal curve study was conducted with both data from HSIS databases and an enhanced dataset with supplemental information from the RFIP and NOAA. This study, however, focused on using a new LiDAR-based dataset from Utah. In the previous sections, the project team examined the differences between an analysis that relies solely on a “traditional” dataset, advanced dataset, and one with data supplemented from other non-traditional sources. These discussions also highlighted the added values of an enhanced dataset. However, in this Utah study, with a completely new data source, the project team explored the potentials of LiDAR-based data in safety study.

The data collection process was described in chapter 6. Table 72 provides the descriptive statistics of the Utah dataset. The data consist of 1,755 horizontal curves, with crash data from 2009 to 2013 aggregated for each site. Data for the two critical crash scenarios (control loss/road edge departure crashes and opposite direction, no maneuver crashes) are provided along with crash data for all multi-vehicle and all single-vehicle crashes, as well as total crashes. In total, there were approximately 1,220 crashes consisting of approximately 1,041 single-vehicle crashes and 179 multi-vehicle crashes. Additionally, there were approximately 37 multi-vehicle, opposite direction, no maneuver crashes and 296 single-vehicle, control loss/road edge departure crashes.

Table 72. Descriptive statistics for all variables in Utah dataset.

Variable	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
Segment length (mi)	1,755	0.19	0.11	0.05	1.20
Curve radius (ft)	1,755	2,100.03	1,205.19	278	7,350
Degree of curve	1,755	3.73	2.24	0.78	20.61
Horizontal curve angle (degree)	1,755	34.22	24.78	3.43	176.57
Average annual daily traffic (vehicle/day)	1,755	1,162.65	1,406.45	21	9,383
Natural logarithm of average annual daily traffic	1,755	6.55	1.02	3.056	9.147
Winter closure indicator	1,755	0.10	0.30	0	1
Shoulder width (ft)	1,755	2.98	1.85	0	15.5
Right barrier length (mi)	278	0.08	0.07	0.001	0.51
Left barrier length (mi)	194	0.07	0.05	0.001	0.29
Proportion of total barrier presence in curve on both sides (%)	1,755	12	29	0.00	200
Mean temperature (°F)	1,755	48.19	5.16	36.5	62.0
Number of days with temperature greater than 90 °F (day)	1,755	35.87	31.61	0	120.6
Number of days with temperature less than 32 °F (day)	1,755	166.42	39.14	62.8	238.2
Total snowfall (inches)	1,755	61.53	58.04	1.3	322.3
Number of days with 0.1 inch of precipitation (day)	1,755	38.12	15.01	14.2	87.2
Number of days with 0.5 inch of precipitation (day)	1,755	7.20	4.45	2	32
Number of days with 1 inch of precipitation (day)	1,755	1.28	1.17	0	9.8
All crashes (crash)	1,755	0.70	1.51	0	18
Total SV crashes (crash)	1,755	0.59	1.28	0	16
Total MV crashes (crash)	1,755	0.10	0.43	0	8
Total MV scenario 22 (crash)	1,755	0.02	0.16	0	2
Total SV scenario 48 (crash)	1,755	0.17	0.63	0	12

Model Estimation Results and Discussions

The first step of the model development process consisted of exploring and understanding the key statistical characteristics of the data. The focus of this process was to determine the possible correlations between explanatory variables. Variables were plotted against each other in pairs in an X–Y, two-dimensional coordinate system for visual and, if necessary, quantitative examinations. The project team evaluated the correlations between variables based on both the distribution of the data points on the graph and the R^2 value of a linear trend line. Figure 37 through figure 41 show those variables with high levels of correlation. The graphs suggest that number of days with temperature of at least 90 °F (Ndays90Plus), number of days with temperature of at most 32 °F (Ndays32less), mean temperature (MeanTemp), and the total snowfall (TotalSnow) were highly correlated. This finding is certainly reasonable and expected, given the fact that those curves with lower temperature are often located at higher elevations and have more snowfall (TotalSnow), more cold days (Ndays32less), and fewer hot days (Ndays90plus). Among other variables, variables representing numbers of days with at least 0.1, 0.5, and 1 inch of precipitation were correlated as well. The knowledge of correlations between variables provided critical guidance to the model development process.

The project team developed negative binomial models for the following five datasets:

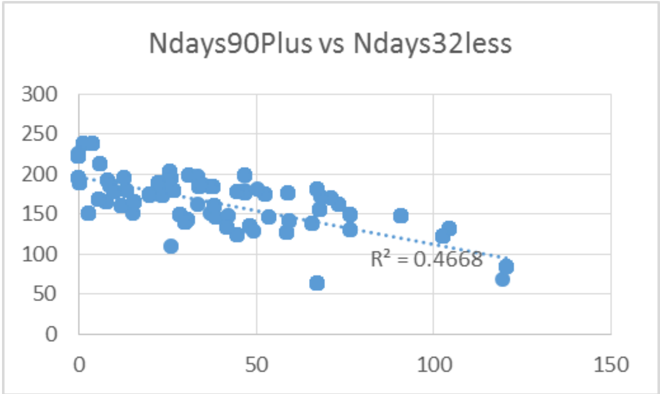
- All crashes.
- Single-vehicle crashes.
- Multi-vehicle crashes.
- Multi-vehicle, opposite direction, no maneuver crashes.
- Single-vehicle, control loss/road edge departure crashes.

For each set of data, the project team estimated two models:

- Model 1—considering all predisposing factors collected for Utah dataset.

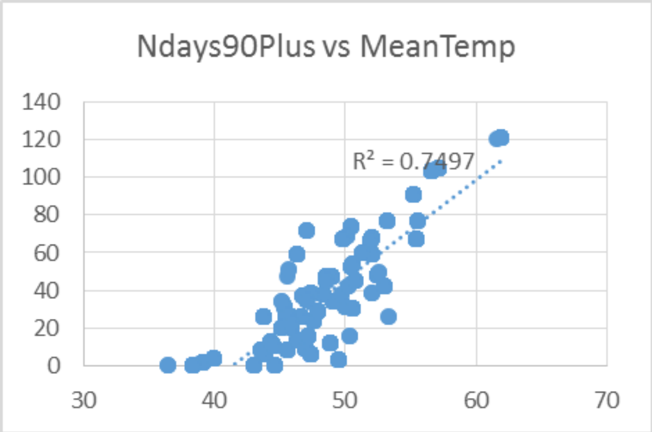
- Model 2—considering only a selected set of variables that provides statistical significant estimates and adequate model fit.

The model development and model estimation results for each crash categories are discussed in the remainder of this section.



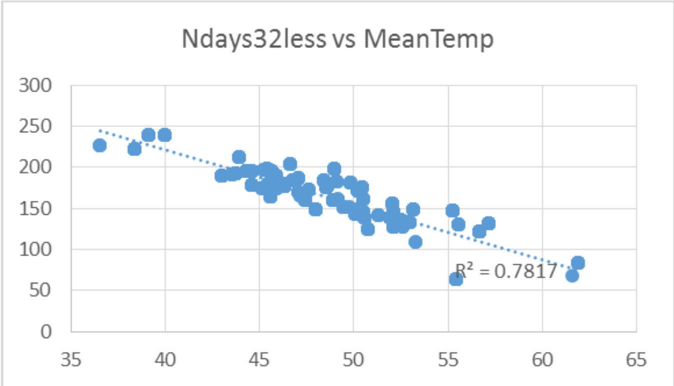
Source: FHWA.

Figure 36. Chart. Correlation plot between number of days with temperature greater than 90 °F and less than 32 °F.



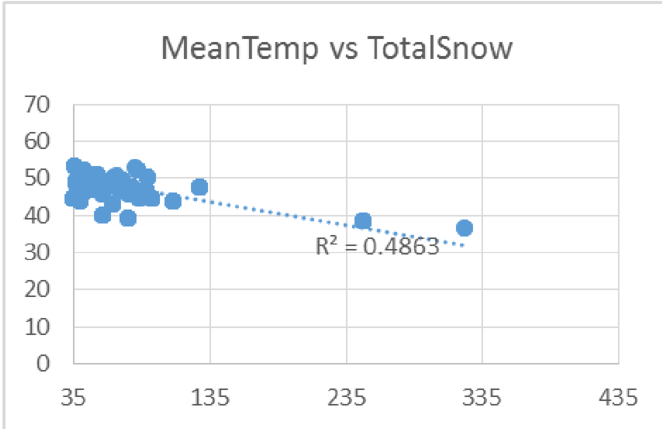
Source: FHWA.

Figure 37. Chart. Correlation plot between number of days with temperature greater than 90 °F and mean temperature.



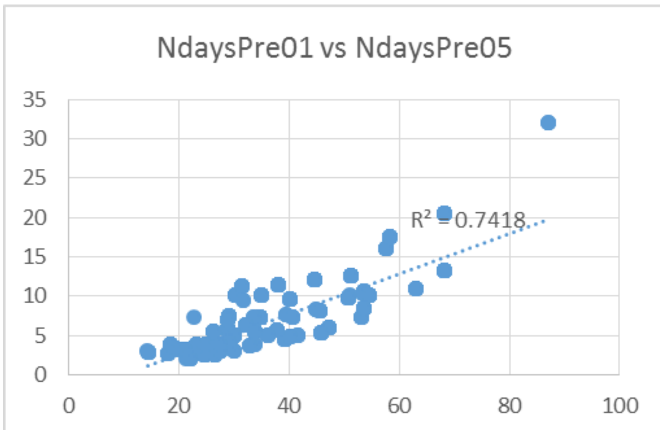
Source: FHWA.

Figure 38. Chart. Correlation plot between number of days with temperature less than 32 °F and mean temperature.



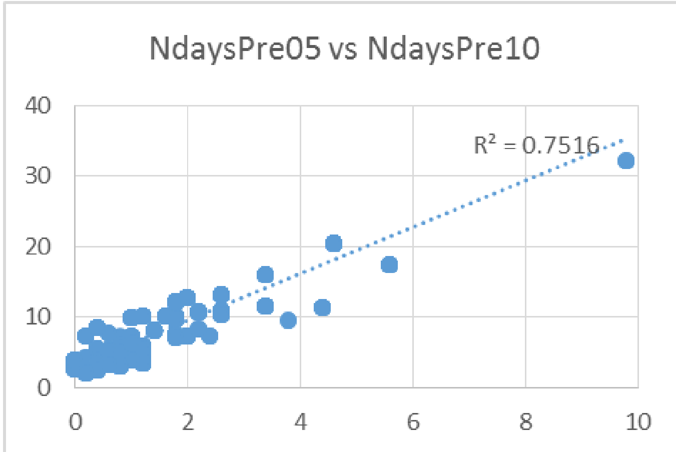
Source: FHWA.

Figure 39. Chart. Correlation plot between mean temperature and total snow.



Source: FHWA.

Figure 40. Chart. Correlation plot between number of days with 0.1 inch of precipitation and 0.5 inch of precipitation.



Source: FHWA.

Figure 41. Chart. Correlation plot between number of days with 0.5 inch of precipitation and 1 inch of precipitation.

Models for All (Total) Crashes

The model development process started with the simplest specification in which only traffic volume was included as an explanatory variable in the form of the natural logarithm of AADT to capture the nonlinear relationship between traffic volume and expected crashes. After this first specification came specifications that added other variables to the model. The project team added these variables individually to the model, tested their combined interactions, and estimated and evaluated the new model with each addition of a new variable or a new combination of variables. Both goodness of fit from a statistical standpoint and engineering judgment were critical parts of this process. While the 95-percent level of confidence provided a general reference point, the project team did not use this number as a fixed threshold for variable inclusion. While some variables did not achieve this level of confidence, the project team still included them in the model if there were reasons to do so from an engineering perspective.

The prior knowledge about correlations between variables became critical during this process. Even though the correlated variables were still added to the model together for the purpose of confirming the effects of correlation, each of them was evaluated individually, and eventually only the specification resulting in the best fit was used. As previously discovered, mean temperature (MeanTemp), number of days with temperature of at least 90 °F (Ndays90Plus), number of days with temperature of at most 32 °F (Ndays32less), and total snowfall (TotalSnow) were highly correlated. With this fact in mind, the project team tested each one of these variables and found that mean temperature (MeanTemp) provided the best fit for the all (total) crashes model. Similarly, variables representing numbers of days with at least 0.1, 0.5, and 1 inch of precipitation (NdaysPre01, NdaysPre05, and NdaysPre10 respectively) were also correlated. NdaysPre10 provided the best fit. The project team also collected information on winter closure of the horizontal curve (wntr_clo), coded closure as an indicator and tested it with the model. Although this variable provided a good fit when it stood with only traffic volume (logaadt) and degree of curve (deg_curv), it became statistically insignificant when MeanTemp (or any of the variables that have strong correlations with MeanTemp) were present in the model. The reason for this is likely an inherent correlation between winter closure and mean temperature or snowfall. Those sections of roadway that are closed are so because they are located at higher elevation and have lower temperature and more snowfall. The project team eventually dropped the winter closure variable from all models due to this finding. The project team did not find roadside barriers to have an acceptable level of statistical significance in the model for all (total) crashes dataset, or any other models. Therefore, the project team excluded this data element from all models in this analysis as well. Table 73 shows the key parameters for all (total) crashes model.

Table 73. Model parameters for all (total) crashes.

Variable List	Estimated Coefficient Model 1	Estimated Coefficient Model 2
Log AADT	0.898**	0.931**
Shoulder width	–0.019	–0.030*
Degree of curve	0.083**	0.094**
Number of 0.1-inch precipitation days	0.005	--
Number of 0.5-inch precipitation days	0.045*	--
Number of 1-inch precipitation days	–0.028	0.094**
Mean temperature	–0.046*	–0.025**
Total snowfall	–0.003**	--
Number of 90 °F or greater days	–0.002	--
Number of 32 °F or less days	–0.002	--
Winter closure indicator	0.129	--
Barrier percentage	0.039	--
Constant	–2.953*	–4.448**
Segment length	1 (offset)	1 (offset)
Log-likelihood	–1,602.99	–1,610.87
Dispersion parameter	0.444	0.479

*Significant with 80-percent confidence.
**Significant with 95-percent confidence.
--No data.

In this model, all variables except one are statistically significant with 99-percent confidence ($p < 0.01$). Mean temperature has a negative coefficient. This indicates that the expected number of all crashes increases as mean temperature (MeanTemp) decreases. In this context, MeanTemp is likely to act as a surrogate for other weather elements that have inherent correlations with it (e.g., Snow fall). The project team found a negative coefficient for shoulder width (shld_wid). This means that the expected number of crashes on rural horizontal curves increase as shoulder width decreases, which is an intuitive finding. Additionally, number of days with at least 1 inch of precipitation (NdaysPre10) represents the adverse weather conditions for driving that comes with rain. This variable has a positive coefficient, suggesting that expected number of crashes increases as precipitation increases. Meanwhile, a positive coefficient for the degree of curve (deg_curv) suggests that the expected number of crashes has a negative relationship with curve radius. The project team expected this finding because sharper curves are likely to have adverse effects on driving, and therefore on safety as well. However, it is worth noting that this model does not capture the effects of tangents before and after the horizontal curves. From design consistency and driver expectancy standpoints, a curve with a 1,000-ft radius between two long tangents is not the same as a curve with the same basic features but located on a winding section of roadway.

MODELS FOR ALL SINGLE-VEHICLE CRASHES

Table 74 summarizes the key parameters for the final model for all single-vehicle crashes. The project team found total snowfall (TotalSnow) to have better fit than mean temperature (MeanTemp) in the model for all single-vehicle crashes. While shoulder width is not significant with 95-percent confidence, it is at 90 percent. In a single-vehicle crash scenario, a wider shoulder provides more room (and hence more time) for correction and crash avoidance maneuver, should a vehicle deviate from its intended path a long a horizontal curve.

Table 74. Negative binomial regression models single-vehicle crashes on rural two-lane horizontal curves.

Variable List	Estimated Coefficient Model 1	Estimated Coefficient Model 2
Log AADT	0.855***	0.888***
Shoulder width	–0.016	–0.035**
Degree of curve	0.087***	0.099**
Number of 0.1-inch precipitation days	0.004	--
Number of 0.5-inch precipitation days	0.052**	--
Number of 1-inch precipitation days	–0.048	0.090***
Mean temperature	–0.046*	--
Total snowfall	–0.003***	0.001***
Number of 90 °F or greater days	–0.001	--
Number of 32 °F or less days	–0.002	--
Winter closure indicator	0.034	--
Barrier percentage	0.027	--
Constant	–2.909*	–5.561***
Segment length	1 (offset)	1 (offset)
Log-likelihood	–1,488.78	–1,498.80
Dispersion parameter	0.454	0.489

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
***Significant with 95-percent confidence.
-- No data.

Models for All Multi-Vehicle Crashes

The project team evaluated explanatory variables and variable combinations for the multi-vehicle crash model in the same manner as all crashes and all single-vehicle crashes. The findings indicate that the relationship between the expected number of multi-vehicle crashes and traffic volume, shoulder width and degree of curve is very similar to those found in all crashes and all single-vehicle crashes models. While the magnitudes of the effects these elements have on safety vary slightly, the sign and the statistical significance level are very similar. Table 75 provides a summary of the model parameters for multi-vehicle crashes.

Table 75. Negative binomial regression models for multi-vehicle crashes on rural two-lane horizontal curves.

Variable List	Estimated Coefficient for Model 1	Estimated Coefficient for Model 2
Log AADT	1.167***	1.163***
Shoulder width	–0.050	–0.058*
Degree of curve	0.056**	0.072***
Number of 0.1-inch precipitation days	0.014	--
Number of 0.5-inch precipitation days	0.014	0.036***
Number of 1-inch precipitation days	0.048	--
Mean temperature	–0.012	--
Total snowfall	–0.006**	--
Number of 90 °F or greater days	–0.007	--
Number of 32 °F or less days	–0.004	--
Winter closure indicator	0.664*	--
Barrier percentage	0.225	--
Constant	–8.084**	–9.325***
Segment length	1 (offset)	1 (offset)
Log-likelihood	–446.35	–471.17
Dispersion parameter	0.527	0.619

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
***Significant with 95-percent confidence.
-- No data.

Model for Multi-Vehicle, Opposite Direction, No Maneuver Crashes

Similar to the all multi-vehicle crashes dataset, temperature and snowfall do not have a strong statistical association with the expected number of crashes for this model (multi-vehicle, opposite direction, no maneuver crashes on horizontal curves). The project team added MeanTemp to the model and tested it. However, it did not result in a statistically significant parameter estimate, having a *p*-value of 0.43. Although the significance levels for the estimates of TotalSnow, Ndays32less, and Ndays90Plus vary slightly, their parameter estimates fall into the same range in terms of *p*-values. The results for shoulder width (and its various forms) were also similar with *p*-values for average shoulder width (shld_wid), right and left shoulder width (r_shoulder and l_shoulder) of 0.48, 0.54, and 0.56, respectively. With these findings, shoulder width was dropped from the final model for this crash type. The final model parameters are presented in table 76.

Table 76. Negative binomial regression models for pre-crash scenario 22 (opposite direction/no maneuver) for multi-vehicle crashes on rural two-lane horizontal curves.

Variable List	Estimated Coefficient for Model 1	Estimated Coefficient for Model 2
Log AADT	1.054***	1.096***
Shoulder width	0.002	0.151***
Degree of curve	0.161***	--
Number of 0.1-inch precipitation days	0.020	0.016*
Number of 0.5-inch precipitation days	0.101	--
Number of 1-inch precipitation days	–0.508*	--
Mean temperature	0.126	--
Total snowfall	0.002	--
Number of 90 °F or greater days	–0.010	--
Number of 32 °F or less days	0.011	--
Winter closure indicator	–0.629	--
Barrier percentage	0.428	--
Constant	–18.995*	–11.316***
Segment length	1 (offset)	1 (offset)
Log-likelihood	–138.16	–140.94
Dispersion parameter	0.00	0.00

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
***Significant with 95-percent confidence.
-- No data.

Model for Single-Vehicle, Control Loss/Road Edge Departure Crashes

The model parameter estimation result for single-vehicle control loss/road edge departure crashes is summarized in table 77. It is quite similar to the models for all crashes and single-vehicle crashes in terms of variable composition. However, the project team found precipitation to have the opposite sign with a –0.082 coefficient for NdaysPre10, even though the estimate is only significant at the 84-percent level of confidence with a *p*-value of 0.16.

Overall, the model estimation results are as expected and are reasonable from an engineering perspective, except for one precipitation variable: the model for single-vehicle, control loss/road edge departure crashes. Traffic volume, degree of curve, and adverse weather conditions have negative effects on safety (or positive relationships with expected number of crashes).

Table 77. Negative binomial regression models for pre-crash scenario 48 (control loss and road edge departure) for single-vehicle crashes on rural two-lane horizontal curves.

Variable List	Estimated Coefficient for Model 1	Estimated Coefficient for Model 2
Log AADT	1.013***	1.045***
Shoulder width	–0.034	–0.047*
Degree of curve	0.212***	0.215***
Number of 0.1-inch precipitation days	0.008	--
Number of 0.5-inch precipitation days	0.036	--
Number of 1-inch precipitation days	–0.169	–0.082*
Mean temperature	–0.053	–0.037***
Total snowfall	–0.003	--
Number of 90 °F or greater days	0.004	--
Number of 32 °F or less days	0.002	--
Winter closure indicator	0.140	--
Barrier percentage	–0.086	--
Constant	–6.180*	–6.363***
Segment length	1 (offset)	1 (offset)
Log-likelihood	–662.41	–664.44
Dispersion parameter	0.872	0.894

*Significant with 75-percent confidence.
**Significant with 90-percent confidence.
***Significant with 95-percent confidence.
--No data.

CHAPTER 10. ALTERNATIVE APPROACHES TO ESTIMATING EXPOSURE ON RURAL TWO-LANE HIGHWAYS

After exploring and assessing multiple traditional and non-traditional data sources, as outlined in the previous chapters of this report, it still proved difficult to find data on traffic patterns at more disaggregate levels than estimates of daily traffic. This is commonly true for rural areas, such as the ones this study explores. This chapter includes descriptions of the methodologies and results of additional research on availability, possible benefits, and challenges of different traffic volume data collection alternatives.

STUDY 1: ADDITIONAL TRAFFIC SOURCES FOR UTAH AND NORTH CAROLINA DATASETS

AADT values are key inputs to various transportation engineering and planning decisions.[(53,54)] AADTs represent the average 24-hour traffic volume at a given location over a full 365-day period. AADT is also a key variable in statistical road safety models, where the influence of predisposing factors is related to expected crash frequency (sometimes disaggregated by crash type and severity) through some type of regression modeling. It is possible to estimate AADTs more precisely at sites having permanent automatic traffic recorders (ATRs) that accurately record traffic flows throughout the year. However, for the majority of rural roadway segments in a region or State, estimating AADTs frequently involves extrapolating short-term local counts over time and space. These latter types of estimates have significant uncertainty, and are unable to capture differences in traffic volume distributions throughout the day, which can have significant effects on safety performance. For example, two sites with the same AADT may have significantly different safety performance due to differences in day/night volume distributions.

The lack of information on daily travel patterns, as well as suspected uncertainty in daily volume estimates, remains the “elephant in the room” when analyzing rural road safety. The literature on estimating traffic volumes on roads that do not have continuous traffic volume counts is also quite limited. However, there are some recent studies that have looked at improving the AADT estimates for roads without traffic counts using spatial statistics. Researchers found that models that take both spatial trends and spatial autocorrelation into account provided acceptable traffic volume predictions for locations where no observed traffic volume data exists.[(54,55)] To further explore these promising approaches, this study focuses on spatial interpolation of day and night traffic volumes throughout the States of Utah and North Carolina using the ATR traffic volume data that is available at specified locations in both the States.

Study 1 Objectives

The objective of this study is to explore the use of kriging interpolation techniques to estimate day and night traffic volume information at rural horizontal curve locations in Utah and rural unsignalized intersection locations in North Carolina. It is suspected that, by having accurate more disaggregated estimates of traffic volumes at these locations, it is possible to estimate safety with higher levels of confidence. We will explore this by comparing statistical road safety models with and without the disaggregated, day/night volume estimates.

The specific steps of the study are as follows:

- Develop and select a best fitting variogram model with the traffic volume data as a primary variate.
- Select a preferred kriging method from the analysis of the cross-validation error statistics through variogram models.
- Use the selected kriging method to predict standardized day and night traffic volumes at unmeasured locations.
- Perform cross-validation of interpolated day and night traffic volume data results obtained from the preferred kriging approach.
- Incorporate day and night traffic volume estimates into statistical safety models and compare the alternative model specifications with models having more traditional AADT estimates.

Methods and Data

This section describes data sources and examines the summary statistics. The project team collected traffic data that they used to estimate variogram models at ATR stations throughout Utah (for the horizontal curve analysis) and North Carolina (for the unsignalized intersection analysis). Traffic volumes disaggregated by hour are available on the UDOT website for 99 locations throughout the entire State of Utah and for the entire analysis period (2009–2013). The project team obtained ATR data for stations across North Carolina from NCDOT through a formal data request. Both States collected traffic volumes at the ATR locations by hour of the day, each day, and compiled them to generate monthly and yearly totals by the hour of the day.

The project team developed the spatial prediction methodology they used to estimate day and night traffic volumes at locations without ATRs in this study using a kriging-based geostatistical approach. The methodological framework consists of the following three steps:

- Data preparation and transformation—categorizing and performing exploratory data analysis and normality tests for the observed traffic volume data during day and night at locations with ATRs.
- Variogram modeling and kriging interpolation—fitting and selecting appropriate variogram models, performing kriging interpolations, and estimating kriging error.
- Performance assessment of kriging method—employing “leave-one-out” cross-validation and a method similar to the K-fold cross-validation that involves removing 10 percent of the observed (i.e., ATR-measured) sample from the dataset, and then predicting the values at those locations using information from the remaining observations.

The following subsections describe each of these three steps in detail.

Data Preparation and Transformation

Hourly traffic volume data obtained from the ATR stations were disaggregated into estimates of daytime and nighttime traffic volumes. Day and night were categorized based on the sunrise and sunset times every month. The time period from 1.5 hours after sunset to 1.5 hours before the sunrise was considered to be night, with the remaining hours classified as daytime. There is also a twilight period that exists between day and night hours. Twilight periods were considered as day for this analysis.[56,57] The lengths of day and night vary greatly throughout the year, because of the high latitude of the United States. In the middle of June, the nights are approximately 6 hours long, whereas in December, nights are about 11 hours long. To take into account this significant variation in day and night durations, the project team divided the entire dataset into four seasons, based on months of the year. The project team then determined day and night volumes by season of the year across each of the 5 years used in the analysis (2009–2013). Table 78Table 78 shows the day and night times for North Carolina and Utah depending on the season of the year.

Table 78. Day and night times by season of the year in North Carolina and Utah.

Season/Month of the Year	Day	Night
North Carolina spring: March–May	5 a.m.–9:59 p.m.	10 p.m.–4:59 a.m.
North Carolina summer: June–August	5 a.m.–9:59 p.m.	10 p.m.–4:59 a.m.
North Carolina fall: September–November	6 a.m.–6:59 p.m.	7 p.m.–5:59 a.m.
North Carolina winter: December–February	5 a.m.–6:59 p.m.	7 p.m.–4:59 a.m.
Utah spring: March–May	5 a.m.–9:59 p.m.	10 p.m.–4:59 a.m.
Utah summer: June–August	4 a.m.–9:59 p.m.	10 p.m.–3:59 a.m.
Utah fall: September–November	6 a.m.–7:59 p.m.	8 p.m.–5:59 a.m.
Utah winter: December–February	6 a.m.–6:59 p.m.	7 p.m.–5:59 a.m.

The normal distribution of data is a basic requirement of the kriging-based geostatistical approach.[58,59] Kriging assumes that data exhibits stationarity. The notion of stationarity underpins geostatistics, and allows us to assume that there is the same degree of variation from place to place.[60] Kriging also assumes that the correlation (covariance or semivariogram) between any two locations depends only on the distance between them, not on their exact locations. Kriging leads to an optimum estimator and yields best results when the data are normally distributed. Thus, the inconsistency present in the observed data should be identified and fixed (i.e., transformed) prior to model development and analysis. This includes detecting and removing outliers, performing normality tests for the observed traffic volume data, and applying data transformations for non-normal datasets. A log-transformation is very common and often used for the data that have skewed or non-normal distributions.

To meet the assumption of data normality, the distribution in the histogram should be bell-shaped and the skewness value should be around zero. In this study, the project team applied a log transformation to the traffic volume data when the datasets did not satisfy the normal distribution assumption. Once the data transformation was completed, the project team tested the transformed data more formally to determine whether the evidence was sufficient to accept the normal distribution assumption. The Shapiro–Wilk test and the Shapiro–Francia test use a W and W' test statistic, respectively, to determine the likelihood that a random sample comes from a normal distribution. Small values of W and W' are evidence of departure from normality. For additional verification, the project team used a visual examination of the normal QQ plot to graph the data distribution against the standard normal distribution. After transforming and testing all the observed data for the normal distribution, the project team used the resulting day and night traffic volume datasets for the variogram modeling and kriging interpolation. Table 79 and table 80 provide the standard descriptive statistics for the day and night traffic volume data with and without the log transformation. Figure 42 through figure 57 show the frequency distributions and normal probability plots for the untransformed and transformed datasets.

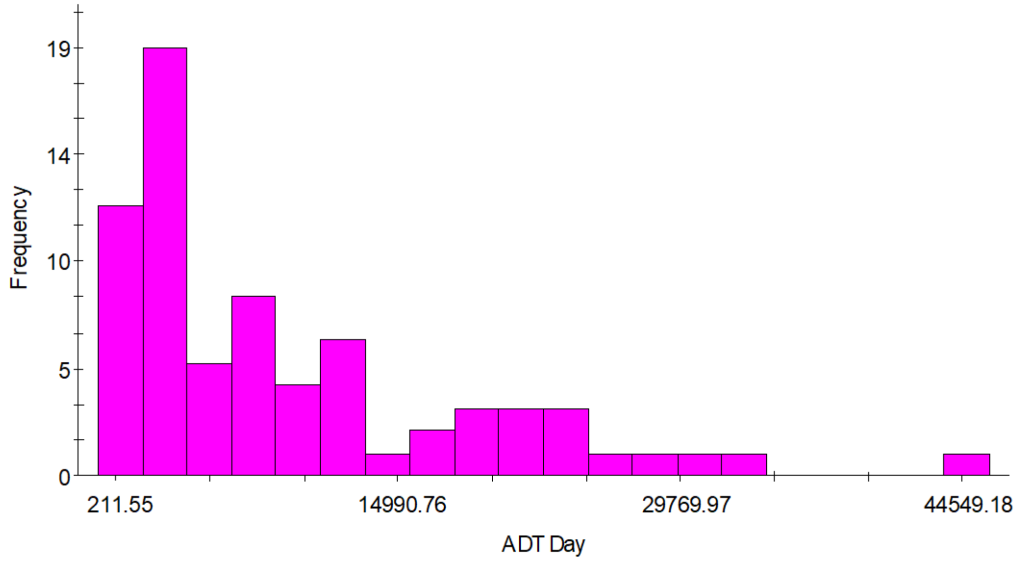
Table 81 and table 82 provide the test statistics for the hypothesis tests carried out to verify the normal distribution assumption. The transformed data for day and night traffic volumes met the normality assumptions based on values for skewness, kurtosis, and the two hypothesis tests.

Table 79. Standard descriptive statistics for the day and night non-transformed and transformed datasets—North Carolina.

Summary Statistics	Non-Transformed Traffic Volumes During the Day	Transformed Traffic Volumes During the Day	Non-Transformed Traffic Volumes During the Night	Transformed Traffic Volumes During the Night
Mean	8,895.99	3.643	1,222.46	2.715
Standard deviation	9,569.02	0.584	1,563.91	0.636
Sample variance	91,566,113	0.341	2,445,805	0.405
Minimum value	211.55	2.33	30.86	1.49
Maximum value	44,549.18	4.65	9,189.19	3.96

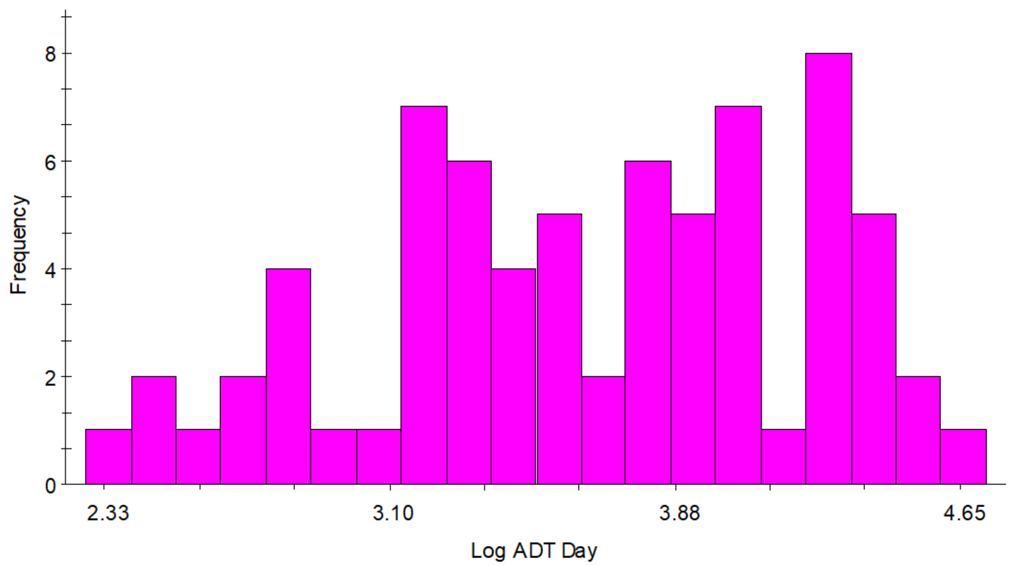
Table 80. Standard descriptive statistics for the day and night non-transformed and transformed datasets—Utah.

Summary Statistics	Non-Transformed Traffic Volumes During the Day	Transformed Traffic Volumes During the Day	Non-Transformed Traffic Volumes During the Night	Transformed Traffic Volumes During the Night
Standard deviation	29,496.19	0.592	4,231.30	0.648
Sample variance	70,025,639	0.350	17,903,923	0.419
Minimum value	428.65	2.63	31.27	1.50
Maximum value	146,411.84	5.17	20,358.27	4.31



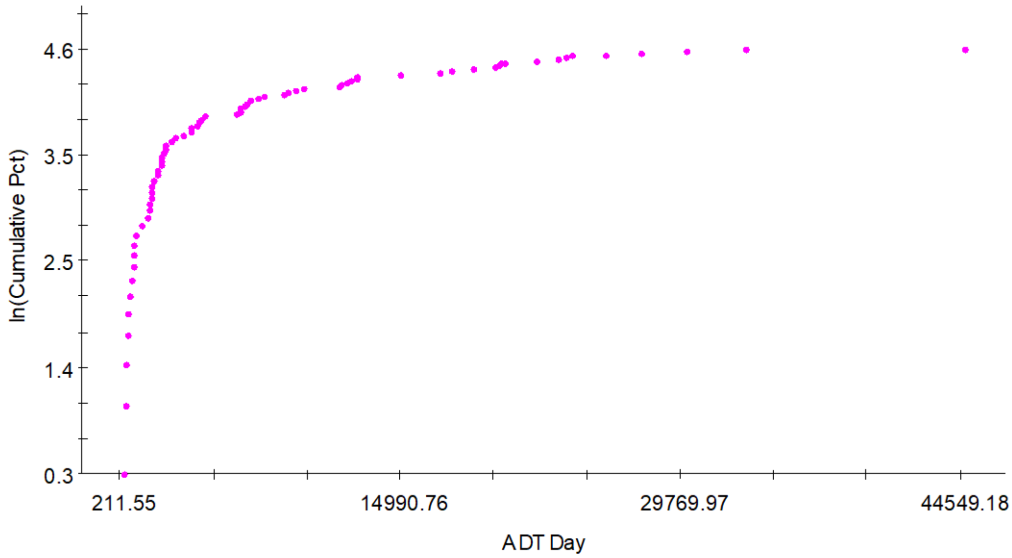
Source: FHWA.

Figure 42. Graph. Frequency distribution plot for ADT day non-transformed dataset for North Carolina.



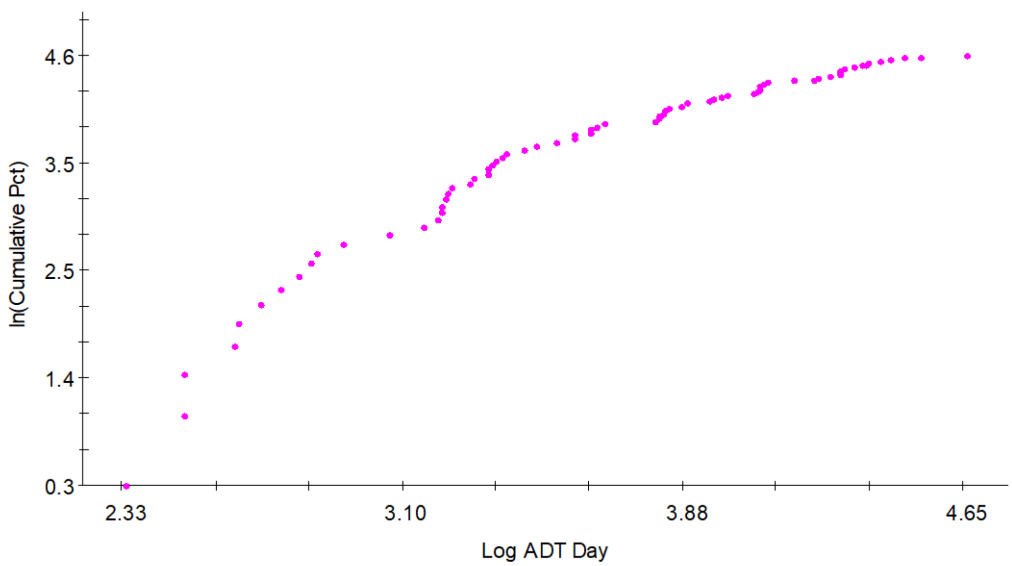
Source: FHWA.

Figure 43. Graph. Frequency distribution plot for ADT day transformed dataset for North Carolina.



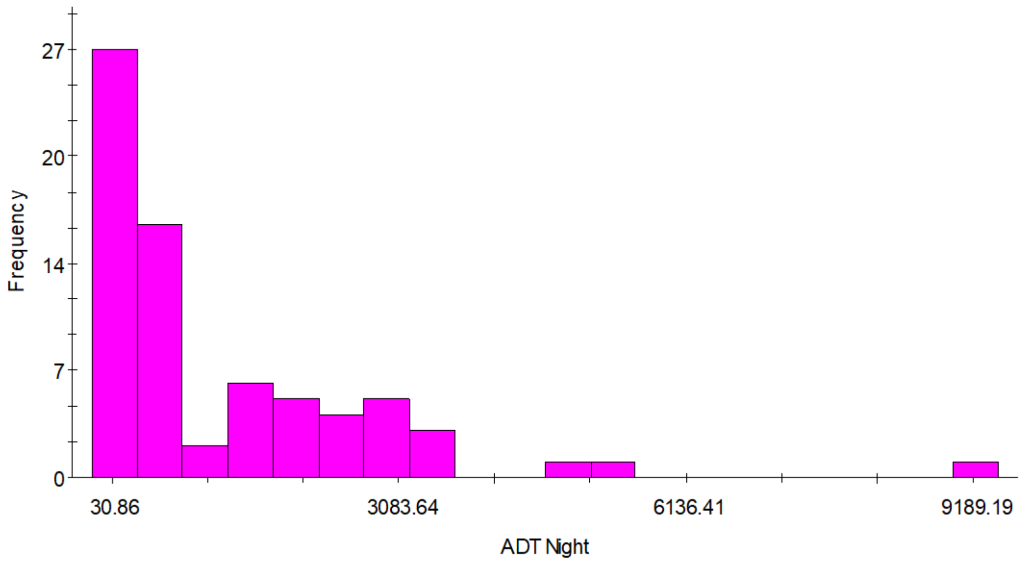
Source: FHWA.

Figure 44. Graph. Normal probability plot for ADT day non-transformed dataset for North Carolina.



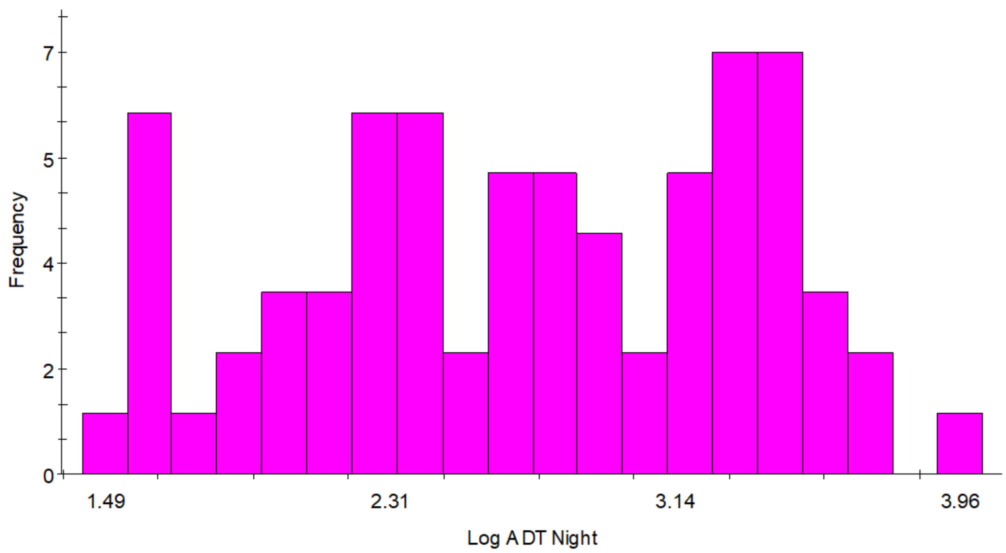
Source: FHWA.

Figure 45. Graph. Normal probability plot for ADT day transformed dataset for North Carolina.



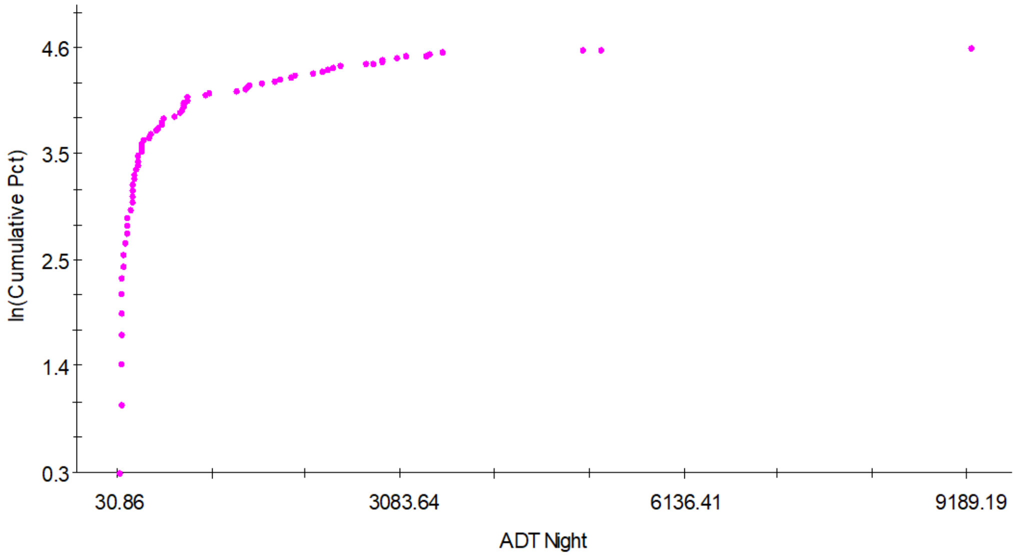
Source: FHWA.

Figure 46. Graph. Frequency distribution plot for ADT night non-transformed dataset for North Carolina.



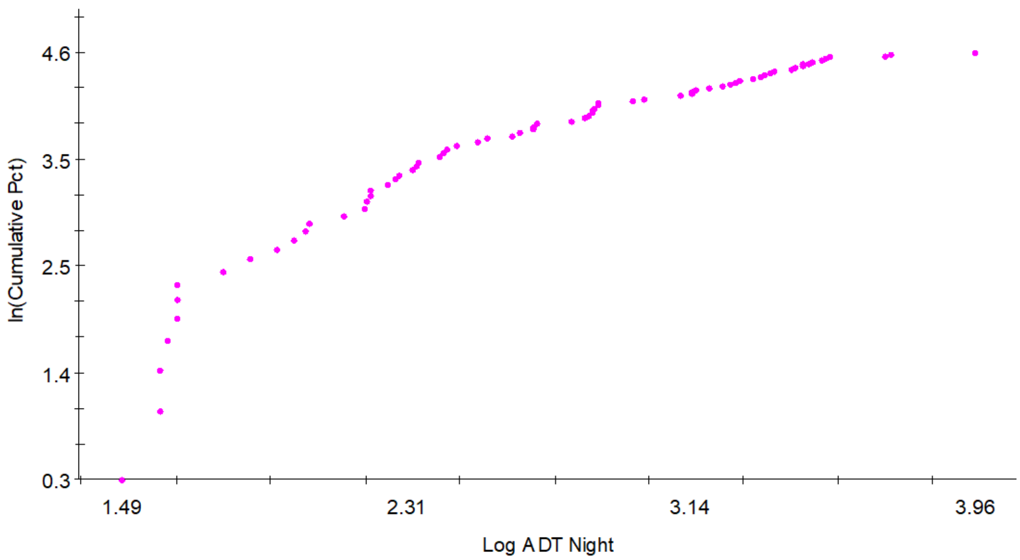
Source: FHWA.

Figure 47. Graph. Frequency distribution plot for ADT night transformed dataset for North Carolina.



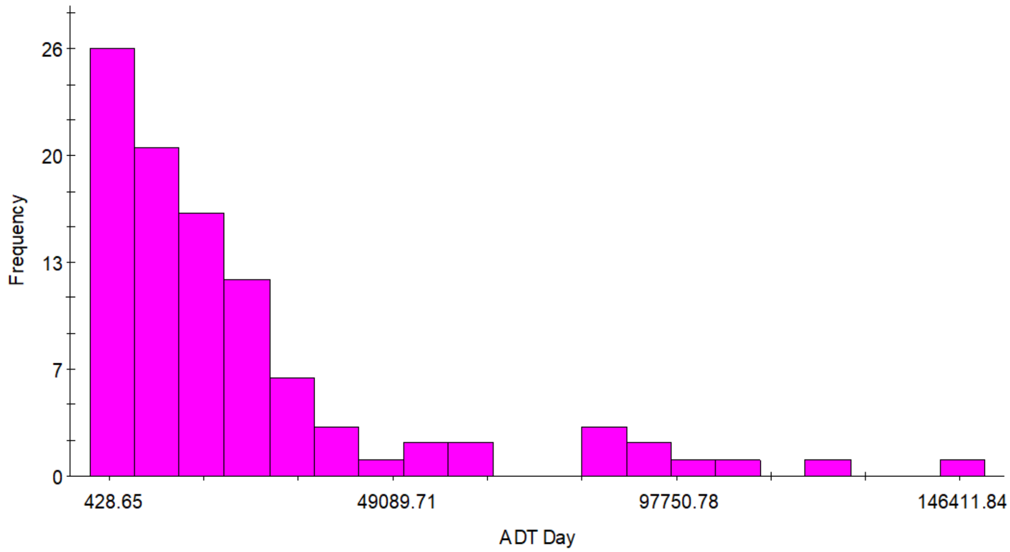
Source: FHWA.

Figure 48. Graph. Normal probability plot for ADT night non-transformed dataset for North Carolina.



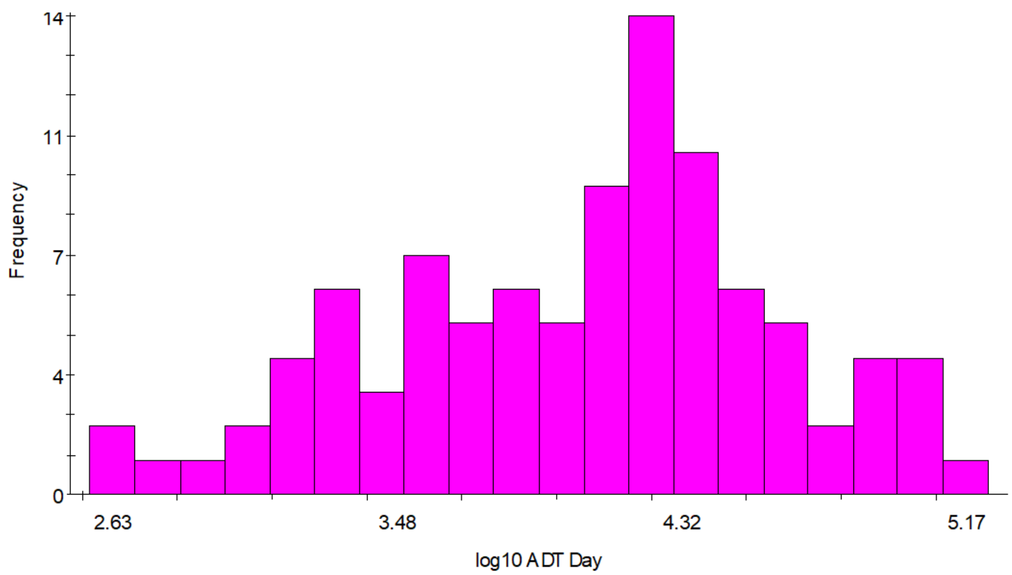
Source: FHWA.

Figure 49. Graph. Normal probability plot for ADT night transformed dataset for North Carolina.



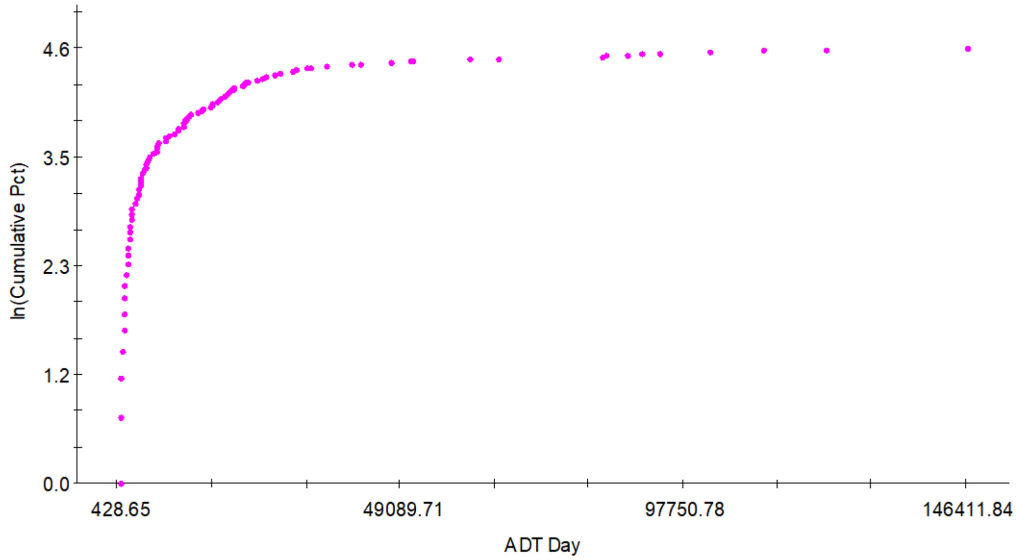
Source: FHWA.

Figure 50. Graph. Frequency distribution plot for ADT day non-transformed dataset for Utah.



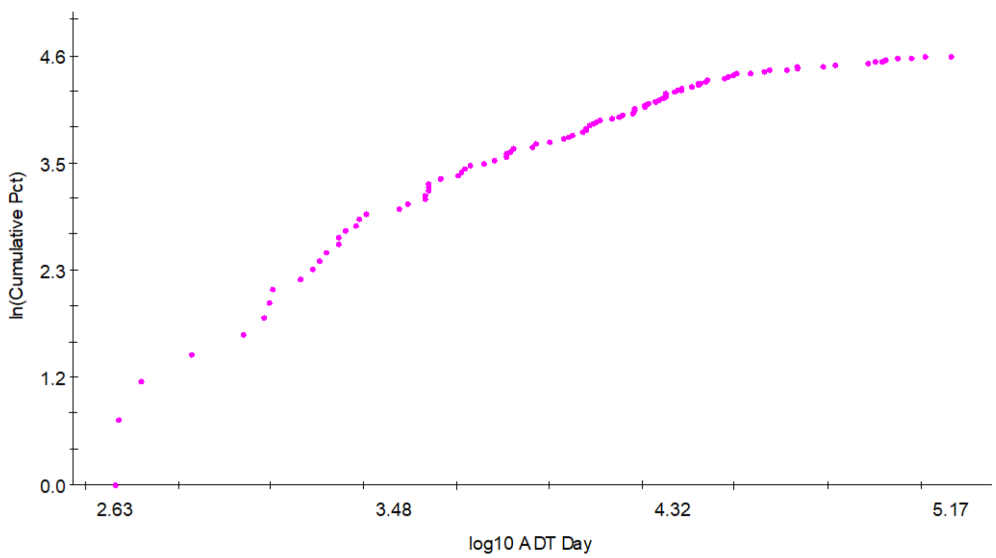
Source: FHWA.

Figure 51. Graph. Frequency distribution plot for ADT day transformed dataset for Utah.



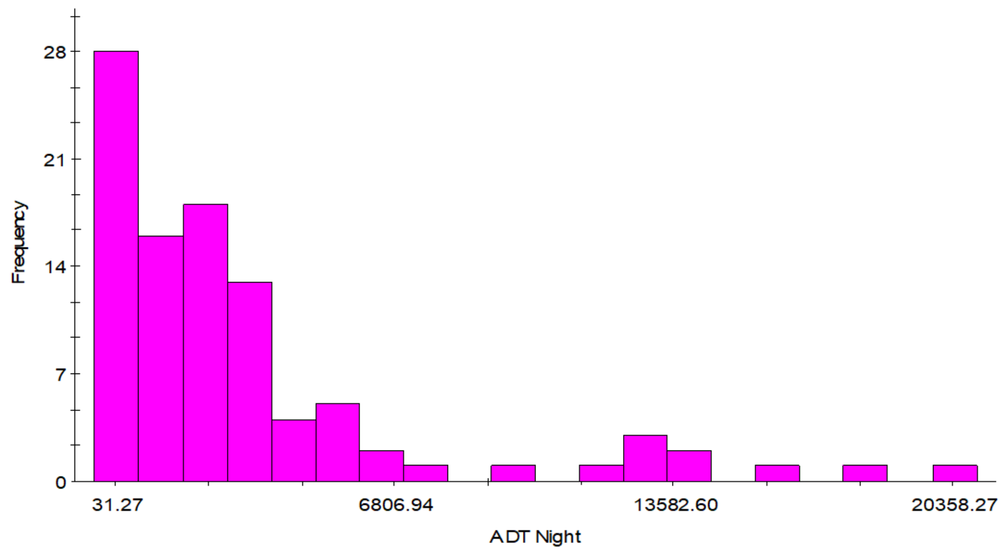
Source: FHWA.

Figure 52. Graph. Normal probability plot for ADT day non-transformed dataset for Utah.



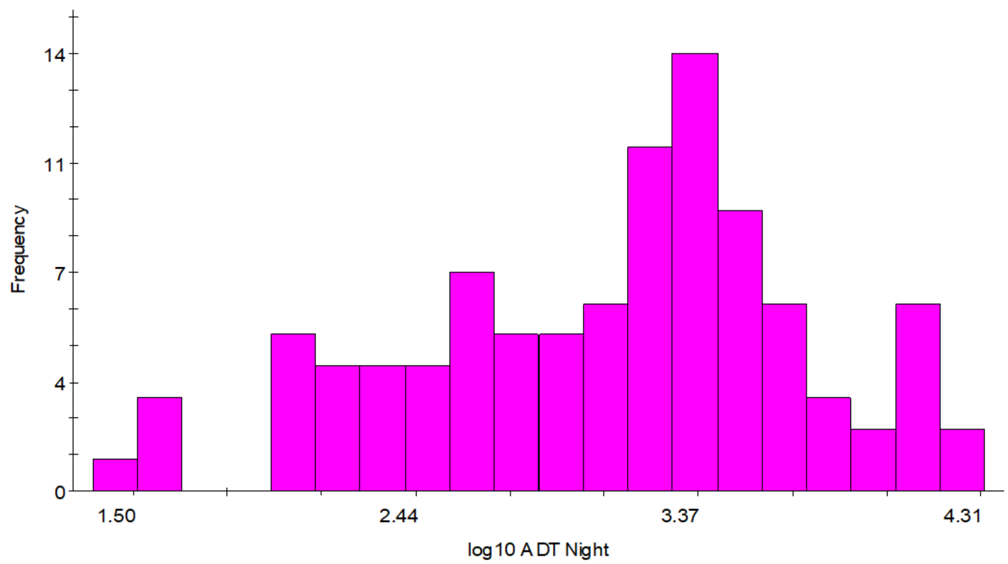
Source: FHWA.

Figure 53. Graph. Normal probability plot for ADT day transformed dataset for Utah.



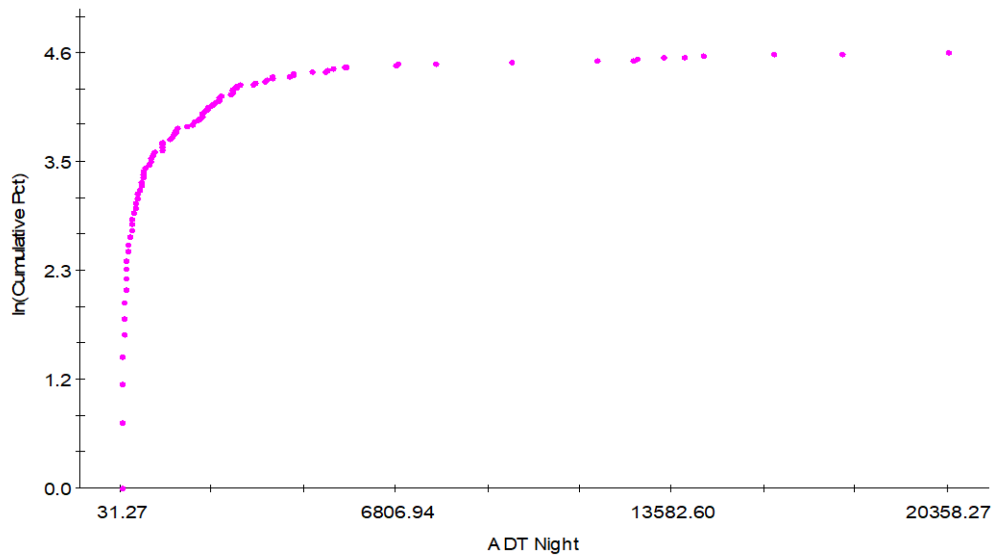
Source: FHWA.

Figure 54. Graph. Frequency distribution plot for ADT night non-transformed dataset for Utah.



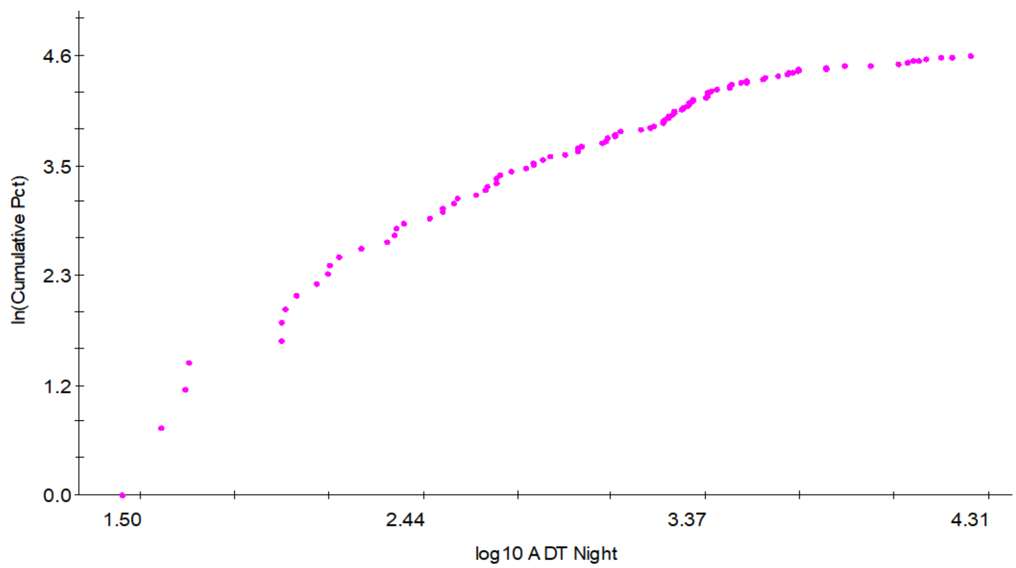
Source: FHWA.

Figure 55. Graph. Frequency distribution plot for ADT night transformed dataset for Utah.



Source: FHWA.

Figure 56. Graph. Normal probability plot for ADT night non-transformed dataset for Utah.



Source: FHWA.

Figure 57. Graph. Normal probability plot for ADT night transformed dataset for Utah.

Table 81. Statistics for the tests done to verify normal distribution hypothesis for the day and night traffic volume datasets—North Carolina.

Statistics	Day, Non-Transformed	Day, Transformed	Night, Non-Transformed	Night, Transformed
Skewness	1.41	–0.33	2.44	–0.19
Kurtosis	1.66	–0.83	8.27	–1.01
Shapiro–Wilk test W/V/*p*-value	11.023	2.272	16.781	2.383
Shapiro–Wilk test W/V/*p*-value	0.000	0.037	0.000	0.029
Shapiro–Wilk test W/V/*p*-value	0.822	0.969	0.722	0.969
Shapiro–Francia test W'/V'/*p*-value	12.214	2.079	19.146	2.156
Shapiro–Francia test W'/V'/*p*-value	0.000	0.078	0.000	0.068
Shapiro–Francia test W'/V'/*p*-value	0.823	0.963	0.730	0.962

Table 82. Statistics for the tests done to verify normal distribution hypothesis for the day and night traffic volume datasets—Utah.

Statistics	Day, Non-Transformed	Day, Transformed	Night, Non-Transformed	Night, Transformed
Skewness	2.14	–0.31	2.18	–0.44
Kurtosis	4.34	–0.46	4.34	–0.35
Shapiro–Wilk test W/V/*p*-value	23.716	1.622	24.896	2.324
Shapiro–Wilk test W/V/*p*-value	0.000	0.141	0.000	0.030
Shapiro–Wilk test W/V/*p*-value	0.710	0.984	0.696	0.975
Shapiro–Francia test W'/V'/*p*-value	26.364	1.416	27.630	2.213
Shapiro–Francia test W'/V'/*p*-value	0.000	0.245	0.000	0.058
Shapiro–Francia test W'/V'/*p*-value	0.712	0.980	0.698	0.971

Variogram Modeling and Kriging Interpolation: Utah

A variogram is generally more useful than a covariance function and it has become a key tool in geostatistics. If h is the separation between the samples of data (i.e., ATR counts in our context) in both distance and direction, $Z(x)$ and $Z(x + h)$ are the values of Z at places x, $x + h$, and E denotes the expectation, then the semivariance is given by the equation in figure 58

$$\gamma(h) = \frac{1}{2} var[Z(x) - Z(x+h)] = \frac{1}{2} E[\{Z(x) - Z(x+h)\}^2]$$

Figure 58. Equation. Semivariance as a function of distance.

Like covariance, semivariance depends only on *h*, and as a function of *h*, it is the variogram $\gamma(h)$. In other words, the variogram is a graph of semivariance versus separation distance. Where autocorrelation is present, semivariance is lower at smaller separation distances (autocorrelation is greater). There are two types of variograms that should be considered. Initially, an experimental variogram is estimated from observed data, $Z(x_i)$, where $i = 1,2,\ldots n$. It is usually computed by the method of moments as shown in figure 59.(60)

$$\gamma(h) = \frac{1}{2m(h)} \sum_{j=1}^{m(h)} \{z(x_j) - z(x_j + h)\}^2$$

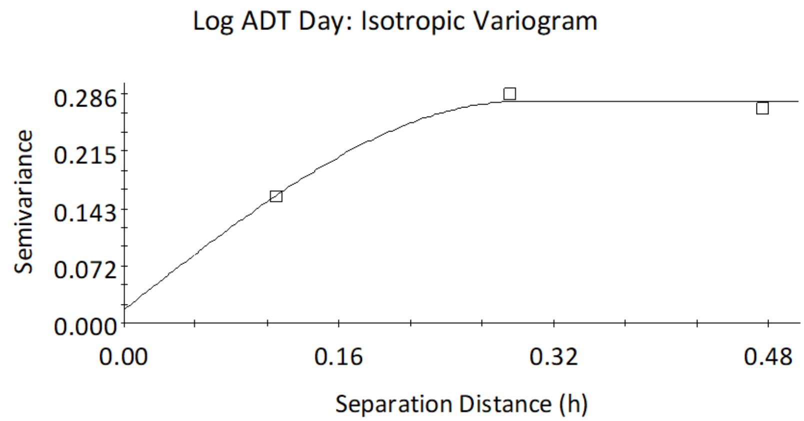
Figure 59. Equation. Computation of experimental variogram.

Where $m(h)$ is the number of paired comparisons at lag *h*.

This is known as an experimental variogram, which is obtained by the ordered set of values by incrementing the lag size in steps. A functional variogram model is then fitted to the experimental variogram. If all the plausible functional variogram models seem to fit well, then the one with the smallest residual sum of squares (*RSS*) or smallest mean square is selected. Once a proper variogram model is selected for the observed data, kriging is employed for the generation of interpolated surfaces and the estimation of the corresponding kriging error.

The variogram (graph) typically yields a curve, which can be modeled using three terms—a nugget variance (C_0), a sill ($C_0 + C$), and a range (*A*). In the variogram, the model levels out at a certain distance. This distance where the model first flattens out is known as the range. Sample points or locations separated by distances closer than the range are spatially autocorrelated, whereas locations farther apart than the range are uncorrelated. The value that the variogram attains at the range on the Y-axis is called the sill. The partial sill is the sill minus the nugget.

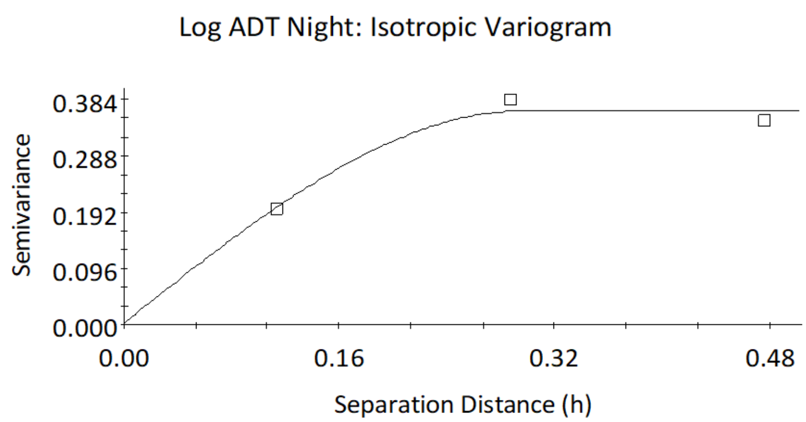
Geostatistical software, GS+™, is used to plot, fit, and verify the accuracy of the experimental variogram models. Three of the most common types of isotropic variogram models are provided in available software used for variogram modeling: Spherical, Exponential, and Gaussian. The best fitted variogram model is the one with the smallest RSS and the largest percent of variance explained. Experimental variograms are computed with the three robust estimators discussed above, and Matheron's method of moments for the observed data of traffic volumes with the skewness coefficients close to zero. Figure 60 shows the experimental variograms that fit the observed data well for the transformed day and night traffic volume data in the State of Utah. Table 83 notes the parameters of models fitted to the experimental variograms of the day and night transformed traffic volume data for Utah. For the analysis, the variation is considered to be isotropic.



Spherical model (Co = 0.01620; Co + C = 0.27740; Ao = 0.30; r2 = 0.985; RSS = 1.509E-04)

Source: FHWA.

Figure 60. Graph. Experimental variograms computed for transformed day traffic volume data.



Spherical model (Co = 0.00300; Co + C = 0.36500; Ao = 0.30; r2 = 0.965; RSS = 6.991E-04)

Source: FHWA.

Figure 61. Graph. Experimental variograms computed for transformed night traffic volume data.

Table 83. Parameters of models fitted to the experimental variograms of log-transformed day and night traffic volume data in Utah.

	Estimates of Parameters			Diagnostics	
Dataset/*Model*	C_0	$C_0 + C$	A	***RSS***	**Variance explained (%)**
Day traffic volume data/ *spherical*	0.0162	0.2774	0.295	1.509E–04	98.5
Night traffic volume data/ *spherical*	0.003	0.365	0.297	6.991E–04	96.5

Kriging is a generic term for a range of least squares methods that provide best linear unbiased predictions, in the sense of minimum variance. Kriging provides a means of estimating unknown variable values at unsampled locations in space (in our context, day and night volumes at locations without ATRs), where no measurements are available based on known sampling values at the surrounding locations (in our context, day and night volumes measured at locations with ATRs).[60–62] An ordinary kriging technique from the family of geostatistical methods was used in this study to interpolate day and night traffic volume data and estimate kriging error. This technique is frequently used. It requires knowledge of the variogram function and data for its implementation. The kriging estimator is expressed as show in figure 62.

$$Z(x_0) = \sum_{i=1}^{N} w_i z(x_i)$$

Figure 62. Equation. Kriging estimator.

where $z(x_i)$, $i = 1,2,\ldots,N$ are observed values of variable z at points $x_1, x_2,\ldots, x_N$, $Z(x_0)$ is an estimated value of Z at desired location, x_0; w_i represents weights associated with the observation at the location x_i with respect to x_0, and N indicates the number of observations within the domain of the search neighborhood of x_0 used for performing the estimation of $Z(x_0)$.

The kriging variance, ${\sigma^2}_z(x_0)$ in the ordinary kriging can be computed with figure 63.

$${\sigma^2}_z(x_0) = \mu_z + \sum_{i=1}^{N} w_i \gamma(h_{0i}) \qquad for \quad \sum_{i=1}^{N} w_i = 1$$

Figure 63. Equation. Kriging variance in ordinary kriging.

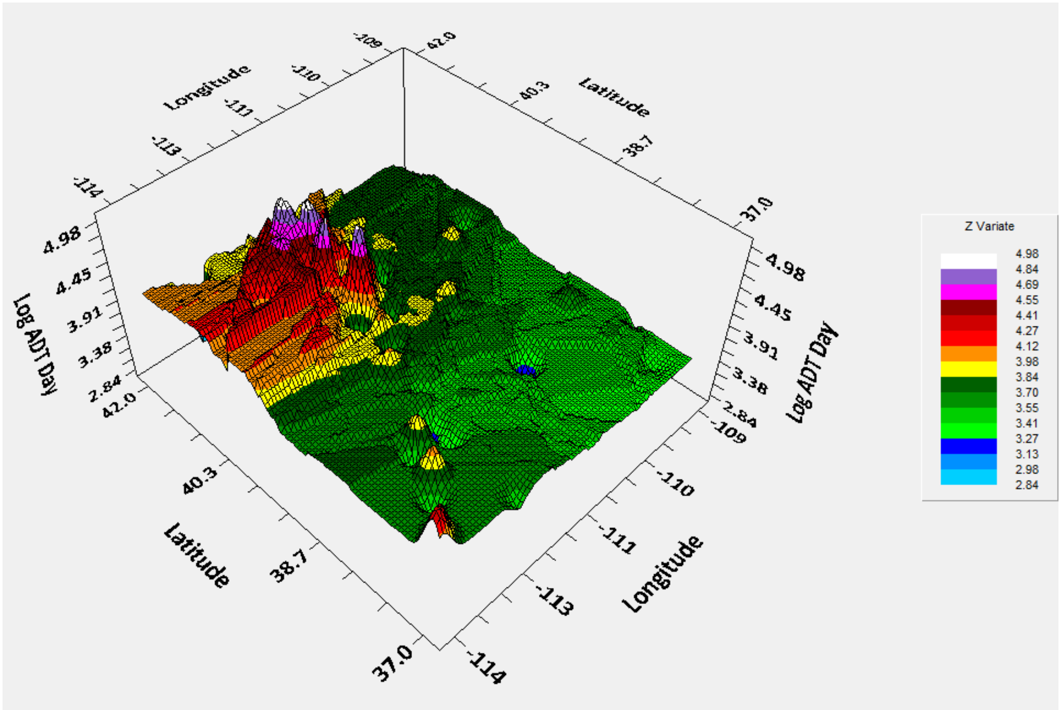
where $\gamma(h)$ is the variogram value for the distance h, h_{0i} is the distance between observed data points x_1, and x_2, μ_z is the Lagrangian multiplier in the Z scale, h_{0j} is the distance between unobserved data point x_0 and observed data point x_i and N is the number of sample locations. When a log transformation is applied to the data, ordinary kriging is converted into log–normal kriging. The log-transformed predicted values obtained in the log–normal kriging are then back-transformed to their original states. It is assumed that these back-transformed values are an unbiased predictor of the kriging interpolation. Kriging interpolation maps for day and night traffic volumes in Utah are shown in figure 64 and figure 65.

Performance Assessment of Kriging Results

A performance assessment of the final form of the variogram model and kriging interpolations was executed using a validation scheme known as a cross-validation procedure. In cross-validation analysis provided by GS+™, each measured point in the spatial domain (in our case, ATR measurements) is individually removed from the domain and its value is estimated using an estimated variogram model and kriging interpolation as though it were never there. Then, the point is replaced and the next point is removed and estimated. This procedure continues for all data points. This validation scheme is also known as "leave-one-out cross-validation," and it helps to evaluate the prediction performance of kriging by comparing estimated and actual values. In addition to the leave-one-out cross-validation, K-fold cross-validation involves removing 10 percent of the sample from the dataset, and then predicting the values at those locations using information from the remaining observations. This process is repeated with multiple “10 percent subsamples” until all the available data points are used at least once in a subsample. Measured and predicted values are compared. A comparison of the average difference between predicted and observed values is made through the leave-one-out cross-validation method. Cross-validation statistics serve as diagnostics to demonstrate whether the performance of the adopted model is acceptable. The statistics are used to check whether the prediction is unbiased and as close as possible to the measured values. The variability of the predictions is also assessed.

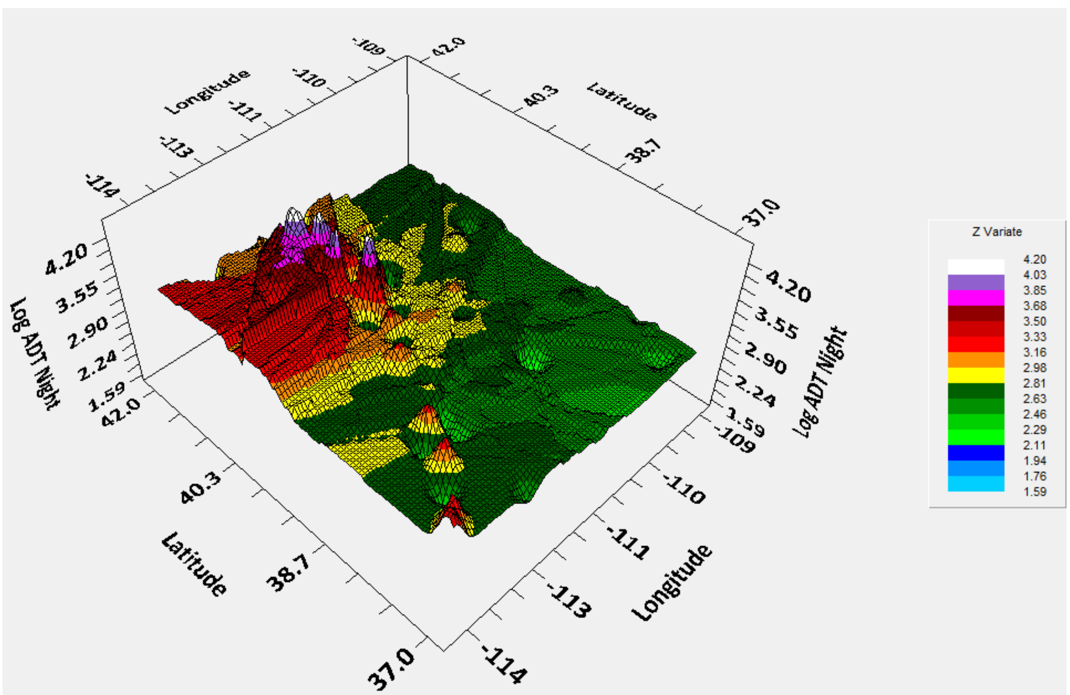
When the sample size is relatively small, other cross-validation techniques, such as random subsampling, may be more effective. These techniques depend on the spatial and/or temporal resolution and correlation in the data. For example, if the randomly selected sites for cross-validation are close to other stations, they will likely have high correlations with those nearby stations, and the prediction model would appear to appear to work quite well (and possibly be overstated). Alternatively, the predictions may perform poorly at other locations selected for validation that are farther from any remaining data points. In both cases, the estimates obtained from the kriging methodology and the resulting prediction error may be a function of the random sampling procedures, which is undesirable.

This study used K-fold with random subsampling (to address the small sample sizes) and leave-one-out cross-validation techniques to compute the average standard error in model predictions. The diagnostics and the scatter diagrams from the cross-validation procedures are provided in table 84, figure 66, and figure 67. The predictions matched observed values from a directional and “order of magnitude” perspective (e.g., regression coefficient between observed and predicted is positive and close to unity), but there was significant “noise” in model predictions (R^2 values around 0.3). Future directions in improving these models are discussed at the end of this section.



Source: FHWA.

Figure 64. Graphic. Kriging interpolation maps for transformed day traffic volume data for Utah.

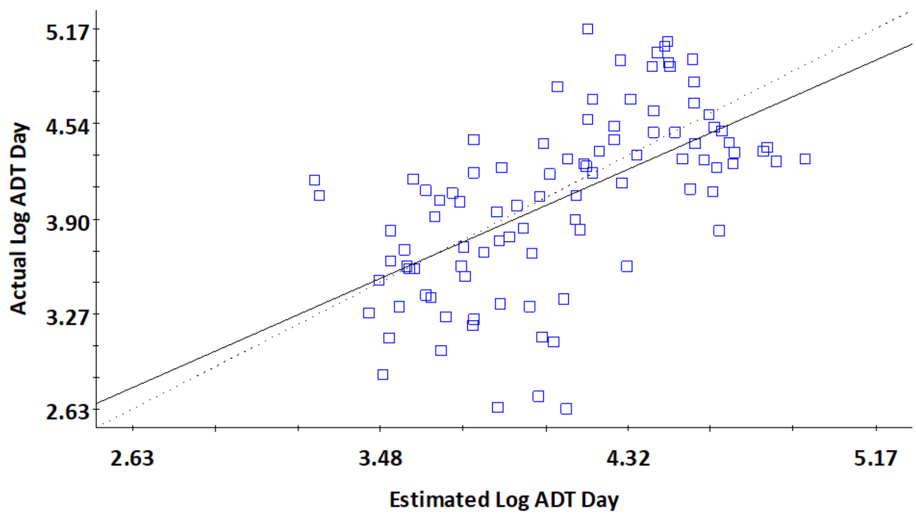


Source: FHWA.

Figure 65. Graphic. Kriging interpolation maps for transformed night traffic volume data for Utah.

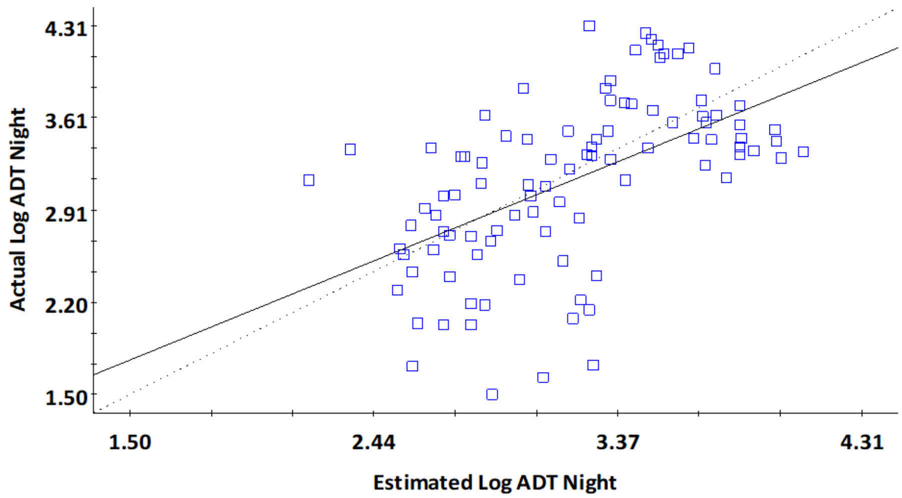
Table 84. Diagnostics from cross-validation of kriging interpolation done for the day and night traffic volume data (leave-one-out cross-validation: Utah).

Dataset/ *Model*	Regression Coefficient	R^2	y-Intercept	Standard Error	Standard Error Prediction
Day traffic volume data/ *spherical*	0.862	0.347	0.51	0.121	0.479
Night traffic volume data/ *spherical*	0.808	0.297	0.55	0.128	0.543



Source: FHWA.

Figure 66. Chart. Scatter diagrams of observed values of day traffic volume data at ATR stations plotted against values predicted by ordinary kriging in GS+™ during cross-validation.



Source: FHWA.

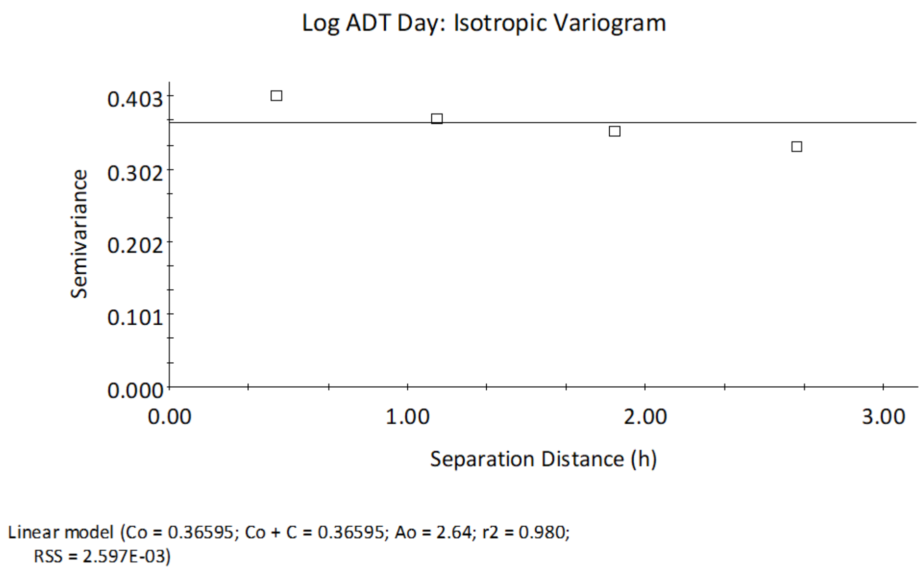
Figure 67. Chart. Scatter diagrams of observed values of night traffic volume data at ATR stations plotted against values predicted by ordinary kriging in GS+™ during cross-validation.

Variogram Modeling and Kriging Interpolation: North Carolina

Traffic volume data were requested for all ATR locations throughout the State of North Carolina. The project team ultimately obtained traffic data for 72 ATR stations throughout the State. They used these data to explore spatial prediction and analysis techniques for North Carolina that were the same as those the project team used in Utah. As already outlined in previous sections, the basic tool of geostatistics and kriging is the variogram. The variogram captures the spatial dependence between the samples (data points), ATR stations in our case, by plotting semivariance against separation distance. The basic premise of any spatial interpolation is that close samples/points tend to be more similar than distant samples/points. This is also known as spatial autocorrelation. In kriging, the spatial autocorrelation is modeled using a variogram instead of assuming a direct, linear relationship with separation distance.

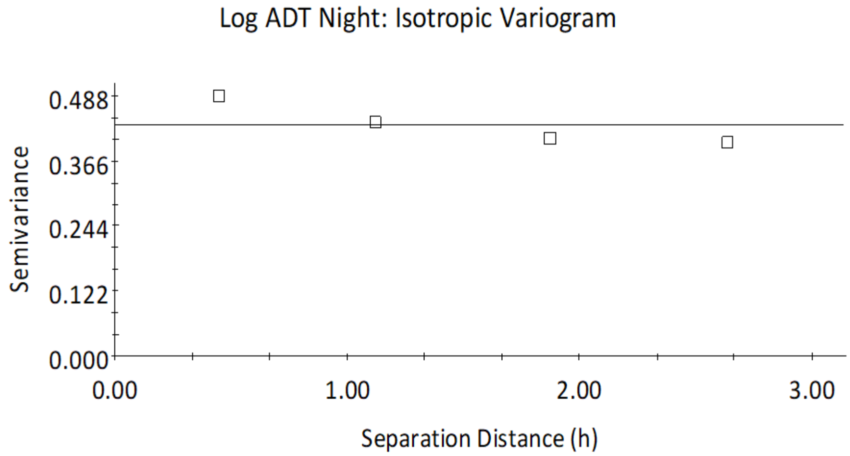
Semivariance equals one-half the squared difference between points separated by a distance, with no direction preferences. As the distance between the samples increases, the semivariance increases, again because the project team assumed the nearby data points to be more related than distant points. This is true, however, only up to a given separation distance. Beyond this distance, points are essentially considered unrelated (i.e., high values for semivariance) and the level by which they are unrelated does not change as distance continues to increase. For example, if 10 mi is the critical separation distance, two points separated by 10 mi are likely to be just as similar (or not similar) as points than are separated by 30, 50, 100 mi, or any distance greater than 10 mi. In other words, spatial autocorrelation exists only for pairs of points separated by less than the range value. The more quickly variogram rises from the origin to the sill, the more quickly autocorrelation declines.

For the North Carolina data, the project team implemented variogram modeling in the same way as for the Utah data. They used the data for the 72 ATR recorder stations, and the respective latitude and longitude coordinates, for the spatial prediction. Results of the variogram modeling for day and night traffic volumes are shown in figure 68 and table 85. These variograms show a straight line. In other words, as the separation distance increases from zero, the semivariance of pairs of data points remains the same. Data from the spatially separated ATRs in North Carolina are uncorrelated (i.e., there is no spatial autocorrelation that exists for pairs of points, regardless of separation distance). There are multiple possible reasons for this finding. There may not be enough samples (i.e., ATRs) with varying separation distances in the North Carolina data to estimate the exact range and sill values for the data. As a result, the project team observed no autocorrelation, even between points that are close to each other. Another reason may be the fewer number of ATR stations that were available for the analysis in general, compared to the large geographic area of North Carolina. Without a variogram, it is not possible to apply spatial prediction and interpolation techniques for the rural unsignalized intersections in North Carolina.



Source: FHWA.

Figure 68. Graph. Experimental variograms computed for transformed day traffic volume data.



Linear model (Co = 0.43415; Co + C = 0.43415; Ao = 2.64; r2 = 0.882; RSS = 4.649E-03)

Source: FHWA.

Figure 69. Graph. Experimental variograms computed for transformed night traffic volume data.

Table 85. Parameters of models fitted to the experimental variograms of log-transformed day and night traffic volume data in North Carolina.

Dataset/*Model*	Estimates of Parameters			Diagnostics	
	C_0	$C_0 + C$	A	RSS	Variance Explained (%)
Day traffic volume data/ *linear*	0.3659	0.3659	2.6379	2.59E–03	98.0
Night traffic volume data/ *linear*	0.4341	0.4341	2.6379	4.64E–03	88.2

Exploratory Impacts of Day/Night Volume Estimates on Safety Modeling Results

Given the successful attempt at developing a variogram and kriging model to estimate day and night traffic volumes at locations without ATRs in Utah, the project team explored the impact of this new information on statistical road safety modeling. The same dataset consisting of horizontal curves along rural two-lane highways in Utah described in previous chapters of this report was used for safety model estimation, but with new information on day and night traffic volumes. The hypothesis was that sites with the same geometric characteristics would differ in their safety performance if they differ in their distributions of day and night traffic volumes. Specifically, the project team expected that horizontal curves with higher proportions of traffic at night would experience more crashes than similar curves with higher proportions of traffic during the day. A negative binomial regression model, with the expected number of total crashes (i.e., all types and severities) on the left-hand side and selected variables on the right-hand side that included AADT and the predicted night-to-day ratio from the kriging map, was used to test this hypothesis at a preliminary level. Model estimation results are provided in table 86.

Table 86. Estimated model parameters for all total crashes (all types and severities) with day/night volumes.

Variable List	Estimated Coefficient	Standard Error	*p*-Value
Log AADT	0.901	0.042	0.000
Ratio of predicted night/day volume	0.451	2.448	0.854
Degree of curve	0.062	0.015	0.000
Number of 0.5-inch precipitation days	0.030	0.008	0.000
Constant	–7.452	0.430	0.000
Segment length	1 (offset)	1 (offset)	1 (offset)
Log-likelihood	–1,644.2918	–1,644.2918	–1,644.2918
Dispersion parameter	0.6617	0.6617	0.6617

The positive coefficient for the AADT (logaadt) in table 86 suggests that as the traffic volume increases on the roadway, the expected number of crashes also increases, but at a nonlinear rate. The parameter estimate of less than unity is consistent with previous work on rural two-lane highways. The coefficient for the ratio of night to day traffic volume is positive, with an estimated value of 0.451. This variable is not statistically significant, but is included in the models to give a perspective of the magnitude of the relationship between the expected number of crashes and the distribution of traffic volumes by day and night hours throughout the year. The positive coefficient verifies the earlier hypothesis that horizontal curves with higher proportions of traffic at night are expected to experience more crashes than similar curves with higher proportions of traffic during the day. Other coefficients are also in the direction expected. A positive coefficient for the degree of curve (deg_curv) suggests that as the radius of curve increases, the expected number of crashes will decrease. The number of days with at least 0.5 inch of precipitation (NdaysPre05) throughout the year could represent more exposure to adverse weather conditions. This variable has a positive coefficient, which suggests that as precipitation increases, the total number of crashes is expected to increase.

In conclusion, the spatial modeling approach implemented in this study, which estimates average annual day and night traffic volumes in rural locations where permanent counters are not available shows promise. The approach was successful in Utah and unsuccessful in North Carolina, with the latter likely due to inadequate ATR coverage. The hypothesis that horizontal curves with higher proportions of traffic at night are expected to experience more crashes than similar curves with higher proportions of traffic during the day was verified with a positive parameter for night-to-day volume ratio in a negative binomial regression model of total crash frequency. The parameter estimate, however, was noisy and statistically insignificant, most likely attributable to the errors in the kriging predictions. Additional modifications and extensions to the kriging and safety modeling approaches offered here could potentially improve the kriging predictions, and therefore reduce the standard error associated with the predicted night and day volumes. These include the following:

- Incorporating variables in addition to spatial proximity (e.g., functional classification, surrounding area characteristics) into the kriging model.

- Predicting night-to-day ratios directly, instead of night and day traffic volumes.

- Developing separate models for day and night crashes as a function of day and night volumes, respectively.

STUDY 2: IMPACTS OF QUASI-INDUCED DEMAND ESTIMATES AND SOCIODEMOGRAPHICS ON SAFETY ESTIMATION FOR RURAL UNSIGNALIZED INTERSECTIONS IN NORTH CAROLINA

As noted in the previous study, transportation safety researchers have traditionally used the AADT as the measure of exposure for SPFs or crash prediction models. These models predict the expected number of crashes per year (or for a given number of years) for a segment or location based on a single value of AADT. However, it is common knowledge that traffic volumes vary by time of day and day of week. Additionally, the proportion of drivers within an age group varies by these factors. Estimates of these variabilities in traffic and drivers are difficult to obtain, and therefore safety models are unable to incorporate these more disaggregated measures of exposure even though it is very likely that these characteristics of a site influence safety.

Kirk and Stamatiadis described the quasi-induced demand approach.(63) The approach uses data on crash history to estimate exposure from "not-at-fault drivers" in multi-vehicle crashes. The premise is that not-at-fault drivers are a random representation of all drivers on the roadway and their characteristics do not present a bias. At-fault drivers tend to be over-represented in terms of certain age groups and physical characteristics, and are not representative of true exposure. Kirk and Stamatiadis used distributed trip diaries to validate quasi-induced demand data identified from crash data and found that the method to provided promising results.(63) The authors recommended further validation of the method. This analysis attempts to validate quasi-induced demand in North Carolina using not-at-fault driver data from multi-vehicle crashes and time of day distributions validated through analysis of ATR data.

Table 87 provides a summary of crash frequency by age group by year based on select inclusion criteria from data that the HSIS provided for crashes in North Carolina. The observations in each cell in column 3 include data for each vehicle involved in a North Carolina crash between 2008 and 2012. The data in column 4 include only vehicles involved in multi-vehicle crashes, meaning that the project team excluded vehicles from single-vehicle crashes. The difference between columns 3 and 4 indicates that younger drivers are more over-represented in single-vehicle crashes (i.e., at-fault drivers tend to be younger). In general, the proportion of drivers in an age group goes down for multi-vehicle crashes for the group less than 41 years old, and goes up for the group greater than 40 years old. This is true for the oldest category, greater than 64 years old, as well. However, single-vehicle crashes generally include only at-fault drivers, while multi-vehicle crashes include at-fault and not-at-fault drivers, as there are two or more vehicles involved.

Due to the nature of the data, the at-fault status of drivers can only be determined by police-reported contributing factors provided in the vehicle file. There are 35 contributing factor codes, with code 0 indicating no contributing factors for the vehicle in question. In most cases, one driver in the vehicle file had at least one contributing factor (multiple contributing factors were common) while the other driver or drivers were not coded with a contributing factor. Instances where both drivers had contributing factors were common in the dataset. Further analysis indicated no instances where neither driver in multi-vehicle crashes had a contributing factor. Therefore, the researchers included only vehicles where the driver had no contributing factors in the not-at-fault category. Across the 5 years of data, this resulted in a final dataset of 73,758 vehicles for not-at-fault drivers. When considering missing data, such as driver age, the final analysis includes 72,412 vehicles. The project team excluded the 1,346 vehicles with missing information from the analysis; they assumed that the information was missing at random and the exclusion would not introduce bias into the exposure estimates.

Table 87. Data distributions by inclusion criteria.

Year	Age Group	All Crashes	All MV Crashes	MV Not at Fault Crashes
2008	24 or less	26.4	24.3	19.1
2008	25–40	29.1	27.8	30.4
2008	41–64	31.4	32.5	37.9
2008	65 or more	13.0	15.4	12.6
2008	Unknown	5.8	6.2	4.4
2009	24 or less	26.6	24.3	18.8
2009	25–40	29.2	27.6	29.9
2009	41–64	31.6	32.4	38.2
2009	65 or more	12.6	15.7	13.1
2009	Unknown	5.3	6.0	4.4
2010	24 or less	26.7	25.6	19.4
2010	25–40	30.8	28.5	29.9
2010	41–64	34.2	35.3	41.2
2010	65 or more	8.2	10.6	9.5
2010	Unknown	0.0	0.0	0.0
2011	24 or less	26.7	25.4	19.1
2011	25–40	30.6	28.5	30.0
2011	41–64	34.4	35.3	40.8
2011	65 or more	8.4	10.6	10.1
2011	Unknown	0.0	0.0	0.0
2012	24 or less	26.5	25.0	19.1
2012	25–40	30.1	28.2	29.3
2012	41–64	34.5	35.2	41.0
2012	65 or more	8.9	11.6	10.6
2012	Unknown	0.0	0.0	0.0

Initially, the research approach attempted to identify situations where it was possible to identify not-at-fault drivers from crashes where both drivers had contributing factors. This could include situations where the contributing factor for one driver was more egregious than the other driver. For example, if driver one disregarded a traffic signal and driver two was inattentive, driver two may still be not-at-fault, and would have been randomly involved in the crash. After further consideration, the project team could find no systematic solution for accurately defining random not-at-fault drivers. Therefore, the project team retained no drivers with contributing factors in the database.

Additionally, the project team made an effort to verify the key assumption of the quasi-induced demand approach by defining situations where not-at-fault drivers may not necessarily be random. For example, if the at-fault driver was noted to be following too closely or was noted to have disregarded a traffic control device, the not-at-fault driver's characteristics may not necessarily be random. A younger or older driver may have slower perception–reaction times and may be less likely to avoid a crash in these instances compared to drivers in the middle age categories. However, there was no evidence to suggest a reason to exclude any crashes with a driver having no contributing factors from the dataset. Additionally, the methodology was consistent with that outlined by Kirk and Stamatiadis (2001).[(63)] In their analysis, the authors removed crashes only where both drivers were coded as having at-fault characteristics or neither driver was coded as having an at-fault characteristic.

Column 5 of table 87 includes only vehicles that have no contributing factors from multi-vehicle crashes (i.e., not-at-fault vehicles). The difference between column 4 and 5 indicates that both older and younger drivers are over-represented in crash data that include at-fault drivers. The distributions in column 5 identify the approximate percentages of the driving population by age group (i.e., the values for quasi-induced demand). Table 88 and table 89 provide quasi-induced demand for all North Carolina counties based on frequency and percentages, respectively. The data provide breakdowns by age, roadway class, time of day, day of week, and daylight conditions. The age groups the project team used, as well as roadway class, time of day, and day of week, are consistent with the groupings Kirk and Stamatiadis used.[(63)] The age distribution in table 89 is very similar to the diary exposure presented by Kirk and Stamatiadis (all percentages are within 2 to 3 percent for their group). Table 89 indicates a higher nighttime exposure (18.2 versus 13.6 percent), but is relatively consistent with the crash-based quasi-induced exposure reported by Kirk and Stamatiadis. However, it was unclear if the authors included dusk, dawn, and the category "other" in the nighttime crash totals.

The crash data used in table 87 through table 89 represent all crashes occurring in North Carolina on rural two-lane highways. The 42 intersections with minor road stop control used in this study represent only a few counties scattered across the Piedmont and northern Sandhills regions. There is a logical concern that statewide data may not represent the trends occurring at study intersections, and the crash database was further limited to the study-site counties. The study sites included data in Franklin, Granville, Harnett, Johnston, Moore, Transylvania, and Wake Counties. Table 90 and provide the quasi-induced demand data for only study-site counties. The resulting percentages indicate that statewide data represented trends at the study-site counties quite well. Therefore, statewide data are further explored in table 92 and table 93. Table 92 and table 93 further break the data from three age categories to four age categories. Table 93 provides the final percentages that can be used to subdivide AADT for analysis. It is possible to cross-examine the hourly data (in 6-hour increments) with the ATR data to assess the validity of the methodology.

Table 88. All North Carolina counties induced demand.

Category	Values	Ages <34	Ages 35–64	Ages 65+	Total
Age	—	27,511	37,983	6,918	72,412
Roadway class	Principal arterial	1,574	2,268	447	4,289
Roadway class	Minor arterial	4,769	7,114	1,342	13,225
Roadway class	Major collector	10,854	15,542	2,890	29,286
Roadway class	Minor collector	4,184	5,360	875	10,419
Roadway class	Local	6,045	7,524	1,339	14,908
Time of day	Daylight	21,822	31,358	6,072	59,252
Time of day	Nighttime	5,689	6,625	846	13,160
Day of week	Weekday	21,512	30,218	5,436	57,166
Day of week	Weekend	5,999	7,765	1,482	15,246
Hour	12:00 a.m. to 6:00 a.m.	698	669	52	1,419
Hour	6:00 a.m. to 12:00 p.m.	7,463	11,939	2,185	21,587
Hour	12:00 p.m. to 6:00 p.m.	13,764	19,050	3,814	36,628
Hour	6:00 p.m. to 12:00 a.m.	5,586	6,325	867	12,778

—Not applicable.

Table 89. All North Carolina counties induced demand percentage.

Category	Values	Ages <34	Ages 35–64	Ages 65+	Total
Age	—	37.99	52.45	9.55	100.00
Roadway class	Principal arterial	2.17	3.13	0.62	5.92
Roadway class	Minor arterial	6.59	9.82	1.85	18.26
Roadway class	Major collector	14.99	21.46	3.99	40.44
Roadway class	Minor collector	5.78	7.40	1.21	14.39
Roadway class	Local	8.35	10.39	1.85	20.59
Time of day	Daylight	30.14	43.30	8.39	81.83
Time of day	Nighttime	7.86	9.15	1.17	18.17
Day of week	Weekday	29.71	41.73	7.51	78.95
Day of week	Weekend	8.28	10.72	2.05	21.05
Hour	12:00 a.m. to 6:00 a.m.	0.96	0.92	0.07	1.96
Hour	6:00 a.m. to 12:00 p.m.	10.31	16.49	3.02	29.81
Hour	12:00 p.m. to 6:00 p.m.	19.01	26.31	5.27	50.58
Hour	6:00 p.m. to 12:00 a.m.	7.71	8.73	1.20	17.65

—Not applicable.

Table 90. North Carolina study counties induced demand.

Category	Values	Ages <34	Ages 35–64	Ages 65+	Total
Age	—	3,528	4,605	733	8,866
Roadway class	Principal arterial	75	101	20	196
Roadway class	Minor arterial	531	766	138	1,435
Roadway class	Major collector	1,536	2,119	325	3,980
Roadway class	Minor collector	599	682	109	1,390
Roadway class	Local	782	911	141	1,834
Time of day	Daylight	2,720	3,669	628	7,017
Time of day	Nighttime	806	925	104	1,835
Day of week	Weekday	2,748	3,665	567	6,980
Day of week	Weekend	780	940	166	1,886
Hour	12:00 a.m. to 6:00 a.m.	82	81	3	166
Hour	6:00 a.m. to 12:00 p.m.	943	1,409	254	2,606
Hour	12:00 p.m. to 6:00 p.m.	1,711	2,282	371	4,364
Hour	6:00 p.m. to 12:00 a.m.	792	823	105	1,720

—Not applicable.

Table 91. North Carolina study counties induced demand percentage.

Category	Values	Ages <34	Ages 35–64	Ages 65+	Total
Age	—	39.79	51.94	8.27	100.00
Roadway class	Principal arterial	0.85	1.14	0.23	2.22
Roadway class	Minor arterial	6.01	8.67	1.56	16.24
Roadway class	Major collector	17.39	23.98	3.68	45.05
Roadway class	Minor collector	6.78	7.72	1.23	15.73
Roadway class	Local	8.85	10.31	1.60	20.76
Time of day	Daylight	30.73	41.45	7.09	79.27
Time of day	Nighttime	8.68	10.53	2.01	20.73
Day of week	Weekday	30.99	41.34	6.40	78.73
Day of week	Weekend	8.80	10.60	1.87	21.27
Hour	12:00 a.m. to 6:00 a.m.	0.93	0.91	0.03	1.87
Hour	6:00 a.m. to 12:00 p.m.	10.65	15.91	2.87	29.43
Hour	12:00 p.m. to 6:00 p.m.	19.32	25.77	4.19	49.28
Hour	6:00 p.m. to 12:00 a.m.	8.94	9.29	1.19	19.42

—Not applicable.

Table 92. All North Carolina counties induced demand.

Category	Values	Ages <24	Ages 25–44	Ages 44–64	Ages 65+	Total
Age	—	14,091	28,041	23,362	6,918	72,412
Roadway class	Principal arterial	747	1,659	1,436	447	4,289
Roadway class	Minor arterial	2,396	5,023	4,464	1,342	13,225
Roadway class	Major collector	5,422	11,370	9,604	2,892	29,288
Roadway class	Minor collector	2,283	4,081	3,180	875	10,419
Roadway class	Local	3,200	5,792	4,577	1,339	14,908
Time of day	Daylight	11,201	22,451	19,528	6,072	59,252
Time of day	Nighttime	2,886	5,574	3,829	845	13,134
Day of week	Weekday	10,906	22,295	18,529	5,436	57,166
Day of week	Weekend	3,185	5,746	4,833	1,482	15,246
Hour	12:00 a.m. to 6:00 a.m.	368	634	365	52	1,419
Hour	6:00 a.m. to 12:00 p.m.	3,387	8,733	7,282	2,185	21,587
Hour	12:00 p.m. to 6:00 p.m.	7,408	13,374	12,032	3,814	36,628
Hour	6:00 p.m. to 12:00 a.m.	2,928	5,300	3,683	867	12,778

—Not applicable.

Table 93. All North Carolina counties induced demand percentage.

Category	Values	Ages <24	Ages 25–44	Ages 44–64	Ages 65+	Total
Age	—	19.5	38.7	32.3	9.5	100.0
Roadway class	Principal arterial	1.0	2.3	2.0	0.6	5.9
Roadway class	Minor arterial	3.3	7.0	6.2	1.9	18.3
Roadway class	Major collector	7.5	15.8	13.3	4.0	40.6
Roadway class	Minor collector	3.2	5.7	4.4	1.2	14.4
Roadway class	Local	4.4	8.0	6.3	1.9	20.7
Time of day	Daylight	15.5	31.0	27.0	8.4	81.9
Time of day	Nighttime	4.0	7.7	5.3	1.2	18.1
Day of week	Weekday	15.1	30.8	25.6	7.5	78.9
Day of week	Weekend	4.4	7.9	6.7	2.0	21.1
Hour	12:00 a.m. to 6:00 a.m.	0.5	0.9	0.5	0.1	2.0
Hour	6:00 a.m. to 12:00 p.m.	4.7	12.1	10.1	3.0	29.8
Hour	12:00 p.m. to 6:00 p.m.	10.2	18.5	16.6	5.3	50.6
Hour	6:00 p.m. to 12:00 a.m.	4.0	7.3	5.1	1.2	17.6

—Not applicable.

The project team also explored socio-economic data at the tract level and county level for the 42 North Carolina intersections included in the dataset. The American FactFinder website houses data from the Census Bureau, obtained from nearly 100 annual surveys and censuses. The website contains non-identifying information from the Decennial Census, American Community Survey, American Housing Survey, and three Annual Economic Surveys, among others.

A focused set of data were obtained at the tract level and county level, including the following:

- Total population.
- Total households.
- Household income.
- Percentage male.

Using the GPS coordinates of the intersection locations, the project team recorded the tract-level and county-level data for each intersection. Table 94 provides a summary of the socioeconomic data. The summary data indicated a wide range, even within the small sample of counties and tracts with study intersections.

Table 94. Socioeconomic summary data.

Level	Element	Minimum	Maximum	Mean	Standard Deviation
Tract	Total population	2,122	15,146	5,868	2,680
Tract	Percent male	42.1	55.8	48.6	0.03
Tract	Total households	66	501	145	73
Tract	Household income	22,372	107,292	50,374	18,135
County	Total population	32,825	905,573	285,709	330,210
County	Percent male	47.8	53.2	49.2	0.01
County	Total households	959	22,485	7,606	7,962
County	Household income	40,678	65,826	50,443	8,651

Initially, the intent was to use socioeconomic data to provide surrogate information for typically unobserved characteristics related to crashes. However, the project team considered socioeconomic data to have the potential to act as a surrogate for minor road AADT. Minor road AADTs are not commonly known, and these roads may not necessarily be low-volume roadways. Therefore, the project team predicted the minor road AADTs based on socioeconomic data to determine if it was possible to develop an adequate model. Table 95 provides the best model for predicting minor road AADT. The minor road AADT was considered as a continuous variable using ordinary least squares regression as well as count models (i.e., Poisson and negative binomial). The results indicate that the tract median household income is marginally associated with minor road AADT; however, the model itself is not a good fit and is not significant. Surprisingly, total tract households and tract population were not predictors of minor road AADT. Therefore, socioeconomic data were not pursued further for predicting minor road AADT.

Table 95. Model for minor road AADT as a function of household income (in thousands of dollars).

Variable	*b*	Standard Error	*t*	*p*	L95	U95
Household income (in $1,000)	8.99	5.18	1.73	0.091	–1.49	19.46
Constant	654.57	277.17	2.36	0.023	94.39	1,214.74

Note: $N = 42$; $F_{(1,40)} = 3.00$; $P(F > 3.00) = 0.091$; $R^2_{adj} = 0.047$.

The correlation matrix in table 96 provides the relationship between tract and county level data for total population, percentage male, total households, and household income with major and minor road AADT and crash outcomes. The matrix, which presents time-of-day-level data, shows a strong positive relationship between AADT values and crash outcomes, and a weak relationship between socioeconomic data and crash outcomes. Relationships between socioeconomic data and AADT variables are also very weak; however, on average, relationships tend to be stronger between socioeconomic data and minor road AADT. Regression models between socio-economic variables and minor road AADT indicate no relationship (i.e., the overall model is not significant). Further analysis indicates no socioeconomic values are associated with crash outcomes. The project team discontinued further exploration of socioeconomic data.

Table 96. Correlation matrix for socioeconomic data.

Variable	MV	Right	Major AADT	Minor AADT	TPop	TMale	THH	TInc	CPop	CMale	CHH	CInc
MV	1.00	--	--	--	--	--	--	--	--	--	--	--
Right	0.83	1.00	--	--	--	--	--	--	--	--	--	--
MajAADT	0.65	0.38	1.00	--	--	--	--	--	--	--	--	--
MinAADT	0.60	0.43	0.75	1.00	--	--	--	--	--	--	--	--
TPop	–0.10	–0.06	–0.02	0.03	1.00	--	--	--	--	--	--	--
TMale	–0.17	–0.11	–0.17	–0.06	0.33	1.00	--	--	--	--	--	--
THH	–0.15	–0.15	0.04	–0.05	0.40	0.05	1.00	--	--	--	--	--
TInc	–0.00	–0.11	0.07	0.15	0.19	0.07	–0.13	1.00	--	--	--	--
CPop	0.01	–0.13	0.04	0.15	0.17	0.07	–0.04	0.68	1.00	--	--	--
CMale	0.10	0.15	0.02	0.03	–0.01	0.17	–0.07	–0.17	–0.38	1.00	--	--
CHH	0.01	–0.12	0.03	0.15	0.13	0.02	–0.02	0.69	0.99	–0.41	1.00	--
CInc	–0.02	–0.16	0.03	0.14	0.23	0.15	–0.07	0.69	0.97	–0.29	0.95	1.00

--No data.

The 42 North Carolina intersections used to evaluate the AADT were distributed by the quasi-induced demand. Table 97 and table 98 present the data elements considered in the analysis, as well as summary statistics. Table 97 presents the continuous data elements and table 98 presents the categorical data elements. Additionally, the presence of turn lanes was considered; however, there was only one major road right-turn lane and one minor road left-turn lane present at the study intersections. The data in table 97 and table 98 are average values for variables that change across time. Table 99 presents summary statistics for annual data. Comparing table 97 and table 98 shows that the mean value of continuous variables does not change due to aggregation; however, the standard deviation for the variables does change.

Table 97. Summary data for continuous variables.

Variable	Minimum	Maximum	Mean	Standard Deviation
Multi-vehicle crashes	0	22	5.69	4.41
Straight crossing path crashes	0	12	3.26	2.72
Major road AADT	950	9,560	3,767	2,020
Minor road AADT	174	3,120	1,111	617
Speed limit (mph)	35	55	50.83	5.62
Number of access points within 0.25 mi on major road	1	30	11.36	6.77
Shoulder width (ft)	0	3	1.55	0.78
Lane width (ft)	9	12	10.51	0.81
Number of access points within 250 ft on minor road	0	5	2	1.27
Distance to advance intersection warning sign (ft)	0	1,050	655	231
Distance to advance stop ahead warning sign (ft)	0	1,200	562	291
Reduction from posted to advisory speed (mph)	0	20	3.81	5.82
Average annual temperature (°F)	55.1	60.82	59.51	1.41
Number of days 90 °F and greater	10	59.8	50.06	8.73
Number of days 32 °F and lower	64	118	79.73	14.59
Total annual snowfall (inches)	0.74	9.76	3.69	2.02
Number of days with 0.1 inch of precipitation	67.6	98	78.95	4.36
Number of days with 0.5 inch of precipitation	24	41.6	30.93	3.20
Number of days with 1 inch of precipitation	8.8	17.4	11.83	1.46
Intersection angle (degrees)	49	90	71.44	11.58
Average ISD	355.75	1,321	942.80	236.76
ISD quality	1.5	3	1.92	0.35
Number of deficient ISD quadrants	0	4	1.07	1.13

Table 98. Summary data for categorical variables.

Variable	Condition	Frequency	Percentage
Pavement quality	Deficient	6	14.29
Pavement quality	Not deficient	36	85.71
Stop line	Present	4	9.52
Stop line	Not present	38	90.48
Stop signs	One present	24	57.14
Stop signs	Two present	18	42.85
Advance intersection warning sign	Present	39	92.86
Advance intersection warning sign	Not present	3	7.14
Advance stop ahead warning sign	Present	34	80.95
Advance stop ahead warning sign	Not present	8	19.05
Edge line extensions	Present	40	95.24
Edge line extensions	Not present	2	4.76
Advisory speed plaque	Present	14	33.33
Advisory speed plaque	Not present	28	66.67
Major road RRPM	Present	29	69.05
Major road RRPM	Not present	13	30.95
Minor road RRPM	Present	34	80.95
Minor road RRPM	Not present	8	19.05
Terrain	Flat	11	26.12
Terrain	Rolling or mountainous	31	73.88
Major road functional classification	Principal arterial	1	2.38
Major road functional classification	Major arterial	5	11.91
Major road functional classification	Minor arterial	10	23.81
Major road functional classification	Major collector	14	33.33
Major road functional classification	Minor collector	3	7.14
Major road functional classification	Local	9	21.43

Table 99. Summary data for time-based continuous variables.

Variable	Minimum	Maximum	Mean	Standard Deviation
Multi-vehicle crashes	0	7	1.14	1.40
Straight crossing path crashes	0	4	0.65	0.92
Major road AADT	900	9,700	3,767	2,039
Minor road AADT	174	3,120	1,111	503
Average annual temperature (°F)	53.4	61.8	59.51	61.8
Number of days 90 °F and greater	3	91	50.06	21.33
Number of days 32 °F and lower	43	137	79.73	19.89
Total annual snowfall (inches)	0	22.5	3.69	5.27
Number of days with 0.1 inch of precipitation	48	120	78.95	10.26
Number of days with 0.5 inch of precipitation	13	49	30.93	6.06
Number of days with 1 inch of precipitation	3	23	11.83	2.45

AADT MODELS BASED ON EXPOSURE AGGREGATION

North Carolina intersection data were used to explore the impacts of AADT aggregation, considering the following levels of aggregation:

- A—AADT by year for the 5-year period.
- B—AADT averaged across the 5-year period.
- C—AADT broken into 6-hour periods by proportions developed from the quasi-induced demand methodology (annual average 6-hour traffic volume). Crashes were aggregated across the 5-year period for the 6-hour window at each intersection (e.g., average AADT for all crashes occurring from 12:00 a.m. to 5:59 a.m.).
- D—AADT broken into weekday versus weekend by proportions developed from the quasi-induced demand methodology. Crashes were aggregated across the 5-year period for weekdays or weekends at each intersection (e.g., average AADT for all weekend crashes).

Table 100 provides the coefficients and model fit for each of the aggregate levels considering major and minor road AADT independently. Table 101 provides the coefficients and model fit for each of the aggregation levels considering total entering volume and the proportion of entering volume on the minor road approach. Both tables consider multi-vehicle crashes for this illustration.

Table 100. Model coefficients for independent major and minor road AADTs.

Variable	A	B	C	D
Constant	–6.511 (1.487)	–4.798 (1.548)	–6.344 (0.754)	–5.149 (0.811)
Log major AADT	0.546 (0.149)	0.449 (0.173)	0.542 (0.140)	0.468 (0.140)
Log minor AADT	0.312 (0.186)	0.412 (0.181)	0.534 (0.143)	0.400 (0.140)
Alpha	0.454	0.171	0.210	0.147
Pseudo R^2	0.035	0.073	0.218	0.148
Log-likelihood	–298.16	–103.76	–215.32	–155.12
N	210	42	168	84

Note: Numbers in parentheses are standard errors.

Table 101. Model coefficients for total entering volume and proportion on minor road.

Variable	A	B	C	D
Constant	–7.781 (1.895)	–6.772 (1.874)	–8.173 (0.931)	–6.494 (0.958)
Log total entering	0.909 (0.207)	0.954 (0.207)	1.118 (0.115)	0.895 (0.109)
Proportion minor	0.848 (0.906)	1.780 (0.937)	2.241 (0.786)	1.561 (0.775)
Alpha	0.445	0.153	0.172	0.128
Pseudo R^2	0.036	0.086	0.228	0.157
Log-likelihood	–297.82	–102.30	–212.37	–153.52
N	210	42	168	84

Note: Numbers in parentheses are standard errors.

Due to differing sample sizes, it is difficult to make generalizations by aggregation level across models. Keeping in mind that sample size can influence model fits, the results indicate the following:

- In all cases, the total entering volume and proportion of traffic on the minor road provide a better model fit than independent AADT values for the major and minor road.

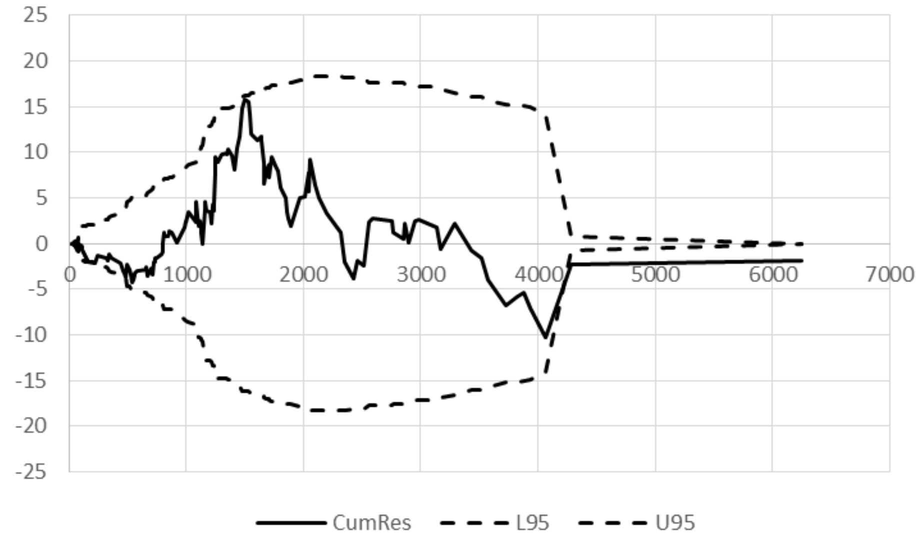
- Aggregating across years seems to provide one observation for each site, which provides minor benefit in terms of model fit. The pseudo R^2 is somewhat improved (larger) and the dispersion parameter is somewhat improved (smaller). The coefficient for total entering volume is consistent and the coefficient for proportion of traffic on the minor road increases substantially, but insignificantly.

- Aggregation into 6-hour increments provides a much-improved pseudo R^2 over the averaged annual model (Model B), but the dispersion parameter becomes worse. It is likely that the sample size negatively impacts both the pseudo R^2 and the dispersion parameter, indicating that the improvement to the pseudo R^2 is even better than first sight suggests.

- Aggregation into 6-hour increments results in the proportion of minor road traffic variable being statistically significant (with 95-percent confidence).

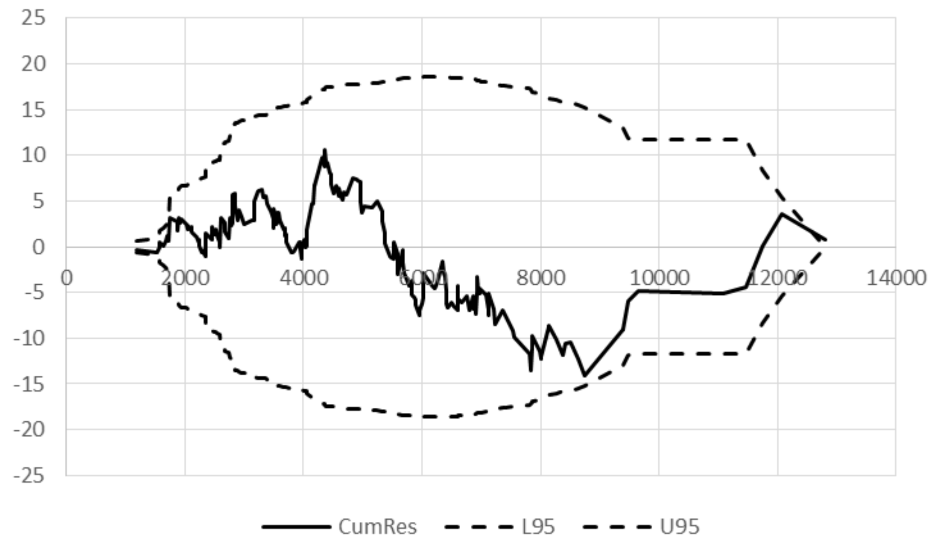
- Weekday versus weekend AADT aggregation results in model improvement over the averaged annual model, but is not as good as the 6-hour increment model.

Figure 70 and figure 71 provide the CURE plots for Model C and Model A, respectively. The X-axis shows that total entering volumes differ between the 6-hour increment model and the AADT model. Areas of increase and decrease in the CURE plots denote areas where crashes are under- and over-predicted by the model in comparison to observed. While the sample size differs for the two models, the total number of crashes is the same between them. The magnitude of bias between the two models is similar, but in opposite directions (maximum of 15 for Model C and −15 for Model A). Both models indicate a systematic over-prediction in the mid-ranges of AADT, but there is a larger under-prediction in the lower mid-range for Model C. Model C does very well in the extremely low volume regions, which indicate nighttime AADT and nighttime crashes.



Source: FHWA.

Figure 70. Plot. CURE plot for Model C.



Source: FHWA.

Figure 71. Plot. CURE plot for Model A.

Table 102 provides the final full models for multi-vehicle and straight crossing path crashes. The final models were estimated for annual crashes using AADT and for aggregate 6-hour crashes using the annual average 6-hour traffic volume. Each model presented is the best model for that crash type and level of aggregation. The results indicate that the models were consistent in which variables were statistically significant across crash types and level of aggregation. The model for multi-vehicle crashes with AADT had more significant variables, but not substantively more. Although the sample size reduced from 239 to 137, the number of significant variables did not change substantially between multi-vehicle crashes and straight crossing path crashes. Additionally, the coefficients for significant variables were consistent across levels of data aggregation and somewhat consistent across crash type (i.e., they were not significantly different). One notable outcome is that the estimated coefficient for total entering volume became close to 1.0 as more data were included in the full models for aggregate 6-hour crash data. This indicates that with more precise exposure data, the impact of exposure becomes nearly linear.

Table 102. Final full models.

Variable	Annual Crashes MV	Annual Crashes SCP	Aggregate 6-Hour Crashes MV	Aggregate 6-Hour Crashes SCP
Constant	–7.272**	–6.316**	–8.271**	–8.561**
Total entering	0.744**	0.596**	1.067**	1.034**
Proportion from minor road	--	--	--	1.785*
Posted 55	0.451**	0.663**	0.409**	0.581**
Low access density	–0.764**	–0.943**	–0.637**	–0.797**
Narrow lane	0.488**	0.605**	0.545**	0.576**
Stop line	--	–0.611*	--	–0.960**
Stop ahead	0.475**	--	0.449**	--
More snow	0.458**	--	--	--
Low ISD quality	0.491**	0.556**	0.386**	0.513**
Local	–0.405*	--	--	--
Alpha	0.084	<0.001	<0.001	<0.001
LL	–277.91	–210.20	–195.96	–153.86
N	210	210	168	168
Pseudo_R^2	0.101	0.091	0.288	0.272

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
--Variable not included in final model specification.

The advantage of annual data over aggregate data is the consideration of time effects. The model for annual multi-vehicle crashes had a significant coefficient for a higher annual snowfall indicator. When aggregating across years, this effect was lost (since sites in North Carolina had a similar average annual snowfall over the 5-year period). Additionally, annual factors can be included in models to account for unobserved trends that occur across years (which often includes weather effects). Table 103 shows the multi-vehicle annual crashes model with annual factors included in the model (with and without the indicator for higher annual snowfall). Comparing table 102 to table 103, as the annual factors are added to the model, the coefficient for the snowfall indicator becomes marginally significant (significant with 90-percent confidence). When removing the indicator for snowfall in table 103, the effects for annual indicators generally increase, with the indicator for 2010 becoming significant with 95-percent confidence. This is consistent with the summary data, which indicate that the average snowfall for North Carolina sites was 1.9 inches for years other than 2010 and 10.7 inches for 2010.

Table 103. Multi-vehicle crashes with annual factors.

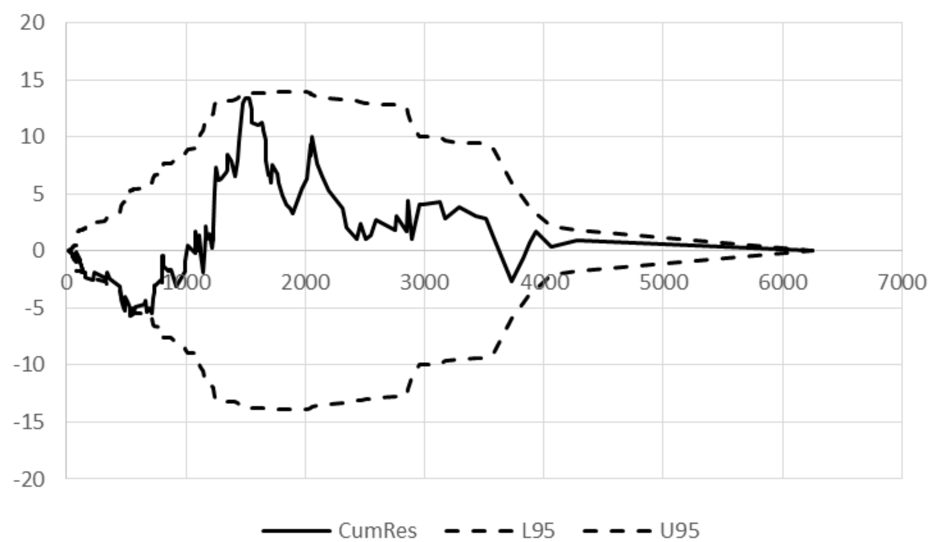
Variable	With Snow Indicator	Without Snow Indicator
Constant	–7.388**	–7.387**
Total entering	0.733**	0.739**
Posted 55	0.457**	0.427**
Low access density	–0.771**	–0.761**
Narrow lane	0.491**	0.473**
Stop ahead	0.477**	0.473**
More snow	0.487*	--
Low ISD quality	0.498**	0.450**
Local	–0.418*	–0.380*
Year2009	0.107	0.165
Year2010	0.191	0.554**
Year2011	0.202	0.214
Year2012	0.449**	0.449*
Alpha	0.082	0.110
LL	–275.66	–277.32
N	210	210
Pseudo_R^2	0.108	0.103

*Significant with 80-percent confidence.
**Significant with 90-percent confidence.
--No data.

The model statistics in table 102 indicate that the aggregate 6-hour crashes models fit the data better than the annual crashes models. However, there is little difference in model coefficients, except for exposure-level characteristics (i.e., traffic volume). Comparison of predicted values to observed values provides a better indication of how models are performing. Table 104 provides a comparison of the full models in terms of differences between the predicted and observed crashes. The maximum number of observed crashes is provided, along with residuals, mean absolute deviation, and mean square error. The statistics indicate that the mean absolute deviation and mean square error are slightly better for the 6-hour multi-vehicle crashes model, although the maximum residuals are larger for this model. The CURE plots for these models are presented in figure 72 and figure 73. Figure 72, which provides the CURE plot for 6-hour multi-vehicle crashes, and figure 73, which provides the CURE plot for annual multi-vehicle crashes. The CURE plots indicate that there is bias throughout the full model for 6-hour multi-vehicle crashes (although insignificant). The CURE plot for annual multi-vehicle crashes indicates little bias throughout the range of data for the model. This shows the difficulty in selecting the best model, and that the overall results indicate that if more disaggregate data are not available, annual data are sufficient.

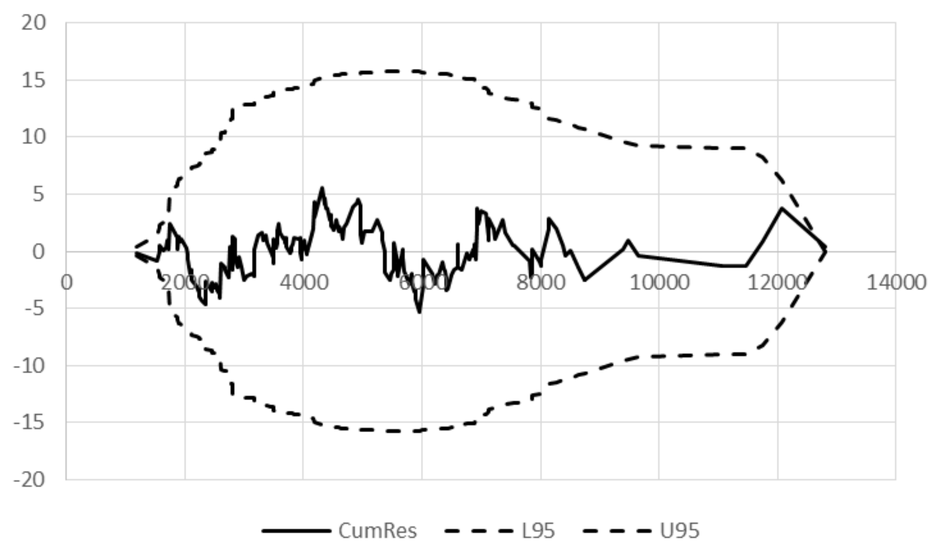
Table 104. Statistics for comparing models.

Statistic	Annual Multi-Vehicle Crashes	6-Hour Multi-Vehicle Crashes
Maximum observed	7	12
Maximum/minimum residual	2.99/–1.23	3.35/–3.73
Mean absolute deviation	0.88	0.77
Mean square error	1.29	1.25



Source: FHWA.

Figure 72. Plot. CURE plot for 6-hour multi-vehicle crashes full model.



Source: FHWA.

Figure 73. Plot. CURE plot for annual multi-vehicle crashes full model.

CHAPTER 11. SUMMARY OF FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

The overall objective of this effort was to increase understanding of causative, precipitating, and predisposing factors of crashes occurring on horizontal curves and at unsignalized intersections along rural two-lane roads. Specifically, we explored methods to understand how and to what level human, vehicle, and environmental (including roadway environment) elements contribute to crash occurrence and what opportunities exist to more effectively identify and understand the direct and indirect effects of these elements using enhanced data.

In the first part of this study, the project team developed generalized conceptual crash model frameworks, informed by a review of supporting published literature on conceptual crash models and contributing factors, alternative approaches to accident analysis, and the role of constraints in systemic approaches to accident analysis. The framework proved useful from several perspectives, including:

- Identifying and organizing all factors that influence the likelihood of a crash and defining the event sequences that lead to a crash.
- Providing terminology that will encourage clear communication across accident analysis disciplines as research on crash causation continues.
- Visualizing the nature by which a certain factor influences the likelihood of a crash or by which an event directly causes a crash.
- Identifying data needs (versus data availability) for studying the precipitating events, system constraints, predisposing factors, and target groups associated with a specific crash type.

Future applications of the conceptual crash model framework in studies exploring crashes at a more microscopic level seems promising, and is now possible with the availability of NDS data.

After marrying the conceptual crash model framework with available data, the project team proposed two sets of studies to implement ideas for expanding beyond limited datasets and variables traditionally used by highway and traffic engineers:

- The Benefits and Trade-Offs of Enhanced Data and Refined Crash Type Definitions on Models Relating Expected Crash Frequency to Predisposing Roadway, Traffic, and Weather Factors.
- Developing Crash Causal Types Using Data from Detailed, On-Scene Crash Investigations.

The overall conclusion of the second part of this effort is that expanding beyond traditional databases used for crash-based evaluations (i.e., crash, volume, and traditional roadway inventory databases maintained by State and local transportation agencies) can provide further insight into these crashes. Enhanced data collection and subsequent analysis were demonstrated for three high priority crash scenarios on rural two-lane roads: “straight crossing path crashes” at unsignalized intersections, combination “control loss/no vehicle action” and “road edge departure/no maneuver” single-vehicle crashes on horizontal curves, and “opposite direction/no maneuver crashes” on horizontal curves. The project team conducted four separate studies in the second part of this effort: unsignalized intersection study, Washington horizontal curve study, Utah horizontal curve study, and NMVCCS study. The conclusions for each of these studies are presented in the following sections.

CONCLUSIONS FROM THE UNSIGNALIZED INTERSECTION STUDY

The analysis of crashes at unsignalized intersections compared the results of an analysis of SCP crashes utilizing traditional data with traditional data supplemented by other sources of data (i.e., enhanced data). Additionally, the analysis compared models linking predisposing factors to SCP crashes to models linking predisposing factors to all multi-vehicle crashes. Variables available from traditional data sources did not fully describe the predisposing factors associated with the frequency of the multi-vehicle and SCP crashes. Parameter estimates for variables only available from non-traditional data sources were influential and statistically significant, indicating that models that do not consider these variables suffer from omitted variable bias. These variables included detailed information on traffic control devices (e.g., stop sign number/placement, stop ahead warning sign presence, speed advisory sign presence), geometrics (e.g., approach grade, intersection sight distance, horizontal alignment on intersection approaches), and weather (e.g., frequency of below freezing temperatures, snowfall frequency/amount). Variables with statistically significant parameters included in model specifications developed using the traditional databases were also present and included in model specifications using the enhanced databases, but the magnitude of their estimated effect was different (e.g., the estimated safety benefits of increasing lane widths were larger in the models specified with the enhanced dataset than with the traditional dataset). In addition, the estimated dispersion parameter was nearly 70 percent smaller for the models specified with the enhanced dataset than with the traditional dataset. Future researchers should consider developing recommendations or research protocols (similar to what is available for creating high quality SPFs and CMFs) that identify “minimum data elements” for different crash types. The generalized conceptual crash modeling frameworks presented in chapter 3 and implemented in chapter 4 would be useful for informing these protocols. Several positive outcomes of implementing these types of data protocols would be likely:

- Improve the reliability of results for both before–after and cross-sectional road safety studies conducted using traditional analysis methods.
- Improve the effectiveness of emerging analysis methods intended to increase the repeatability of observational road safety study results.
- Allow researchers to consider more refined crash types definitions based on countermeasures of interest.

The reduced sample size that comes with looking at SCP crashes (versus all multi-vehicle crashes) was offset by a gain in efficiency. While the sample size was smaller, the project team still observed similar levels of statistically significant variables for this crash type versus all multi-vehicle crashes. This is likely due to efficiency gains of looking at similar crash types that have similar generating processes. There is likely more noise when aggregating additional crash types.

The primary conclusion from this portion of the study is, regardless of the intended use of crash prediction models (or SPFs), it is imperative that researchers have a sound theoretical understanding of what factors are associated with the crash type of interest, and how the predisposing factors of interest interact with one another.

Additionally, the project team used several supplemental databases and field collection to create the enhanced dataset used for analyses. Field collection included traffic volumes on the minor routes, intersection sight distance, and vertical gradient. Each of these three elements differed in the amount of time and manpower required. Obtaining the traffic volumes for minor routes with no AADT data was the most time-consuming effort, while obtaining intersection sight distance required the most manpower.

The project team used driver vantage imagery and Google® Earth™ extensively to obtain supplemental data, such as sign location in reference to the intersection and messages on signs. The project team obtained these data quickly and accurately, in a desktop environment. Further, driver vantage imagery was available from 2007 to 2014 for most locations, allowing for observation of changes over time. Analysts collected desktop-based data for many intersections in one day, whereas field-based data collection would have allowed for one or two sites per day.

NOAA weather data supplemented the traditional data and the project team used many land-based stations by relative location to study intersections. The unsignalized intersections were mostly located in central and eastern North Carolina, as well as northwestern Ohio, where the terrain is relatively flat or rolling. For these locations, using the nearest land-based station data was justifiable. However, for locations with mountains and large hills, the terrain may impact the weather patterns such that the study site may not have similar weather to the nearest station. For the horizontal curve study, the project team used the nearest locations for ease of data collection, using GPS coordinates. Future research should consider the impact of using proximity-based stations versus terrain-based stations to determine if this has an impact.

CONCLUSIONS FROM WASHINGTON HORIZONTAL CURVE STUDY

Washington undertook an effort to collect roadside features as part of an RFIP. The data collection fell between 2006 and 2012, and focused heavily on rural two-lane roads. In this study, the project team used the WSDOT RFIP data in combination with other roadway databases available through HSIS and WSDOT. The hypothesis was that detailed roadside information is important to study horizontal curve crashes because it directly impacts available sight distance when roadside features are on the inside of a horizontal curve. Roadside design may also influence the sequence of events leading to certain crash outcomes. Without addressing these features, analysis results may falsely “over-attribute” the frequency of certain crash outcomes to characteristics of the roadway. In addition to the roadside inventory, the project team also incorporated detailed weather data into the dataset. The project team combined coordinate-based roadside inventories with linear-referenced roadway inventories using GIS mapping software. The project team used the same software to compute average offsets from the centerline to continuous roadside features (e.g., barrier, fence) as well as individual counts of, and offsets to, other roadside features (e.g., trees) efficiently.

For single-vehicle crashes, analysis results showed very stable findings when comparing parameters estimated using traditional databases, advanced databases, and databases enhanced with non-traditional variables. Statistically significant variables in the model specification developed using the traditional dataset included AADT, lane width, shoulder width, and posted speed limit. The project team used statistically significant variables in the final model specifications including AADT, lane width, posted speed limit, and horizontal curve radius. The project team arrived at the final model specifications by using enhanced datasets including AADT, posted speed limit, horizontal curve radius, vertical curve presence, number of days at 90 °F or more, number of days at 32 °F or less, number of days with more than 1 inch of rainfall, and average tree diameter. The estimated dispersion parameter was only 6 percent smaller for the traditional single-vehicle models specified with the enhanced dataset than with the traditional dataset; it was nearly 30 percent smaller for the more refined single-vehicle crash type definition (i.e., combination “control loss/no vehicle action” and “road edge departure/no maneuver” adapted from the previously published pre-crash scenarios). The magnitude of the estimated effects of lane width, shoulder width, and horizontal curvature did decrease as the project team incorporated additional, non-traditional variables into the model, indicating that the effects of these variables maybe overestimated in more limited model specifications. Parameters associated with variables built from the WSDOT RFIP (e.g., guardrail presence and length, tree count, tree diameters, fixed object count) were generally in the direction expected for single-vehicle crashes, and were statistically significant for the more refined single-vehicle crash type definition. However, the sample sizes and analysis did not support any direction-specific conclusions related to sight distance restrictions from these roadside features when located on the inside of a horizontal curve. As with the unsignalized intersection analysis, the overdispersion parameter was smallest for the models estimated with the enhanced databases.

For multi-vehicle crashes, the model parameters were less stable and more difficult to explain. The expected number of multi-vehicle crashes were shown to increase as lane width and shoulder width increased. This estimated effect was larger in magnitude as additional variables were included in the model specification. The direction of the regression parameter quantifying the speed limit effect was negative for the all multi-vehicle models, indicated a decrease in the expected number of crashes as speed limit increases. It was, however, positive and statistically significant at a higher level of confidence for the more refined multi-vehicle crash type definition (i.e., opposite direction/no maneuver crashes), indicating the opposite effect of an increase in the expected number of crashes as speed limit increases. The same general findings and recommendations from the unsignalized intersection seem to hold here. Parameter estimates for variables only available from non-traditional data sources were influential and statistically significant, indicating that models that do not consider that these variables suffer from omitted variable bias. Additionally, the reduced sample size that comes with looking at the more refined crash type definitions was offset by a gain in efficiency. While the sample size was smaller, the project team still observed similar levels or higher levels of statistically significant variables for the more refined crash type models.

CONCLUSIONS FROM UTAH HORIZONTAL CURVE STUDY

As with the unsignalized intersection study, this set of two sub-studies explored the potential to increase understanding the causative, precipitating, and predisposing factors of crashes occurring on horizontal curves through the use of new data sources for safety research. This process provided some insights into these alternative data sources and recommendations for processing and analyzing the data for the purpose of studying road safety.

The horizontal curve study used two different datasets, one from Utah and one from Washington State. Between these two datasets, the LiDAR-based dataset from Utah differs the most from other electronically coded datasets that researchers have traditionally used for most safety studies (e.g., HSIS, State-maintained databases). These differences originate from the fact that the project team employed a new technology for collecting and processing the data. Every element of this dataset also incorporates GPS. This offers the opportunities of experimenting, transitioning, and potentially adopting a GPS-based spatial referencing system instead of the linear referencing approach (i.e., route and milepost). The transition from linear-based referencing to geo-coding for building safety analysis datasets is likely to become more prevalent as datasets become more robust.

It is important to distinguish that the dataset from Utah was designed for asset management purposes. Many important data elements for a safety study might not be available even though it has many others that are not very meaningful from a safety research perspective. This is a primary conclusion of this effort. Databases collected for non-safety purposes may have meaningful information for safety but, because the data were collected for another purpose, the database may be missing critical information needed for safety analysis, or additional post-processing may be required. In other words, safety-specific processing of original point cloud data (in terms of features needed and level of accuracy) is likely necessary to fully realize the benefit of LiDAR data from a safety perspective. The following sections provide a summarized discussion of some of the issues the project team encountered when working with Utah dataset.

Horizontal Curve Data

UDOT provided the project team with the horizontal curve data file in a GIS shape file format. The file contained key pieces of information on horizontal curves and tangents (which also appear as curves with very large radii) for the entire State-maintained roadway network. Every data point came with GPS coordinates and elevation. However, the data points often did not represent a complete curve but, instead, one curve was broken into numerous short segments with different radii and curve degrees. Further investigation uncovered that the contractor's curve algorithm may have been falsely associating changes in steering with the start/end of a new horizontal curve. The project team therefore developed algorithms for merging these short segments into complete curves and estimating the PC and PT locations, curve length, and deflection angle from these pieces. Although the algorithm worked well for many locations, there were issues related to accuracy when curves had small deflection angles. The algorithm also had problems accurately detecting and estimating curves for winding stretches of roadway where no or very short tangents exist between curves. It is also worth noting that the horizontal alignment data for these particular locations, in their original form, were often inaccurate. Improved algorithms for processing the curve data and better estimating the complete horizontal curves would be beneficial to future safety studies that rely on this type of dataset.

GPS-Based Crash Location Identification

Utah crash data have GPS coordinates for each crash location. This could potentially solve a problem with crash location identification using a linear referencing system (based on route number and mileposts) where major reconstructions cause milepost shifts that impact multi-year analyses. Misidentified, misreported, and miscoded mileposts are another potential problem with using the linear referencing system. GPS coordinates could help locate and link crashes to roadway entities (e.g., segment, intersection, interchange) more accurately. This approach to identifying crashes could especially be useful for linking crashes to smaller roadway entities that require increased accuracy, such as intersections or interchange ramps. However, this method also has its potential complications. In this study, each horizontal curve location was identified by the GPS coordinates of the PC and PT. Therefore, it was relatively straightforward to identify and assign crashes to each curve by their relative locations to the PC and PT based on their GPS coordinates. This method will likely be accurate for intersections and interchanges. However, linking crashes to a long roadway segment with a complex alignment may require more information than just the coordinates of the beginning and ending points. Accurate GPS coordinates of points along the horizontal alignments would be necessary for moving toward a GPS-based spatial referencing system.

Surface Area and Lane Width

Researchers often consider lane width one of the key components for crash prediction models. The Utah dataset, built from contractor-process LiDAR data, did not provide lane width information or any other data elements that the project team could use to calculate lane width. Surface area, provided in the pavement information file, was the closest element from which to determine lane width. It was possible to calculate average surface width from the area width. However, the data file provided the surface areas for only very long segments of roadway (e.g., several miles). The calculated surface width would be the average value for a long stretch of roadway without knowing the actual surface width for given point within that segment. Given the fact that lane width and shoulder width change, especially for horizontal curves within a stretch of roadway, the project team did not consider the average value very meaningful for a safety analysis.

Barrier Offset

Utah data provided detailed locations and types of all roadside barriers, as the primary function of the database was asset management. While barrier length could be easily calculated from the barrier mileposts, the barrier offset was neither available nor could it be determined. The barrier locations are fairly accurate because they come with GPS coordinates. However, the project team could not reliably obtain their relative locations to the roadway edge (or roadway centerline). Although the dataset comes with route centerlines in a GIS shape file, their locations were not as accurate as needed to calculate the distance between the centerlines to the roadside barriers with reasonable accuracy and reliability. Utah DOT plans to have their data collection contractor report barrier offset during its most recent round of LiDAR data collection.

Accuracy of GPS Coordinates

The project team examined the data and found that GPS coordinates associated with the data elements are not always accurate, especially when it comes to using the coordinates to calculate short distances with a reasonable accuracy. This appears to happen more often in mountainous areas. This could be related to the fact that the mountains might have negative effects on the GPS signal quality. Therefore, careful examination and evaluation of data accuracy are necessary when working with data that depend on GPS coordinates.

CONCLUSIONS FROM NMVCCS STUDY

This study focused on developing crash “causal types,” or similar crashes grouped together based on their key precipitating events, predisposing factors, and target groups. The crashes were analyzed individually using NMVCCS data, photographs, and narratives developed from detailed, on-scene crash investigations. The intent of the study was to explore whether or not examining individual crash reports, photographs, and other available evidence to better understand factors leading to a crash, instead of relying solely on data summaries and modeling of electronically coded variables, holds promise from a safety research perspective. The NMVCCS database provides detailed information on the crash, manner of collision, drivers and vehicles involved and the crash location. Each case report comes with detailed descriptions, sketches, and photographs. However, such a level of detail also becomes a challenge for building statistical models. This database certainly offers valuable details about the crash that no other crash database does. However, more efforts are necessary to collect data from the case reports and process it in a format that researchers could use for statistical analyses, or to have access to enough cases fitting a specific causal type grouping to be able to further analyze that grouping with statistical analysis.

It proved possible to combine location information with detailed crash data from NMVCCS and identify crashes for a specific area type and pre-crash scenario combination (to the best of our knowledge, this study was the first to utilize location information for NMVCCS crashes). The project team used detailed data, coded from on-scene crash investigations, to try to uncover more microscopic interpretations of known empirical associations or develop causal type groupings of similar crashes based on the combination of their key precipitating events, predisposing factors, and target groups. In the end, the project team developed seven and three causal type groupings for the high-priority horizontal curve and unsignalized intersection crashes, respectively. However, most causal type grouping focused on one or more types of driver error. From a highway and traffic engineering perspective, the role that traffic and roadway factors played in the crash event sequence was of particular interest, but was difficult to determine. Therefore, the overall utility and practicality of this approach still remains unclear, particularly with emerging datasets and data sources such as NDS and the emerging technologies identified in the future research section of this chapter.

ESTIMATING EXPOSURE

After exploring and assessing multiple traditional and non-traditional data sources as outlined in the previous chapters of this report, it still proved difficult to find data on traffic patterns at more disaggregate levels than estimates of daily traffic. This is commonly true for rural areas like the ones this study explores. This lack of information is troublesome because traffic patterns are a key indicator of user exposure to crashes at various times of day. For example, two sites with the same AADT may have significantly different safety performance due to differences in day/night volume distributions. The lack of information on daily travel patterns, as well as suspected uncertainty in daily volume estimates by time of day, remains the “elephant in the room” when analyzing rural road safety. The project team therefore executed a final set of studies as part of this research to accomplish the following:

- Demonstrate and assess the use of kriging techniques to estimate day and night traffic volumes at rural, horizontal curve locations in Utah and rural, unsignalized intersection locations in North Carolina.

- Investigate the use of the quasi-induced demand approach, which uses data on crash history and "not-at-fault" drivers to estimate more disaggregate measures of exposure based on traffic volumes at rural, unsignalized intersection locations in North Carolina.

- Explore the use of socioeconomic data as a surrogate for typically unobserved characteristics related to crashes.

The kriging techniques implemented in this study to estimate average annual day and night traffic volumes in rural locations where permanent counters are not available showed promise. The approach was successful in Utah and unsuccessful in North Carolina, the latter likely due to inadequate ATR coverage. This study verified a tested hypothesis that horizontal curves with higher proportions of traffic at night are expected to experience more crashes than similar curves with higher proportions of traffic during the day with a positive parameter for night-to-day volume ratio in a negative binomial regression model of total expected crash frequency. The parameter estimate, however, was noisy and statistically insignificant, which is most likely attributable to the errors in the kriging predictions. Additional modifications and extensions to the kriging and safety modeling approaches are offered to improve the kriging predictions, and therefore reduce the standard error associated with the predicted night and day volumes. These include the following:

- Incorporating variables in addition to spatial proximity (e.g., functional classification, surrounding area characteristics) into the kriging model.

- Predicting night-to-day ratios directly, instead of night and day traffic volumes.

- Developing separate models for day and night crashes as a function of day and night volumes, respectively.

The quasi-induced demand approach was successful in estimating the percentage of daily volumes by driver age group as a function of either roadway class, day of week, time of day in terms of light conditions (i.e., day/night), and time of day in terms of 6-hour time intervals.

The use of these more disaggregated measures of traffic volumes in statistical road safety models of unsignalized intersections were tested using the following specifications of traffic volumes on the major and minor intersection approaches:

- AADT by year for a 5-year period.

- AADT averaged across a 5-year period.

- AADT broken into 6-hour periods by proportions developed from the quasi-induced demand methodology (annual average 6-hour traffic volume). Crashes were aggregated across the 5-year period for the 6-hour window at each intersection (e.g., average AADT for all crashes occurring from 12:00 a.m. to 5:59 a.m.).

- AADT broken into weekday versus weekend by proportions developed from the quasi-induced demand methodology. Crashes were aggregated across the 5-year period for weekdays or weekends at each intersection (e.g., average AADT for all weekend crashes).

Due to differing sample sizes, it was difficult to make generalizations by aggregation level across models. The results indicated the following (note that sample size can influence model fits):

- In all cases, the total entering volume and proportion of traffic on the minor road provided a better model fit than independent AADT values for the major and minor road.

- Aggregating across years provided one observation for each site, which provided minor benefit in terms of model fit. The pseudo R^2 is somewhat improved (larger) and the dispersion parameter is somewhat improved (smaller). The coefficient for total entering volume is consistent and the coefficient for proportion of traffic on the minor road increased substantially.

- Disaggregating traffic volumes into 6-hour increments provided a much-improved pseudo R^2 of the safety model over the averaged annual model, but the dispersion parameter got larger. It is likely that the sample size negatively impacted both the pseudo R^2 and the dispersion parameter.

- Aggregation into 6-hour increments resulted in the proportion of minor road traffic variable being statistically significant (with 95-percent confidence).

- Weekday versus weekend AADT aggregation resulted in model improvement over the averaged annual model, but is not as good as the 6-hour increment model.

The research was unsuccessful at quantitatively linking socioeconomic data to minor road traffic volumes at unsignalized intersections, which is often missing in traditional datasets. Analysis also indicated that socioeconomic variables were not associated with expected crash frequencies at unsignalized intersections. The project team discontinued any further exploration of socioeconomic data following these findings. In the end, kriging and quasi-induced demand techniques both showed some promise based in these exploratory studies, and future research should consider them, given the importance of traffic volume and exposure in most types of safety analysis.

FUTURE RESEARCH

Expansion to Other Crash Types and Situations

This project explores the potential utility of combining traditional and emerging safety data sources and employing analysis methodologies in an attempt to more fully understand precipitating events and predisposing factors of traffic crashes. This study provides specific examples and demonstrations for crashes occurring on horizontal curves and at unsignalized intersections along rural two-lane roads. The concepts presented here could, however, be applied to other situations and other crash types.

Traffic Volume Data

After assessing multiple traditional and non-traditional data sources, it still proved difficult to find data on traffic patterns at more disaggregate levels than just daily traffic, specifically for rural areas explored in this study. This remains a critical weakness in most statistical road safety models. Exposure by time of day and season of year would likely improve estimates of predisposing factor effects from statistical road safety models. This project included several studies to address this need, and in the end, kriging and quasi-induced demand techniques both showed some promise for estimating traffic patterns at more disaggregate levels than just daily. Additional research is needed in the areas of costs and benefits of various traffic volume data collection alternatives in rural areas.

Study Design and Analysis Methodologies

This study explored the potential impact of alternative model specifications made possible by enhanced data, as well as more refined crash type definitions on estimation of model parameters and other model properties. The project team used basic negative binomial regression. A more indepth exploration of this topic that considers a variety of model estimation techniques and study designs, including causal modeling and propensity score potential outcomes frameworks, is necessary. Simulation is also a promising alternative to explore whether relationships uncovered with the types of statistical models implemented in this research are reliable, and to further assess omitted predisposing variable effects.

Spatial and Temporal Resolution

Several spatial and temporal issues were encountered when combining traditional and non-traditional data sources. For example, weather stations provided nearly continuous weather data. Ways to aggregate that data to meaningful measures consistent with time periods used in road safety research (e.g., annual) was one of the challenges encountered during this research. Criteria for linking weather stations to road segments of interest are also necessary. This study utilized proximity measures (e.g., link a horizontal curve to the closest weather station), but there may be other, more appropriate linking criteria. Similar challenges were present for the roadside inventory data, where individual roadside elements had to be aggregated to the horizontal curve level. This same challenge would have also been present had traffic volume data been available at certain sensors in a more continuous format.

New and Emerging Sources of Data

The sources of non-traditional data and the methods by which to collect additional elements are expanded exponentially with the advent of smart device data, drones, and in-vehicle recorders—all of which present opportunity.

Smart Device Data

Smart device data is more appropriately described as data that is collected from phones, tablets, or other electronic devices that imparts a spatial and temporal dimension to the data. An example is traffic congestion information that is collected passively from smart phones in vehicles. Some groups are starting to make these data, particularly traffic volume data, publicly available. However, there is more than just volume that researchers could draw from these data.

Researchers could collect prevailing speed, detailed turning movement counts, pedestrian activity, and even the presence of incidents from these smart devices, and use the data to inform highway safety analysis.

Drones

The use of drones has recently been a growing controversy. Many interest groups are highly in favor of their use, while others, such as commercial pilots, find their use to be a major concern. Though controversial, drones have potential use for many industries. Farmers can use drones to examine their crops in a timelier fashion, while disaster relief teams can use drones to fly in hazardous areas and pinpoint the regions that need rescue teams immediately.

Transportation professionals can also benefit from working with drones, especially when inspecting large infrastructure projects or obtaining data from remote sites. Drones could obtain high resolution photography of a site or corridor being considered for safety analysis. Additionally, drones could inspect features such as lighting fixtures, poles, mast arms, bridges, and culverts. This would minimize traffic and safety concerns by removing crews from the field and placing them in a remote location.

Currently, commercial use of drones is prohibited in the United States, due to safety concerns. The Federal Aviation Administration is currently calculating ways to use drones safely with all other aircraft in the sky. There are many concerns with their use (e.g., airspace safety, invasion of privacy, registration, and monitoring). However, they may provide a new source of data.

In-Vehicle Recorders

Connected vehicles and autonomous vehicles have both made significant progress in the last few years. As these technologies are deployed in part, the equipped vehicles will collect a tremendous amount of information about their placement and interaction with the world around them. If this information can be available to researchers (note the high levels of access restrictions currently in place), it has the ability to be a powerful force to inform analyses.

IMPLICATIONS OF THE FINDINGS

This overall effort has potential implications for researchers and safety practitioners including State and local agencies. This effort has demonstrated that there is value to using data beyond traditional data sources for safety evaluations. The implication of this for researchers is that crash-based studies should attempt to look beyond the readily available data and consider other sources of data. Relying solely on traditional sources may lead to inappropriately attributing causation to the limited set of variables in crash, roadway, and volume databases (i.e., causation between predisposing factors and expected crash frequencies). Although the variables in these databases can be explanatory, they are not exhaustive in their ability to explain expected crash frequencies and may not be reliable predictors of future crash occurrence. Some supplemental data could already be available (with some post-processing) in databases that were collected and built for other purposes (e.g., asset management). While it is likely that only a small number of data elements from these supplemental or alternative data sources are useful for road safety study, they could potentially have significant impact on the ability to predict crash frequencies and the reliability and transferability of such predictions.

The implications for this effort for State and local agencies relate to the data that are collected. Although agencies have limited funds to expand collected data, an agency may want to consider collecting additional elements (e.g., skew) as efforts such as this and others demonstrate the value of these elements to informed decisionmaking. Potentially even more critical than the expansion of elements is the ability to integrate the data already collected by agencies with other sources of data or to leverage it for other uses. Communication between organizational units within public agencies about how data are or will be used could, for example, maximize the ultimate utility of data-related investments.

What is not completely answered by this project is how one assesses the "value added" by this additional data or the value of integrating these data with other sources. The project team collected the enhanced data in this effort with a reasonable amount of effort, particularly given the additional insight that was gained. However, additional research is necessary to quantify explicitly the modeling benefits and related decisionmaking benefits as well as the costs and "points of diminishing returns" associated with using different non-traditional data sources. This could lead to effective recommendations or research protocols (similar to what is available for creating high-quality SPFs and CMFs) that identify "minimum data elements" for different crash types. The current state of this research does, however, demonstrate a need for those researching the safety impacts of countermeasures or exploring the precipitating, predisposing, or causative factors of other crash scenarios to thoughtfully consider expanding their data set to include non-traditional sources of data.

APPENDIX. PRE-CRASH SCENARIO RANKING RESULTS

This appendix includes pre-crash scenario ranking results for each of the following rural two-lane subgroups: single-vehicle, horizontal curve; multi-vehicle, horizontal curve; single-vehicle, unsignalized intersection; and multi-vehicle, unsignalized intersection. All analysis reported in this appendix is based on "fresh runs" of 2005–2008 GES data using SAS programs that identify pre-crash scenarios from the GES provided by Dr. Wassim Najm. The SAS codes represented a "third generation" crash typology that combines information from the "GM 44-crashes typology" and the USDOT "pre-crash scenarios typology" in support of the IVI. Rankings based on both crash frequency and crash cost are provided (crash cost was ultimately used to identify priorities in the body of this work plan). This appendix also provides estimates on the percentages of all crashes and of total crash cost that is captured by the "top three" scenarios in each subgroup.

HIGH-PRIORITY, MULTI-VEHICLE PRE-CRASH SCENARIOS

As shown in table 105 and table 106, high-priority, multi-vehicle crashes include opposite direction/no maneuver (Volpe scenario 22), SCP @ non-signal (scenario 31), and Control loss/no vehicle action (scenario 4), which are identified as the pre-crash scenarios of primary interest on rural two-lane horizontal curves using both crash frequency and crash cost criteria. Those three pre-crash scenarios account for 55 percent of total crashes (see table 114) and 77.1 percent of total crash cost (see table 115) on rural two-lane horizontal curves. With respect to the high-priority, multi-vehicle pre-crash scenarios at rural two-lane unsignalized intersections, the top three pre-crash scenarios (as shown in table 107 and table 108) included SCP @ non-signal (scenario 31), Rear-end/LVS (scenario 27), and Left turn across path/opposite direction at non-signal (scenario 30). Those top three pre-crash scenarios account for 67.5 percent of total crashes (see table 114) that occur on rural two-lane unsignalized intersections and 79.2 percent of total cost (see table 116).

Table 105. High-priority, multi-vehicle pre-crash scenarios on rural two-lane horizontal curves identified using crash frequency.

Multi-Vehicle Pre-Crash Scenario	Curve Crash Frequency
22. Opposite direction/no maneuver	16,793
31. SCP @ non-signal	6,216
4. Control loss/no vehicle action	6,211

Table 106. High-priority, multi-vehicle pre-crash scenarios on rural two-lane horizontal curves identified using overall crash cost.

Multi-Vehicle Pre-Crash Scenario	Curve Crash Cost (millions of dollars)
22. Opposite direction/no maneuver	4,402
4. Control loss/no vehicle action	1,006
31. SCP @ non-signal	653

Table 107. High-priority, multi-vehicle pre-crash scenarios on rural two-lane unsignalized intersections identified using crash frequency.

Multi-Vehicle Pre-Crash Scenario	Intersection Crash Frequency
31. SCP @ non-signal	68,968
27. Rear-end/LVS	29,897
30. Left turn across path/opposite direction at non-signal	15,459

Table 108. High-priority, multi-vehicle pre-crash scenarios on rural two-lane unsignalized intersections identified using overall crash cost.

Multi-Vehicle Pre-Crash Scenario	Intersection Crash Cost (millions of dollars)
31. SCP @ non-signal	6,942
30. Left turn across path/opposite direction at non-signal	1,073
27. Rear-end/LVS	796

HIGH-PRIORITY, SINGLE-VEHICLE PRE-CRASH SCENARIOS

As shown in table 109, the top three pre-crash scenarios for single-vehicle crashes on rural two-lane horizontal curves based on frequency includes Control loss/no vehicle action (scenario 4), Road edge departure/no maneuver (scenario 8), and Animal/no maneuver (scenario 11). These three scenarios account for 87.4 percent of total crashes (see table 113). When using overall crash cost as the criterion, the result changed slightly (as shown in table 110). Control loss/no vehicle action (scenario 4), Road edge departure/no maneuver (scenario 8), and Object contacted/no maneuver (scenario 38) were identified as the top three pre-crash scenarios and account for 89.6 percent of total cost (see table 119).

Table 109. High-priority, single-vehicle pre-crash scenarios on rural two-lane horizontal curves identified using crash frequency.

Single-Vehicle Pre-Crash Scenario	Curve Crash Frequency
4. Control loss/no vehicle action	76,153
8. Road edge departure/no maneuver	40,009
11. Animal/no maneuver	11,606

Table 110. High-priority, single-vehicle pre-crash scenarios on rural two-lane horizontal curves identified using overall crash cost.

Single-Vehicle Pre-Crash Scenario	Curve Crash Cost (millions of dollars)
4. Control loss/no vehicle action	16,524
8. Road edge departure/no maneuver	8,323
38. Object contacted/no maneuver	515

For the single-vehicle crashes that occur at rural two-lane unsignalized intersections, the high-priority pre-crash scenarios based on crash frequency are shown in table 111 and include Control loss/vehicle action (scenario 3), Road edge departure/no maneuver (scenario 8), and Road edge departure/maneuver (scenario 7). These top three pre-crash scenarios account for 53.4 percent of total crashes at rural two-lane unsignalized intersections (see table 113). The high-priority pre-crash scenarios based on overall crash cost are listed in table 112 and include Control loss/vehicle action (scenario 3), Road edge departure/no maneuver (scenario 8), and Pedestrian/no maneuver (scenario 13). These three scenarios account for 58.2 percent of the total cost at rural two-lane unsignalized intersections (see table 121).

Table 111. High-priority, single-vehicle pre-crash scenarios on rural two-lane unsignalized intersections identified using crash frequency.

Single-Vehicle Pre-Crash Scenario	Intersection Crash Frequency
3. Control loss/vehicle action	8,017
8. Road edge departure/no maneuver	6,663
7. Road edge departure/maneuver	5,132

Table 112. High-priority, single-vehicle pre-crash scenarios on rural two-lane unsignalized intersections identified using overall crash cost.

Single-Vehicle Pre-Crash Scenario	Intersection Crash Cost (millions of dollars)
3. Control loss/vehicle action	857
8. Road edge departure/no maneuver	709
13. Pedestrian/no maneuver	701

Table 113. Estimated nationwide annual rural two-lane single-vehicle crashes, disaggregated by Volpe pre-crash scenario.

Single-Vehicle Pre-Crash Scenario	Curve	Curve (%)	Intersection	Intersection (%)
1. No driver present	30	0.0	60	0.2
2. Vehicle failure	3,009	2.1	440	1.2
3. Control loss/vehicle action	1,912	1.3	8,017	21.6
4. Control loss/no vehicle action	76,153	52.1	3,663	9.9
5. Running red light	0	0.0	106	0.3
6. Running stop sign	122	0.1	1,785	4.8
7. Road edge departure/maneuver	1,109	0.8	5,132	27
8. Road edge departure/no maneuver	40,009	27.4	6,663	18.0
9. Road edge departure/backing	832	0.6	72	0.2
10. Animal/maneuver	33	0.0	163	0.4
11. Animal/no maneuver	11,606	7.9	615	1.7
12. Pedestrian/maneuver	20	0.0	790	2.1
13. Pedestrian/no maneuver	300	0.2	1,068	2.9
14. Cyclist/maneuver	34	0.0	1,400	3.8
15. Cyclist/no maneuver	418	0.3	2,068	5.6
16. Backing into vehicle	0	0.0	0	0.0
17. Turning/same direction	0	0.0	0	0.0
18. Parking/same direction	0	0.0	0	0.0

Single-Vehicle Pre-Crash Scenario	Curve	Curve (%)	Intersection	Intersection (%)
19. Changing lanes/same direction	0	0.0	274	0.7
20. Drifting/same lane	0	0.0	0	0.0
21. Opposite direction/maneuver	106	0.1	79	0.2
22. Opposite direction/no maneuver	4,295	2.9	407	1.1
23. Rear-end/striking maneuver	39	0.0	33	0.1
24. Rear-end/lead vehicle accelerating	0	0.0	0	0.0
25. Rear-end/LMV	196	0.1	76	0.2
26. Rear-end/LVD	228	0.2	134	0.4
27. Rear-end/LVS	456	0.3	317	0.9
28. LTAP/OD @ signal	0	0.0	0	0.0
29. Turn right at signal	0	0.0	0	0.0
30. Left turn across path/opposite direction at non-signal	0	0.0	0	0.0
31. SCP @ non-signal	20	0.0	586	1.6
32. Turn @ non-signal	59	0.0	1,004	2.7
33. Evasive maneuver/maneuver	13	0.0	170	0.5
34. Evasive maneuver/no maneuver	938	0.6	225	0.6
35. Rollover	548	0.4	231	0.6
36. Non-collision—no impact	918	0.6	335	0.9
37. Object contacted/maneuver	251	0.2	612	1.7
38. Object contacted/no maneuver	2,615	1.8	536	1.4
Total	146,269	100.0	37,061	100.0

Table 114. Estimated nationwide annual rural two-lane multi-vehicle crashes, disaggregated by Volpe pre-crash scenario.

Multi-Vehicle Pre-Crash Scenario	Curve	Curve (%)	Intersection	Intersection (%)
1. No driver present	0	0.0	0	0.0
2. Vehicle failure	349	0.7	123	0.1
3. Control loss/vehicle action	250	0.5	1,699	1.0
4. Control loss/no vehicle action	6,211	11.7	1,041	0.6
5. Running red light	255	0.5	170	0.1
6. Running stop sign	100	0.2	9,653	5.7
7. Road edge departure/maneuver	3	0.0	4	0.0
8. Road edge departure/no maneuver	290	0.5	86	0.1
9. Road edge departure/backing	0	0.0	0	0.0
10. Animal/maneuver	15	0.0	0	0.0
11. Animal/no maneuver	96	0.2	63	0.0
12. Pedestrian/maneuver	3	0.0	0	0.0
13. Pedestrian/no maneuver	0	0.0	131	0.1
14. Cyclist/maneuver	0	0.0	0	0.0
15. Cyclist/no maneuver	0	0.0	0	0.0

Multi-Vehicle Pre-Crash Scenario	Curve	Curve (%)	Intersection	Intersection (%)
16. Backing into vehicle	800	1.5	5,619	3.3
17. Turning/same direction	1,990	3.7	10,519	6.2
18. Parking/same direction	498	0.9	428	0.3
19. Changing lanes/same direction	2,368	4.5	2,293	1.4
20. Drifting/same lane	1,728	3.3	1,201	0.7
21. Opposite direction/maneuver	355	0.7	150	0.1
22. Opposite direction/no maneuver	16,793	31.6	1,193	0.7
23. Rear-end/striking maneuver	166	0.3	1,014	0.6
24. Rear-end/lead vehicle accelerating	68	0.1	248	0.1
25. Rear-end/LMV	1,967	3.7	3,051	1.8
26. Rear-end/LVD	3,260	6.1	11,186	6.6
27. Rear-end/LVS	5,453	10.3	29,897	17.7
28. LTAP/OD @ signal	165	0.3	0	0.0
29. Turn right at signal	58	0.1	0	0.0
30. Left turn across path/opposite direction at non-signal	2,099	3.9	15,459	9.1
31. SCP @ non-signal	6,216	11.7	68,968	40.7
32. Turn @ non-signal	267	0.5	3,460	2.0
33. Evasive maneuver/maneuver	3	0.0	189	0.1
34. Evasive maneuver/no maneuver	327	0.6	862	0.5
35. Rollover	66	0.1	11	0.0
36. Non-collision—no impact	508	1.0	25	0.0
37. Object contacted/maneuver	17	0.0	0	0.0
38. Object contacted/no maneuver	247	0.5	0	0.0
39. Hit and run	15	0.0	25	0.0
40. Other—rear-end	0	0.0	11	0.0
41. Other—sideswipe	81	0.2	74	0.0
42. Other—opposite direction	0	0.0	0	0.0
43. Other—turn across path	0	0.0	0	0.0
44. Other—turn into path	0	0.0	0	0.0
45. Other—straight paths	0	0.0	0	0.0
46. Other	60	0.1	530	0.3
Total	53,145	100.0	169,385	100.0

Table 115. Annual weighted frequency on curves by injury type.

Multi-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost (%)
1. No driver present	0	0	0	0	0	0	0.0
2. Vehicle failure	244	0	22	83	0	0	0.3
3. Control loss/vehicle action	162	11	55	4	0	18	0.1
4. Control loss/no vehicle action	3,192	886	1,455	568	110	0	12.8
5. Running red light	206	18	30	1	0	0	0.1
6. Running stop sign	0	43	43	14	0	0	0.1
7. Road edge departure/maneuver	0	0	0	0	3	0	0.2
8. Road edge departure/no maneuver	206	1	0	47	11	25	1.1
9. Road edge departure/backing	0	0	0	0	0	0	0.0
10. Animal/maneuver	0	0	15	0	0	0	0.0
11. Animal/no maneuver	90	0	0	6	0	0	0.0
12. Pedestrian/maneuver	0	0	0	3	0	0	0.0
13. Pedestrian/no maneuver	0	0	0	0	0	0	0.0
14. Cyclist/maneuver	0	0	0	0	0	0	0.0
15. Cyclist/no maneuver	0	0	0	0	0	0	0.0
16. Backing into vehicle	602	148	18	32	0	0	0.3
17. Turning/same direction	1,628	202	86	63	11	0	1.4
18. Parking/same direction	484	14	0	0	0	0	0.0
19. Changing lanes/same direction	1,722	356	167	92	4	28	1.1
20. Drifting/same lane	1,396	145	125	62	0	0	0.5
21. Opposite direction/maneuver	212	45	33	36	30	0	2.5
22. Opposite direction/no maneuver	10,636	1,925	2,257	1,372	603	0	56.0
23. Rear-end/striking maneuver	103	31	17	15	0	0	0.1
24. Rear-end/lead vehicle accelerating	65	3	0	0	0	0	0.0
25. Rear-end/LMV	1,579	59	186	141	0	2	0.8

Multi-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost (%)
26. Rear-end/LVD	2,098	816	188	157	1	0	1.4
27. Rear-end/LVS	3,490	1,251	454	255	3	0	2.5
28. LTAP/OD @ signal	106	0	35	19	5	0	0.5
29. Turn right at signal	33	0	25	0	0	0	0.0
30. Left turn across path/opposite direction at non-signal	1,502	218	191	187	0	0	1.1
31. SCP @ non-signal	3,469	1,317	580	798	50	2	8.3
32. Turn @ non-signal	239	18	0	2	7	0	0.6
33. Evasive maneuver/maneuver	0	0	3	0	0	0	0.0
34. Evasive maneuver/no maneuver	135	1	40	84	67	0	5.6
35. Rollover	26	0	0	13	26	0	2.1
36. Non-collision—no impact	464	0	38	6	0	0	0.1
37. Object contacted/maneuver	0	17	0	0	0	0	0.0
38. Object contacted/no maneuver	212	0	35	0	0	1	0.1
39. Hit and run	0	0	15	0	0	0	0.0
40. Other—rear-end	0	0	0	0	0	0	0.0
41. Other—sideswipe	81	0	0	0	0	0	0.0
42. Other—opposite direction	0	0	0	0	0	0	0.0
43. Other—turn across path	0	0	0	0	0	0	0.0
44. Other—turn into path	0	0	0	0	0	0	0.0
45. Other—straight paths	0	0	0	0	0	0	0.0
46. Other	59	0	0	0	1	0	0.1
Total Cost	N/A	N/A	N/A	N/A	N/A	N/A	100.0

Table 116. Annual cost on curves by injury type (millions of dollars).

Multi-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost	Cost (%)
1. No driver present	0	0	0	0	0	0	0	0.0
2. Vehicle failure	1	0	2	24	0	0	27	0.3
3. Control loss/vehicle action	1	0	5	1	0	2	9	0.1
4. Control loss/no vehicle action	15	37	121	162	671	0	1,006	12.8
5. Running red light	1	1	2	0	0	0	4	0.1
6. Running stop sign	0	2	4	4	0	0	9	0.1
7. Road edge departure/maneuver	0	0	0	0	15	0	15	0.2
8. Road edge departure/no maneuver	1	0	0	13	69	3	86	1.1
9. Road edge departure/backing	0	0	0	0	0	0	0	0.0
10. Animal/maneuver	0	0	1	0	0	0	1	0.0
11. Animal/no maneuver	0	0	0	2	0	0	2	0.0
12. Pedestrian/maneuver	0	0	0	1	0	0	1	0.0
13. Pedestrian/no maneuver	0	0	0	0	0	0	0	0.0
14. Cyclist/maneuver	0	0	0	0	0	0	0	0.0
15. Cyclist/no maneuver	0	0	0	0	0	0	0	0.0
16. Backing into vehicle	3	6	2	9	0	0	20	0.3
17. Turning/same direction	7	9	7	18	69	0	110	1.4
18. Parking/same direction	2	1	0	0	0	0	3	0.0
19. Changing lanes/same direction	8	15	14	26	23	3	89	1.1
20. Drifting/same lane	6	6	10	18	0	0	41	0.5
21. Opposite direction/maneuver	1	2	3	10	183	0	199	2.5
22. Opposite direction/no maneuver	49	81	187	391	3,693	0	4,402	56.0
23. Rear-end/striking maneuver	0	1	1	4	0	0	7	0.1
24. Rear-end/lead vehicle accelerating	0	0	0	0	0	0	0	0.0

Multi-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost	Cost (%)
25. Rear-end/LMV	7	2	15	40	0	0	66	0.8
26. Rear-end/LVD	10	34	16	45	6	0	110	1.4
27. Rear-end/LVS	16	53	38	73	15	0	195	2.5
28. LTAP/OD @ signal	0	0	3	5	33	0	41	0.5
29. Turn right at signal	0	0	2	0	0	0	2	0.0
30. Left turn across path/opposite direction at non-signal	7	9	16	53	0	0	85	1.1
31. SCP @ non-signal	16	56	48	227	306	0	653	8.3
32. Turn @ non-signal	1	1	0	1	45	0	47	0.6
33. Evasive maneuver/maneuver	0	0	0	0	0	0	0	0.0
34. Evasive maneuver/no maneuver	1	0	3	24	409	0	437	5.6
35. Rollover	0	0	0	4	161	0	165	2.1
36. Non-collision—no impact	2	0	3	2	0	0	7	0.1
37. Object contacted/maneuver	0	1	0	0	0	0	1	0.0
38. Object contacted/no maneuver	1	0	3	0	0	0	4	0.1
39. Hit and run	0	0	1	0	0	0	1	0.0
40. Other—rear-end	0	0	0	0	0	0	0	0.0
41. Other—sideswipe	0	0	0	0	0	0	0	0.0
42. Other—opposite direction	0	0	0	0	0	0	0	0.0
43. Other—turn across path	0	0	0	0	0	0	0	0.0
44. Other—turn into path	0	0	0	0	0	0	0	0.0
45. Other—straight paths	0	0	0	0	0	0	0	0.0
46. Other	0	0	0	0	6	0	6	0.1
Total Cost	N/A	N/A	N/A	N/A	N/A	N/A	7,854	100.0

Table 117. Annual weighted frequency of intersection crashes by injury type.

Multi-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost (%)
1. No driver present	0	0	0	0	0	0	0.0
2. Vehicle failure	13	28	40	43	0	0	0.2
3. Control loss/vehicle action	1,489	54	91	47	1	18	0.3
4. Control loss/no vehicle action	716	60	219	37	10	0	0.8
5. Running red light	127	11	32	0	0	0	0.0
6. Running stop sign	4,859	2,213	1,865	645	15	57	5.0
7. Road edge departure/maneuver	0	0	4	0	0	0	0.0
8. Road edge departure/no maneuver	26	0	49	11	0	0	0.1
9. Road edge departure/backing	0	0	0	0	0	0	0.0
10. Animal/maneuver	0	0	0	0	0	0	0.0
11. Animal/no maneuver	61	3	0	0	0	0	0.0
12. Pedestrian/maneuver	0	0	0	0	0	0	0.0
13. Pedestrian/no maneuver	106	0	0	0	24	1	1.3
14. Cyclist/maneuver	0	0	0	0	0	0	0.0
15. Cyclist/no maneuver	0	0	0	0	0	0	0.0
16. Backing into vehicle	4,818	703	62	37	0	0	0.6
17. Turning/same direction	7,398	1,542	1,106	447	11	15	3.5
18. Parking/same direction	322	74	32	0	0	0	0.1
19. Changing lanes/same direction	1,773	292	154	74	0	0	0.5
20. Drifting/same lane	1,027	73	49	52	0	0	0.2
21. Opposite direction/maneuver	94	3	40	13	0	0	0.1
22. Opposite direction/no maneuver	738	154	118	129	54	0	3.5
23. Rear-end/striking maneuver	758	192	41	22	0	0	0.2
24. Rear-end/lead vehicle accelerating	64	159	22	3	0	0	0.1
25. Rear-end/LMV	1,973	662	294	103	7	12	1.2

Multi-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost (%)
26. Rear-end/LVD	8,541	1,734	657	254	0	0	2.2
27. Rear-end/LVS	20,699	6,281	1,996	839	4	79	7.2
28. LTAP/OD @ signal	0	0	0	0	0	0	0.0
29. Turn right at signal	0	0	0	0	0	0	0.0
30. Left turn across path/opposite direction at non-signal	8,365	3,512	2,234	1,294	54	0	9.6
31. SCP @ non-signal	39,421	13,556	9,958	5,268	627	137	62.4
32. Turn @ non-signal	3,145	108	150	56	0	0	0.4
33. Evasive maneuver/maneuver	101	84	3	0	0	0	0.0
34. Evasive maneuver/no maneuver	506	170	163	23	0	0	0.3
35. Rollover	0	0	3	0	0	9	0.0
36. Non-collision—no impact	9	0	14	2	0	0	0.0
37. Object contacted/maneuver	0	0	0	0	0	0	0.0
38. Object contacted/no maneuver	0	0	0	0	0	0	0.0
39. Hit and run	0	0	25	0	0	0	0.0
40. Other—rear-end	0	0	11	0	0	0	0.0
41. Other—sideswipe	0	74	0	0	0	0	0.0
42. Other—opposite direction	0	0	0	0	0	0	0.0
43. Other—turn across path	0	0	0	0	0	0	0.0
44. Other—turn into path	0	0	0	0	0	0	0.0
45. Other—straight paths	0	0	0	0	0	0	0.0
46. Other	440	35	35	19	1	0	0.2
Total	N/A	N/A	N/A	N/A	N/A	N/A	100.0

Table 118. Annual cost of intersection crashes by injury type (millions of dollars).

Multi-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost	Cost (%)
1. No driver present	0	0	0	0	0	0	0	0.0
2. Vehicle failure	0	1	3	12	0	0	17	0.2
3. Control loss/vehicle action	7	2	8	13	6	2	38	0.3
4. Control loss/no vehicle action	3	3	18	10	60	0	94	0.8
5. Running red light	1	0	3	0	0	0	4	0.0
6. Running stop sign	22	93	155	184	93	7	554	5.0
7. Road edge departure/ maneuver	0	0	0	0	0	0	0	0.0
8. Road edge departure/ no maneuver	0	0	4	3	0	0	7	0.1
9. Road edge departure/backing	0	0	0	0	0	0	0	0.0
10. Animal/maneuver	0	0	0	0	0	0	0	0.0
11. Animal/no maneuver	0	0	0	0	0	0	0	0.0
12. Pedestrian/maneuver	0	0	0	0	0	0	0	0.0
13. Pedestrian/no maneuver	0	0	0	0	145	0	146	1.3
14. Cyclist/maneuver	0	0	0	0	0	0	0	0.0
15. Cyclist/no maneuver	0	0	0	0	0	0	0	0.0
16. Backing into vehicle	22	30	5	11	0	0	67	0.6
17. Turning/same direction	34	65	92	127	69	2	389	3.5
18. Parking/same direction	1	3	3	0	0	0	7	0.1
19. Changing lanes/ same direction	8	12	13	21	0	0	54	0.5

Multi-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost	Cost (%)
20. Drifting/same lane	5	3	4	15	0	0	27	0.2
21. Opposite direction/maneuver	0	0	3	4	0	0	8	0.1
22. Opposite direction/no maneuver	3	6	10	37	332	0	388	3.5
23. Rear-end/striking maneuver	3	8	3	6	0	0	21	0.2
24. Rear-end/lead vehicle accelerating	0	7	2	1	0	0	10	0.1
25. Rear-end/LMV	9	28	24	29	43	1	135	1.2
26. Rear-end/LVD	39	73	55	72	0	0	239	2.2
27. Rear-end/LVS	95	265	166	239	22	9	796	7.2
28. LTAP/OD @ signal	0	0	0	0	0	0	0	0.0
29. Turn right at signal	0	0	0	0	0	0	0	0.0
30. Left turn across path/opposite direction at non-signal	38	148	186	368	333	0	1,073	9.6
31. SCP @ non-signal	181	572	827	1,500	3,845	16	6,942	62.4
32. Turn @ non-signal	14	5	12	16	0	0	47	0.4
33. Evasive maneuver/maneuver	0	4	0	0	0	0	4	0.0
34. Evasive maneuver/no maneuver	2	7	14	6	0	0	30	0.3
35. Rollover	0	0	0	0	0	1	1	0.0
36. Non-collision—no impact	0	0	1	1	0	0	2	0.0
37. Object contacted/maneuver	0	0	0	0	0	0	0	0.0
38. Object contacted/no maneuver	0	0	0	0	0	0	0	0.0

Multi-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost	Cost (%)
39. Hit and run	0	0	2	0	0	0	2	0.0
40. Other—rear-end	0	0	1	0	0	0	1	0.0
41. Other—sideswipe	0	3	0	0	0	0	3	0.0
42. Other—opposite direction	0	0	0	0	0	0	0	0.0
43. Other—turn across path	0	0	0	0	0	0	0	0.0
44. Other—turn into path	0	0	0	0	0	0	0	0.0
45. Other—straight paths	0	0	0	0	0	0	0	0.0
46. Other	2	1	3	5	6	0	18	0.2
Total	N/A	N/A	N/A	N/A	N/A	N/A	11,127	100.0

Table 119. Annual weighted crash frequency on curves by injury type.

Single-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost (%)
1. No driver present	0	0	26	4	0	0	0.0
2. Vehicle failure	2,064	259	450	208	28	0	1.0
3. Control loss/vehicle action	974	342	332	206	58	0	1.6
4. Control loss/no vehicle action	40,406	10,617	13,496	9,230	1,973	431	58.4
5. Running red light	0	0	0	0	0	0	0.0
6. Running stop sign	56	31	9	26	0	0	0.0
7. Road edge departure/maneuver	682	236	91	74	26	0	0.7
8. Road edge departure/no maneuver	20,727	6,136	6,972	4,876	973	326	29.4
9. Road edge departure/backing	708	81	38	5	0	0	0.0
10. Animal/maneuver	28	0	5	0	0	0	0.0
11. Animal/no maneuver	8,926	1,140	835	699	5	0	1.4
12. Pedestrian/maneuver	0	0	14	4	3	0	0.1
13. Pedestrian/no maneuver	0	56	108	81	56	0	1.3
14. Cyclist/maneuver	0	3	10	14	0	7	0.0
15. Cyclist/no maneuver	62	66	155	135	0	0	0.2
16. Backing into vehicle	0	0	0	0	0	0	0.0
17. Turning/same direction	0	0	0	0	0	0	0.0
18. Parking/same direction	0	0	0	0	0	0	0.0
19. Changing lanes/same direction	505	138	180	47	24	0	0.6
20. Drifting/same lane	0	0	0	0	0	0	0.0
21. Opposite direction/maneuver	0	33	40	18	14	0	0.3
22. Opposite direction/no maneuver	2,527	638	502	625	3	0	1.0
23. Rear-end/striking maneuver	26	0	13	0	0	0	0.0
24. Rear-end/lead vehicle accelerating	0	0	0	0	0	0	0.0
25. Rear-end/LMV	101	0	82	13	0	0	0.0

Single-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost (%)
26. Rear-end/LVD	118	5	61	45	0	0	0.1
27. Rear-end/LVS	312	55	70	19	0	0	0.1
28. LTAP/OD @ signal	0	0	0	0	0	0	0.0
29. Turn right at signal	0	0	0	0	0	0	0.0
30. Left turn across path/opposite direction at non-signal	0	0	0	0	0	0	0.0
31. SCP @ non-signal	20	0	0	0	0	0	0.0
32. Turn @ non-signal	23	0	18	18	0	0	0.0
33. Evasive maneuver/maneuver	0	0	0	13	0	0	0.0
34. Evasive maneuver/no maneuver	732	39	57	84	0	26	0.1
35. Rollover	148	185	67	139	8	0	0.4
36. Non-collision—No impact	695	40	75	67	40	0	1.0
37. Object contacted/maneuver	210	0	26	0	14	0	0.3
38. Object contacted/no maneuver	1,965	154	244	182	70	0	1.8
Total	N/A	N/A	N/A	N/A	N/A	N/A	100.0

Table 120. Annual cost on curves by injury type (millions of dollars).

Single-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost	Cost (%)
1. No driver present	0	0	2	1	0	0	3	0.0
2. Vehicle failure	9	11	37	59	170	0	286	1.0
3. Control loss/vehicle action	4	14	28	59	355	0	460	1.6
4. Control loss/no vehicle action	186	448	1,121	2,628	12,090	51	16,524	58.4
5. Running red light	0	0	0	0	0	0	0	0.0
6. Running stop sign	0	1	1	7	0	0	10	0.0
7. Road edge departure/maneuver	3	10	8	21	161	0	203	0.7
8. Road edge departure/no maneuver	95	259	579	1,388	5,962	39	8,323	29.4
9. Road edge departure/backing	3	3	3	1	0	0	11	0.0
10. Animal/maneuver	0	0	0	0	0	0	1	0.0
11. Animal/no maneuver	41	48	69	199	30	0	388	1.4
12. Pedestrian/maneuver	0	0	1	1	15	0	18	0.1
13. Pedestrian/no maneuver	0	2	9	23	341	0	376	1.3
14. Cyclist/maneuver	0	0	1	4	0	1	6	0.0
15. Cyclist/no maneuver	0	3	13	38	0	0	54	0.2
16. Backing into vehicle	0	0	0	0	0	0	0	0.0
17. Turning/same direction	0	0	0	0	0	0	0	0.0
18. Parking/same direction	0	0	0	0	0	0	0	0.0
19. Changing lanes/same direction	2	6	15	13	145	0	181	0.6
20. Drifting/same lane	0	0	0	0	0	0	0	0.0
21. Opposite direction/maneuver	0	1	3	5	85	0	95	0.3

Single-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost	Cost (%)
22. Opposite direction/no maneuver	12	27	42	178	19	0	277	1.0
23. Rear-end/striking maneuver	0	0	1	0	0	0	1	0.0
24. Rear-end/lead vehicle accelerating	0	0	0	0	0	0	0	0.0
25. Rear-end/LMV	0	0	7	4	0	0	11	0.0
26. Rear-end/LVD	1	0	5	13	0	0	19	0.1
27. Rear-end/LVS	1	2	6	5	0	0	15	0.1
28. LTAP/OD @ signal	0	0	0	0	0	0	0	0.0
29. Turn right at signal	0	0	0	0	0	0	0	0.0
30. Left turn across path/ opposite direction at non-signal	0	0	0	0	0	0	0	0.0
31. SCP @ non-signal	0	0	0	0	0	0	0	0.0
32. Turn @ non-signal	0	0	2	5	0	0	7	0.0
33. Evasive maneuver/maneuver	0	0	0	4	0	0	4	0.0
34. Evasive maneuver/no maneuver	3	2	5	24	0	3	37	0.1
35. Rollover	1	8	6	40	49	0	103	0.4
36. Non-collision—no impact	3	2	6	19	248	0	278	1.0
37. Object contacted/maneuver	1	0	2	0	85	0	88	0.3
38. Object contacted/no maneuver	9	7	20	52	427	0	515	1.8
Total	N/A	N/A	N/A	N/A	N/A	N/A	28,292	100.0

Table 121. Annual weighted frequency of intersection crashes by injury type.

Single-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost (%)
1. No driver present	0	60	0	0	0	0	0.1
2. Vehicle failure	353	24	31	32	0	0	0.4
3. Control loss/vehicle action	5,843	691	1,004	352	100	26	22.0
4. Control loss/no vehicle action	2,515	483	336	275	27	26	7.9
5. Running red light	0	0	0	0	0	0	0.0
6. Running stop sign	1,224	104	337	117	3	0	2.3
7. Road edge departure/maneuver	3,775	545	573	231	2	6	4.3
8. Road edge departure/no maneuver	4,048	829	1,066	614	63	42	18.2
9. Road edge departure/backing	65	0	7	0	0	0	0.0
10. Animal/maneuver	140	0	10	13	0	0	0.1
11. Animal/no maneuver	559	42	0	13	0	0	0.2
12. Pedestrian/maneuver	135	143	380	127	5	0	2.7
13. Pedestrian/no maneuver	0	282	366	328	92	0	18.0
14. Cyclist/maneuver	100	317	779	192	9	3	4.8
15. Cyclist/no maneuver	66	748	804	363	13	74	7.5
16. Backing into vehicle	0	0	0	0	0	0	0.0
17. Turning/same direction	0	0	0	0	0	0	0.0
18. Parking/same direction	0	0	0	0	0	0	0.0
19. Changing lanes/same direction	250	0	24	0	0	0	0.1
20. Drifting/same lane	0	0	0	0	0	0	0.0
21. Opposite direction/maneuver	0	61	0	18	0	0	0.2
22. Opposite direction/no maneuver	277	0	100	30	0	0	0.5
23. Rear-end/striking maneuver	0	27	6	0	0	0	0.0
24. Rear-end/lead vehicle accelerating	0	0	0	0	0	0	0.0
25. Rear-end/LMV	68	0	0	9	0	0	0.1

Single-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost (%)
26. Rear-end/LVD	35	62	27	10	0	0	0.2
27. Rear-end/LVS	226	75	15	1	0	0	0.1
28. LTAP/OD @ signal	0	0	0	0	0	0	0.0
29. Turn right at signal	0	0	0	0	0	0	0.0
30. Left turn across path/opposite direction at non-signal	0	0	0	0	0	0	0.0
31. SCP @ non-signal	393	87	44	61	0	0	0.7
32. Turn @ non-signal	638	94	200	73	0	0	1.1
33. Evasive maneuver/maneuver	164	0	3	3	0	0	0.0
34. Evasive maneuver/no maneuver	192	0	20	13	0	0	0.2
35. Rollover	68	23	91	49	0	0	0.6
36. Non-collision—no impact	212	6	57	61	0	0	0.6
37. Object contacted/maneuver	609	0	0	3	0	0	0.1
38. Object contacted/no maneuver	247	118	45	88	38	0	6.9
Total	N/A	N/A	N/A	N/A	N/A	N/A	100.0

Table 122. Annual cost of intersection crashes by injury type (millions of dollars).

Single-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost	Cost (%)
1. No driver present	0	3	0	0	0	0	3	0.1
2. Vehicle failure	2	1	3	9	0	0	14	0.4
3. Control loss/vehicle action	27	29	83	100	615	3	857	22.0
4. Control loss/no vehicle action	12	20	28	78	167	3	308	7.9
5. Running red light	0	0	0	0	0	0	0	0.0
6. Running stop sign	6	4	28	33	18	0	90	2.3
7. Road edge departure/maneuver	17	23	48	66	15	1	169	4.3
8. Road edge departure/no maneuver	19	35	89	175	387	5	709	18.2
9. Road edge departure/backing	0	0	1	0	0	0	1	0.0
10. Animal/maneuver	1	0	1	4	0	0	5	0.1
11. Animal/no maneuver	3	2	0	4	0	0	8	0.2
12. Pedestrian/maneuver	1	6	32	36	31	0	105	2.7
13. Pedestrian/no maneuver	0	12	30	93	565	0	701	18.0
14. Cyclist/maneuver	0	13	65	55	53	0	187	4.8
15. Cyclist/no maneuver	0	32	67	103	82	9	293	7.5
16. Backing into vehicle	0	0	0	0	0	0	0	0.0
17. Turning/same direction	0	0	0	0	0	0	0	0.0
18. Parking/same direction	0	0	0	0	0	0	0	0.0
19. Changing lanes/same direction	1	0	2	0	0	0	3	0.1
20. Drifting/same lane	0	0	0	0	0	0	0	0.0
21. Opposite direction/maneuver	0	3	0	5	0	0	8	0.2
22. Opposite direction/no maneuver	1	0	8	8	0	0	18	0.5
23. Rear-end/striking maneuver	0	1	0	0	0	0	2	0.0

Single-Vehicle Pre-Crash Scenario	0 No Injury	1 Possible Injury	2 Non-Incapacitating Injury	3 Incapacitating Injury	4 Fatal	5 Injured, unknown Severity	Cost	Cost (%)
24. Rear-end/lead vehicle accelerating	0	0	0	0	0	0	0	0.0
25. Rear-end/LMV	0	0	0	2	0	0	3	0.1
26. Rear-end/LVD	0	3	2	3	0	0	8	0.2
27. Rear-end/LVS	1	3	1	0	0	0	6	0.1
28. LTAP/OD @ signal	0	0	0	0	0	0	0	0.0
29. Turn right at signal	0	0	0	0	0	0	0	0.0
30. Left turn across path/ opposite direction at non-signal	0	0	0	0	0	0	0	0.0
31. SCP @ non-signal	2	4	4	17	0	0	27	0.7
32. Turn @ non-signal	3	4	17	21	0	0	44	1.1
33. Evasive maneuver/maneuver	1	0	0	1	0	0	2	0.0
34. Evasive maneuver/no maneuver	1	0	2	4	0	0	6	0.2
35. Rollover	0	1	8	14	0	0	23	0.6
36. Non-collision—no impact	1	0	5	17	0	0	23	0.6
37. Object contacted/maneuver	3	0	0	1	0	0	4	0.1
38. Object contacted/no maneuver	1	5	4	25	232	0	267	6.9
Total	N/A	N/A	N/A	N/A	N/A	N/A	3,892	100.0

ACKNOWLEDGEMENTS

The project team would like to express appreciation of the following organizations for providing information used in developing this study:

- UNC Highway Safety Research Center.
- UDOT.
- WSDOT.

REFERENCES

1. Eigen, A.M., Valdivieso, R.O., and Ahrari, A. (2014). *Photographic Data Extraction Feasibility and Pilot Study in Support of Roadside Safety and Roadway Departure Research*, Report No. FHWA-HRT-13-088, Federal Highway Administration, Washington, DC.

2. Evans, L. (2004). *Traffic Safety,* Science Service Society, Bloomfield Hills, MI.

3. Hauer, E. (2005). *The Road Ahead*, Journal of Transportation Engineering, Vol. 131, No. 5, Elsevier, Kidlington, Oxford, UK, pp. 333–339.

4. Shinar, D. (2007). *Traffic Safety and Human Behavior,* Elsevier, Kidlington, Oxford, UK.

5. Herbel, S., Laing, L., and McGovern, C. (2010). *Highway Safety Improvement Program Manual*, FHWA-SA-09029, Federal Highway Administration, Washington, DC.

6. Treat, J.R., Tumbas, N.S., McDonald, S.T., Dhinar, D., Hume, R.D., Mayer, R.E., Stansifer, R.L., and Castellan, N.J. (1979). *Tri-level Study of the Causes of Traffic Crashes: Final Report - Executive Summary*, Report No. DOT-HS-034-3-535-79-TAC(S), Institute for Research in Public Safety, Bloomington, IN.

7. Gross, F., Persaud, B., and Lyon, C. (2010). *A Guide to Developing Quality Crash Modification Factors*, FHWA-SA-10-032, Federal Highway Administration, Washington, DC.

8. Shinar, D. (1998). *Speed and Crashes: A Controversial Topic and an Elusive Relationship*. Appendix B of TRB Special Report 254, Transportation Research Board of the National Academies of Science, Washington, DC, pp. 221–276.

9. McCarthy, P. (1998). *Effect of Speed Limits on Speed Distributions and Highway Safety: A Survey of the Literature*. Special Report 254: Managing Speed, Transportation Research Board of the National Academies of Science, Washington, DC, pp. 227–358.

10. Davis, G. (2004). *Possible Aggregation Biases in Road Safety Research and a Mechanism Approach to Accident Modeling*, Accident Analysis and Prevention, Vol. 36, Elsevier, Kidlington, Oxford, UK, pp. 1119–1127.

11. Transportation Research Board. (2009). *Theory, Explanation and Prediction in Road Safety: Identification of Promising Directions and a Plan for Advancement*, Task Force for the Development of a Highway Safety Manual, Workshop Circular, Transportation Research Board, Washington, DC.

12. Head, J.A. (1959). *Predicting Traffic Accidents from Roadway Elements on Urban Extensions of State Highways*, Highway Research Board Bulletin No. 208, Washington, DC, pp. 45–63.

13. Belmont, D.M. (1954). *Effect of Shoulder Width on Accidents on Two-Lane Tangents,* Highway Research Board Bulletin No. 91, Washington, DC, pp. 29–32.

14. Blensly, R.C. and Head, J.A. (1960). *Shoulders and Accident Experience on Two-Lane, Rural Highways: A Summary*, Highway Research Board Bulletin No. 266, Washington, DC, pp. 28–33.

15. Perkins, E.T. (1957). *Relationship of Accident Rate to Highway Shoulder Width*, Highway Research Board Bulletin No. 151, pp. 13–14, Washington, DC.

16. Zegeer, C.V., Deen, R.C., and Mayes, J.G. (1981). *Effect of Lane and Shoulder Width on Accident Reduction on Rural, Two-Lane Roads*, Transportation Research Record No. 806, Transportation Research Board of the National Academies of Science, Washington, DC, pp. 33–43.

17. Griffin, L.I. and Mak, K.K. (1987). *The Benefits to Be Achieved from Widening Rural, Two-Lane Farm-to-Market Roads in Texas,* Texas Transportation Institute, Report No. IAC (86-87)–1039, College Station, TX.

18. Gross, F., Jovanis, P.P., Eccles, K.A., and Chen, K.Y. (2009). *Safety Evaluation of Lane and Shoulder Width Combinations on Rural, Two-Lane, Undivided Roads,* FHWA-HRT 09-031, Federal Highway Administration, Washington, DC.

19. Le, T.Q. and Porter, R.J. (2013). *"Safety Effects of Cross Section Design on Urban and Suburban Roads."* Presented at the Compendium of Papers from the 92nd Annual Meeting of the Transportation Research Board, Washington, DC.

20. Utah Department of Transportation (2015). "Data portal." (website) Salt Lake City, UT. Available online: http://maps.udot.utah.gov/ugate/f?p=111:2:0::NO:::, last accessed August 2016.

21. Insurance Institute for Highway Safety - Highway Loss Data Institute. (2008). *Roadway and Environment. Roadway and Environment Topics, Highway Safety Research & Communications*, IIHS-HLDI, Arlington, VA.

22. Campbell, K.L. (2012). "The SHRP 2 Naturalistic Driving Study: Addressing Driver Performance and Behavior in Traffic Safety." *TR News: Blueprints to Improve Highway Safety Number 282*, September–October 2012, Transportation Research Board of the National Academies of Science, Washington, DC.

23. Federal Highway Administration. (2016). "SHRP2Solutions: Tools for the Road Ahead." (website) McLean, VA. Available online: http://www.fhwa.dot.gov/goshrp2/, last accessed August 2016.

24. Snyder, M.B. and Knoblauch, R.L. (1971). *Pedestrian Safety: The Identification of Precipitating Factors and Possible Countermeasures*, FH-11-7312, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, DC.

25. Cross, K.D. and Fisher, G. (1977). *A Study of Bicycle/Other-Vehicle Accidents: Identification of Problem Types and Countermeasure Approaches*, DOT HS-803 315, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, DC.

26. Hendricks, D.L. et al. (1999). *The Relative Frequency of Unsafe Driving Acts in Serious Traffic Crashes. Summary Technical Report*, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, DC.

27. Davis, G.A. (2002). *"Towards a Unified Approach to Causal Analysis in Traffic Safety Using Structural Causal Models*." Presented at the Transportation and Traffic Theory in 21st Century: Proceedings from the 15th International Symposium on Transportation and Traffic Theory, Adelaide, Australia, pp. 247–266.

28. American Association of State Highway and Transportation Officials (AASHTO). (1996). *Roadside Design Guide, 2nd Edition*, ISBN 1-56051-031-5, AASHTO, Washington, DC.

29. Mak, K.K and Sicking, D.L. (2003). *Roadside Safety Analysis Program (RSAP)—Engineer's Manual.* NCHRP Report 492, Transportation Research Board of the National Academies, Washington, DC.

30. Hollnagel, E. (2004). *Barriers and Accident Prevention*, ISBN-13: 978-0754643012, Ashgate Publishing Ltd., Farnham, Surrey, UK.

31. Levenson, N. (2004). "A New Accident Model for Engineering Safer Systems." *Safety Science, Vol. 42*, pp. 237–270, Elsevier, Kidlington, Oxford, UK.

32. Najm, W.G., Smith, J.D., and Yanagisawa, M. (2007). *Pre-Crash Scenarios Typology for Crash Avoidance Research*, DOT HS 810 767, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, DC.

33. Hauer, E. (2010). "Cause, Effect, and Regression in Road Safety: A Case Study." *Accident Analysis and Prevention, Vol. 42*, p. 4, Elsevier, Kidlington, Oxford, UK.

34. Shinar, D., Treat, J.R., and McDonald, S.T. (1983). "The Validity of Police-Reported Accident Data." *Accident Analysis and Prevention, Vol. 15*, pp. 175–191, Elsevier, Kidlington, Oxford, UK.

35. Austin, K. (1995). "The Identification of Mistakes in Road Accident Records (Part 1, Locational Variables and Part 2, Casualty Variables)." *Accident Analysis and Prevention, Vol. 27*, pp. 261–282, Elsevier, Kidlington, Oxford, UK.

36. Mynatt, M., Bean, J., Kahane, C.J., Rush, C., Traube, E., and Wiacek, C. (2011). *"A Study of the NMVCCS to Identify Critical Precrash Factors in Fatal Crashes*." Presented at the 22nd International Technical Conference on the Enhanced Safety of Vehicles (ESV), No. 11-0168, Washington, DC.

37. Lord, D. and Mannering, F. (2010). "The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives." *Transportation Research Part A, Vol. 44*, pp. 291–305, Elsevier, Kidlington, Oxford, UK.

38. Elvick, R. (2011). "Publication Bias and Time-Trend Bias in Meta-Analysis of Bicycle Helmet Efficacy: A Re-Analysis of Attewell, Glase, and McFadden, 2001." *Accident Analysis and Prevention, Vol. 60*, pp. 245–253, Elsevier, Kidlington, Oxford, UK.

39. American Association of State Highway and Transportation Officials (AASHTO). (2010). *Highway Safety Manual*. AASHTO, Washington, DC.

40. National Climatic Data Center - National Oceanic and Atmospheric Administration. (2015). "Land-Based Station Data." (website) Asheville, NC. Available online: http://www.ncdc.noaa.gov/data-access/land-based-station-data, last accessed March 13, 2016.

41. Caddell, R., Hammond, P., and Reinmuth, S. (2009). "Roadside Features Inventory Program." (website) Tacoma, Washington. Available online: http://www.ce.siue.edu/faculty/hzhou/Highway%20Inventory%20Refereces%5CRFIP_Morin.pdf, last accessed November 2014.

42. "US-6, Helper, UT 84526." Map. Google Maps. Helper, Utah. March 13, 2016. Accessed March 13, 2016. https://goo.gl/maps/h5kjS8ojbsm. Scale: 1:1,200 ft

43. "US-50, Utah." Map. Google Maps. Utah. March 13, 2016. Accessed March 13, 2016. https://goo.gl/maps/rfETJwTcvxA2. Scale 1:12,000 ft

44. "US-6, Helper, UT 84526." Map. Google Maps. Helper, UT. March 13, 2016. Accessed March 13, 2016. https://goo.gl/maps/A1NXSvzKq7K2. Scale 1:2,400 ft

45. "UT-10, Salina, UT 84654." Map. Google Maps. Salina, UT. March 13, 2016. Accessed March 13, 2016. https://goo.gl/maps/XNeDvjVCnoQ2. Scale 1:2,400 ft

46. "Wasatch County, Utah." Map. Google Maps. March 13, 2016. Accessed March 13, 2016. https://goo.gl/maps/nGKzLE9xnxv. Scale 1:2,400 ft

47. Himes, S., Donnell, E., and Porter, E. (2013). "Posted Speed Limit to Include or Not to Include in Operating Speed Models." *Transportation Research Part A, Vol. 52*, pp. 23–33, Elsevier, Kidlington, Oxford, UK.

48. Mitra, X. and Washington, Y. (2012). "On the Significance of Omitted Variables in the Intersection Crash Modeling." *Accident Analysis and Prevention, Vol. 49*, pp. 439–448.

49. Hauer, E. (2013). *"Safety Performance Functions: A Workshop: Summer 2013."* Presented at the Transportation Research Board 92nd Annual Meeting, Washington, DC.

50. Kutner, M., Nachtsheim, C., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models: Fifth Edition*, ISBN 0-07-238688-6, McGraw-Hill Irwin, New York City, NY.

51. Berk, R., and MacDonald, J. (2008). "Overdispersion and Poisson Regression." *National Science Foundation, Journal of Quantitative Criminology, Vol. 4,* pp. 289–308, University of Pennsylvania, Philadelphia, PA.

52. Agresti, A. (2002). *Categorical Data Analysis: Second Edition,* ISBN 0-471-36093-7, John Wiley & Sons, Inc., Hoboken, NJ.

53. Ha, J.-A. and Oh, J.-S. (2014). "Estimating Annual Average Daily Traffic Using Daily Adjustment Factor." *Journal of Emerging Trends in Computing and Information Sciences, Vol. 5*, No.7, pp. 580–587, University of Quebec, Chicoutimi, Quebec, Canada.

54. Gadda, S., Magoon, A., & Kockelman, K. M. (2007). *"Estimates of AADT: Quantifying the Uncertainty."* Presented at the 86th Annual Meeting of the Transportation Research Board, Transportation Research Board, Washington, DC.

55. Selby, B. and Kockelman, K. M. (2011). *"Spatial Prediction of AADT in Unmeasured Locations by Universal Kriging."* Presented at the 90th Annual Meeting of the Transportation Research Board, Transportation Research Board, Washington DC.

56. SunriseSunset. (2015). "Sunrise Sunset Calendar—Selected Major World Cities and Places of Interest." (website). Available online: http://www.sunrisesunset.com/calendar.asp, last accessed October 2015.

57. timeanddate.com (2016). "Salt Lake City, UT, USA — Sunrise, Sunset, and Moon Times for Today." (website). Available online: http://www.timeanddate.com/astronomy/usa/salt-lake-city, last accessed October 2015.

58. Barca, E., Passarella, G., and Uricchio, V. (2008). "Optimal Extension of the Rain Gauge Monitoring Network of the Apulian Regional Consortium for Crop Protection." *Environmental Monitoring and Assessment, Vol. 145,* Issue 1–3, pp. 375–386, Springer Netherlands, Dordrecht, Netherlands.

59. Robertson, G. P. (2008). *GS+ User's Guide Version 10: Geostatistics for the Environmental Sciences,* Gamma Design Software, Plainwell, MI.

60. Oliver, M. A. and Webster, R. (2013). *A Tutorial Guide to Geostatistics: Computing and Modelling Variograms and Kriging,* CATENA 113, Elsevier, Kidlington, Oxford, UK, pp. 57–69.

61. Adhikary, S. K., Yilmaz, A. G., and Muttil, N. (2015). "Optimal Design of Rain Gauge Network in the Middle Yarra River Catchment." *Hydrological Processes, Vol. 29*, John Wiley & Sons, Inc., Hoboken, NJ, pp. 2582–2599.

62. Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics,* ISBN 0-19-505012-6, Oxford University Press, New York, NY.

63. Kirk, A. and N. Stamatiadis (2001). *Evaluation of the Quasi-Induced Exposure: Final Report,* September 2001, Southeastern Transportation Center, Knoxville, TN.

HRDS-20/07-18(WEB)E