SAFETY DATA MANAGEMENT AND ANALYSIS: ADDRESSING THE CONTINUING EDUCATION NEEDS FOR THE PACIFIC NORTHWEST

PROJECT REPORT

by

Kevin Chang – University of Idaho
Shane Brown – Oregon State University
Cynthia Chen – University of Washington
Ali Hajbabaie – Washington State University
Robert Perkins – University of Alaska-Fairbanks

Sponsored by

PacTrans

for
Pacific Northwest Transportation Consortium (PacTrans)
USDOT University Transportation Center for Federal Region 10
University of Washington
More Hall 112, Box 352700
Seattle, WA 98195-2700

In cooperation with US Department of Transportation-Research and Innovative Technology Administration (RITA)

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The Pacific Northwest Transportation Consortium, the U.S. Government and matching sponsor assume no liability for the contents or use thereof.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Safety Data Management and Analysis: Addressing the Continuing Education Needs for the Pacific Northwest		5. Report Date March 30, 2017
		6. Performing Organization Code
7. Author(s) Kevin Chang – University of Idaho Shane Brown – Oregon State University Cynthia Chen – University of Washington Ali Hajbabaie – Washington State University Robert Perkins – University of Alaska-Fairbanks		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Idaho Department of Civil Engineering 875 Perimeter Drive, MS 1022 Moscow, ID 83844-1022		10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTRT13-G-UTC40
12. Sponsoring Agency Name and Address Pacific Northwest Transportation Consortium (PACTRANS) University of Washington More Hall 112, Box 352700 Seattle, WA 98195-2700		13. Type of Report and Period Covered 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Recent advancements in data collection capabilities have allowed transportation-related agencies to collect mountains of safety data. There is an immediate need to find out what types of safety data are being collected, what types of safety analysis can be done with the collected data, and what (other) types of safety data and analysis approaches are required to meet safety objectives.

With the increased complexity of various safety data management and analysis activities, and with most transportation agencies faced with limited staff and financial resources, there is opportunity to provide the transportation workforce, which includes practitioners and academicians alike, with the resources needed to effectively understand, manage, and analyze safety data. Safety data collection, management, integration, improvement, and analysis activities are integral to developing a robust data program that leads to more informed decision making, better targeted safety investments, and overall improved safety outcomes.

This project responds to the current gaps in research and identifies a methodology that will benefit all system users. The objectives included developing a comprehensive understanding of needs and priorities with regard to safety data management and analysis; developing a set of core skills and knowledge required for safety data management and analysis; providing a comprehensive set of safety data workforce development resources that can easily be accessed for use and distribution; and identifying and utilizing proven delivery pipelines to supplement program outreach efforts and activities in the safety data area.

17. Key Words Safety, safety data, workforce development	18. Distribution Statement No restrictions. Copies available from PACTRANS: www.pactrans.org

19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 89	22. Price

Contents

List of Figures

List of Tables

Acknowledgments

The authors wish to thank the PacTrans Regional University Transportation Center for its financial support to conduct this project and to the academicians and practitioners who provided their time and insight to make this research effort possible.

This study also would not have been completed without the significant contributions of Edinson Mejia Bautista (University of Idaho), Peng Chen (University of Washington), Edgar Castro (Oregon State University), and Rasool Mohebifard (Washington State University).

Executive Summary

There is an opportunity to provide the transportation workforce, which includes practitioners and academicians alike, with the resources needed to effectively understand, manage, and analyze safety data, given the increased complexity of safety data and safety data management and analysis activities, and that most transportation agencies are faced with limited staff and financial resources. Safety data collection, integration, and improvement activities are integral to developing a robust data program that leads to more informed decision-making, better targeted safety investments, and overall improved safety outcomes.

This collaborative research effort leveraged the cumulative expertise in transportation safety and transportation education of five institutions: the University of Alaska-Fairbanks, Oregon State University, Washington State University, the University of Washington, and the University of Idaho. As part of this study, user group surveys of practitioners and academicians were conducted, safety data were analyzed, and training tools and techniques were identified and developed.

Several key takeaways were identified as part of this study. From a practitioner standpoint, while the acquisition, flow, storage, and use of data are similar from state to state, many of the details are quite different, such as the data used for state highway projects and data used for local projects within a given state. The most common reported difficulty with the data seems to be crash locations, but all states recognize this and work to validate and modify, if appropriate, location data through some sort of quality control process. Several agencies may be involved in the data's gathering and compilation, and interagency cooperation and coordination are an important part of assuring accuracy and usability of the data. Despite the automated nature of data transfer from one agency to another, errors in the data or in interpretation are possible.

Within the research, practice, and education disciplines, the most common type of safety data used are collision data, as this type of data is accessible, relatable, and applicable. With regard to roadway inventory and traffic volume data, these two types of safety data are used on a regular basis in transportation practice, but less so in the teaching and research settings. Practitioners and academicians provided similar definitions of the term “big data” as a large collection of data compiled for use in transportation applications. However, practitioners explained the term by describing technical applications related to transportation and traffic engineering, while academicians explained the term by associating it with novel research topics such as connected vehicles and the use of real-time data.

From an analysis standpoint, by using different data sources, crash rates, roadway geometric characteristics, traffic data, and weather data, data can be fused to form a single data set for safety analysis. The data set can then be used to estimate the number and severity of crashes on various roadway segments. Roadway segmentation plays an important role in the accuracy of crash prediction models, but there is no consensus on the best way of segmentation or a threshold on short segments. Since crashes are rare events, data need to be aggregated over an extended duration of time. As such, dynamic variables, such as weather conditions and traffic flow rate, are converted to point estimations. This aggregation may introduce some error in crash prediction models. Furthermore, crash prediction models are prone to the under-reporting of less severe crashes. Additional detailed data on vehicle speed, acceleration/ deceleration rate, heading, and braking can be used to improve the accuracy of crash prediction models; however, such data attributes are still very expensive to collect. Connected vehicle technology is expected to greatly facilitate access to such data attributes.

Safety planning efforts and funding often start at the state level and trickle down to local and regional agencies. Analysis of and investment in state highways rather than local roads are often the result of several factors, including statewide priorities on higher volume and higher speed roadways (which typically represent the roadway network of the state department of transportation (DOT) itself), data limitations on local roads, the lack of safety planning expertise at local agencies, the lack of a champion in state DOTs for expanding opportunities for local agencies, and other requirements. This research examined both the challenges and current resources available to local agencies to address road safety. In the past three years, most of the agencies had received training in pedestrian and bicycle safety or complete streets, and the topics of project funding, project prioritization, and data analysis procedures were identified as the three highest needs. The data collected were also used to determine what challenges local agencies face in addressing road safety. Some agencies lack the funds needed to implement safety improvements on their road systems and look to the state Highway Safety Improvement Program (HSIP) or other funding sources for assistance. The increase in competition and the need to do more with less make choosing the right projects more important than ever. Local agencies often lack the safety data or analytical skills necessary to meet crash data analysis requirements. To stretch limited highway safety funding, local transportation agencies are encouraged to identify and implement the optimal combination of countermeasures to achieve the greatest benefits. However, a lack of local staff requires more multitasking and results in increased inefficiencies. One common challenge related to training is its lack of and uneven commitment to training among different organizations.

Although there has been a significant reduction in the number of collision-involved injuries and fatalities in recent years, the United States still faces unacceptably high collision rates in comparison to other developed countries. Road safety is a challenging and evolving field, and preparing students with expertise in road safety research is one important mission that cannot be underestimated by university educators to better solve the forthcoming road safety challenges. However, most domestic universities do not have an independent road safety course in their civil engineering departments. Student knowledge of road safety is built on scattered sessions that are provided by other transportation-related courses such as transportation engineering, planning, and freight and supply chains. Road safety is interdisciplinary in nature as it intersects civil engineering, psychology, mechanical engineering, urban planning, public health, and other disciplines. For these reasons, development of materials for a road safety course that could be used as part of a university curriculum is needed.

1. Introduction

Recent advancements in data collection capabilities have allowed transportation-related agencies to collect mountains of safety data. There is an immediate need to find out what types of safety data are being collected, what types of safety analysis can be done with the collected data, and what (other) types of safety data and analysis approaches are required to meet safety objectives.

Extensive collection efforts exist with regard to roadway, traffic, licensing, and vehicle data. For example, there are more than five million traffic crashes reported annually in the United States, and vehicle crashes on public highways alone result in over 840,000 injuries and 1,700 fatalities annually according to the National Highway Traffic Safety Administration (Driving Safety, 2014). The documentation process for every single one of these crashes must begin at the scene of the incident with information gathered by a member of the law enforcement community or by the private citizen(s) involved in the crash. This information is subsequently transmitted to a local or state agency for data entry, processing, and aggregation for the purpose of future analysis.

With the increased complexity of various safety data management and analysis activities, and with most transportation agencies faced with limited staff and financial resources, there is an opportunity to provide the transportation workforce, which includes practitioners and academicians alike, with the resources needed to effectively understand, manage, and analyze safety data. Safety data collection, management, integration, improvement, and analysis activities are integral to developing a robust data program that leads to more informed decision making, better targeted safety investments, and overall improved safety outcomes.

1.1 Objectives

This project responded to the current gaps in research and identified a methodology to benefit all system users. The objectives included the following:

- Develop a comprehensive understanding of needs and priorities with regard to safety data management and analysis.

- Develop a set of core skills and knowledge required for safety data management and analysis.
- Provide a comprehensive set of safety data workforce development resources that can easily be accessed for use and distribution.
- Identify and utilize proven delivery pipelines to supplement program outreach efforts and activities in the safety data area.

In doing so, this research effort supported regional transportation safety decision-making.

1.2 Approach/Methodology

This research collaboration leveraged the cumulative expertise in transportation safety and transportation education of five institutions: the University of Alaska-Fairbanks (UAF), Oregon State University (OSU), Washington State University (WSU), the University of Washington (UW), and the University of Idaho (UI). The project was divided into six tasks (see figure 1-1) as part of the work plan: (1) literature review and data assessment (led by UAF and OSU), (2) user group survey and assessment (UAF and OSU), (3) analysis of safety data (WSU), (4) development of recommendations for future data collection and application (WSU), (5) training tools (UI and UW), and (6) final reporting (all).

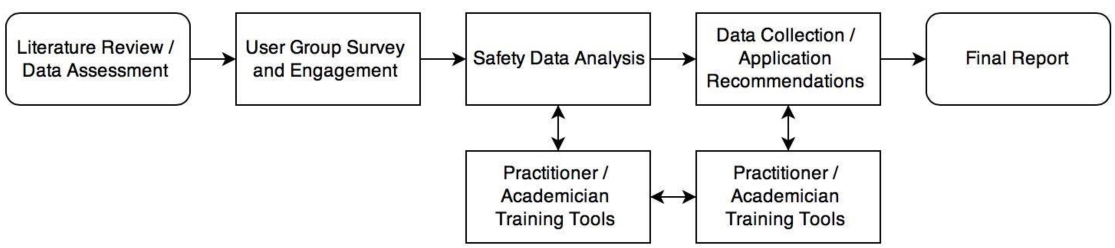

Figure 1-1: Project work plan

1.3 Task Descriptions and Milestones

Task 1: Literature Review and Data Assessment – A literature review was conducted to determine the current state of practice regarding safety data collection (i.e., extent, requirements) and reporting (i.e., to which agencies), with a particular focus on documenting the details of the process, including agency report development, safety data software applications, use of GIS, and other techniques to improve the quality of safety data. Existing curricular materials currently in use at academic institutions with regard to safety data were also reviewed.

Task 2: User Group Engagement – User groups were surveyed to determine safety data and safety data reporting needs. These groups included, but were not limited to, A) local and state agencies (i.e., how they currently work with and apply the data, and their training needs), B) transportation safety researchers and academicians (i.e., what data courses exist and what topics are covered), C) individuals (i.e., what does "big data" mean and how is "big data" being defined), and D) incident responders.

Task 3: Development and Analysis of Safety Data – Once the extent of safety data was determined, the attributes and characteristics of the collected data (e.g., aggregation level, duration) were determined. A particular area of focus identified the types of safety analysis that could be done with the existing data, as well as the types required to meet evolving safety objectives.

Task 4: Data Collection and Application Recommendations – As a logical progression step after Task 3, the characteristics of the data that were required for analyses were identified and the methods to collect missing data were proposed. In addition, additional methods to analyze data to meet safety goals were proposed.

Task 5: Training Tools – Training tools were developed in order to broadly disseminate the results of this research effort to practitioners and academicians. These tools, which can be used as part of continuing education training or classroom teaching at the university level, were thoughtfully designed to be both dynamic and self-sustainable so that there is value for those teaching subject matter related to safety data or safety data management. The modules include components such as definition of learning objectives, identification of reading materials, development of sample test and course content materials, and suggested teaching methods for successful content delivery.

Task 6: Final Reporting - The final report, which describes this research study, conclusions, and recommendations, is represented by this document.

The following chapters represent the work contained in this project. Chapter 2, authored by the University of Alaska-Fairbanks, discusses the safety data gathered from a practitioner perspective. Chapter 3, authored by Oregon State University, discusses the safety data gathered from an academic perspective. Chapter 4, authored by Washington State University, describes the development and analysis of safety data and recommends data collection and application practices. Chapter 5, authored by the University of Idaho, describes the processes used to develop training tools for practitioners and the learning objectives for each training module. Chapter 6, authored by the University of Washington, describes the processes used to develop training tools for academicians and the learning objectives for each training module.

2. Gathering and Using the Data (Practitioner Perspective)

How are roadway crash data acquired, stored, and utilized in engineering and management decisions regarding highway projects? This research answered that question by interviewing the engineers and professionals involved with safety data management from six states and by asking how safety (i.e., crash) data are acquired and used. Since most safety projects are funded by the FHWA through the Highway Safety Improvement Program (HSIP), the general flow of the safety data is similar in each state. Howver, there are many differences in details, especially in computer hardware and software. The methods of data movement between the responder and the department of transportation (DOT) often involve an intermediate agency, such as the department of motor vehicles (DMV), and this varies among states. Likewise, the program to extract these data for the DOT varies. Another pronounced difference is the transfer of HSIP funding to local agencies; also pronounced is the use of historical crash data in the safety performance functions (SPFs). The older method of looking at the crash data from only the location in question is not uncommon, while the more modern method of using data from similar locations via an Empirical Bayes (EB) analysis is becoming more common and is the currently recommended method. Most analysis software is geared to the EB analysis. Historical crash data, before and after countermeasures have been installed, may be used to evaluate SPFs and crash modification factors (CMFs) for particular states and localities, but there are practical problems with this application of crash data because of the time required to acquire adequate data for comparisons.

2.1 Introduction

The Highway Safety Improvement Program (HSIP), which provides federal funding for state and local highway projects, requires a data-driven, strategic approach to improving highway safety on all public roads. Our research focused on the management of safety data. We reviewed how

- Highway safety data are gathered, and
- Data are used in project selection for highway projects.

Our research was based on interviews with traffic and engineering professionals in the states associated with PacTrans and other states. We hope this research will inform discussion of improvements in academic and practical transportation engineering education.

2.1.1 How Data Are Gathered and Stored

Although details vary quite a bit, most states have a similar process of converting a crash report that is initiated by a responder or citizen into a database entry. This database may be controlled by a non-DOT agency, such as the DMV, or the DOT, and either or both may perform quality control (QC) on the entry. Variations may be important, such as whether all crashes are included or only those on certain roads. However, these variations are known by the data users and arise because of administrative rules usually beyond the data users' control. We will refer to this answer to question 1) above as "gathering and storage" and the unit that does it as the "Data Office."

1. Responders generate a report of the crash with data.

   Most states have a preferred electronic system for the responder (hereafter, police, although the responder could be someone with the title sheriff or another law enforcement officer) to enter the data at the scene of a crash. All states have many exceptions: some reports are still hand written, and some police agencies within a state use different systems or programs. Some reports are filed by civilians. All initial police reports are reviewed by their office or the DOT/DMV and revised as needed. Accurate accident location is a frequent problem. Some police cars have a GPS-based location system, but the location is of the police car at the time the report is written, which may not be the crash location.

2. The report goes to a central processing center where data are entered into a generic database. In some states this processing is by the DMV or other non-DOT agency. All states have different software systems for this process. The names are a bewildering mix of acronyms and brand names. Some systems are tightly managed by the DOT, while others are managed by the DMV or another entity such as the Department of Revenue. Fatal accident information also goes into the Fatal Accident Reporting System (FARS), then to the National Highway

Transportation Safety Board (NHTSB), providing more detailed information than for non-fatal crashes. Annually, an HSIP report must be generated by the DOT. Safety data are an important part of that report.

2.1.2 How Data Are Used

Selection of safety improvement projects may be summarized into two fundamental processes:

- Identification of roadway sections and intersections for safety ranking, and
- Candidate safety project selection.

Three sets of distinctions are helpful for examining processes A and B. First, most states have two levels: a headquarters (HQ) and several subordinate regions or districts. Generally, both levels are involved in such decisions. Second, within both HQ and the regions, there is a "traffic and safety" (T&S) division and an "engineering" or "project management" division (PM). Although they are sometimes divided— with engineering handling design and construction and PM handling approval and funding— we will combine engineering with project management. The third distinction is between "DOT" and "local agencies." These latter may be cities, counties, or other entities eligible for federal funding. We will note here that there are safety-related projects that are not HSIP or federally funded, but regarding data, we will not spend much time on these; usually they are a small component of the data-driven safety projects. Finally, there are non-safety-related projects, but again, regarding safety data, if the project has both safety and non-safety, the safety is separated.

Data Use in Location Identification and Ranking

Starting with the crash data in the database controlled by the Data Office, those data are parsed with an extraction program that converts the raw data into information by compiling reports. This may involve an Extract Translate Load (ETL) program or may be done manually. The reports may be automatic, and these are usually summarized for the entities that may need them, a county or region, for example. In addition, users may query the data extraction program for special reports such as data on crashes at a specific intersection location or roadway segment. These users may be T&S or PM staff at either the HQ or region, but some administrative barriers may be in place since some programs are more flexible than others. However, the database may be queried more generally, such as, "list all the roadway intersections with more than one fatal accident in the last five years." The information may be output in graphic display or in Excel files.

The output from the ETL program provides the basis for ranking candidate locations; however, the rankings depend on the queries and algorithms used. The output of the extraction program can express the severity of the safety problem at a location. Often the region T&S takes the lead on this processing.

Data Use in Safety Project Selection

Depending on the program, the processed data may rank and prioritize sections of roadway or intersections, or these may be manually selected. The program may suggest countermeasures, or the engineering staff may propose them.

Often the reports from the extraction program will have many more projects than can be evaluated in detail, so the next step is a "preliminary scoping" of each project. PM will do a rough order of magnitude (ROM) estimate and consider practicalities such as other ongoing projects. Also, at this point, non-crash-data input will be considered, such as citizen complaints and operations and maintenance (O&M) reports.

Once the data set has been pared to a manageable number of projects, the next step is to look at each candidate project and select the "countermeasures" that may improve safety. Each countermeasure will likely have a different anticipated cost, as well as a different anticipated benefit. Thus a benefit/cost (B/C) is developed for each countermeasure. The anticipated costs are estimated on the basis of the historical costs of the various construction options, delays for the public, etc., as well as project-specific issues, such as right of way (ROW). The anticipated benefits are usually a reduction in anticipated crashes and/or their severity. This reduction may be estimated by staff from experience, but the standard method today is the use of a crash modification factor (CMF; see Appendix D.3) supplied by the Highway Safety Manual (HSM; see Appendix D.1). If the project involves both safety and non-safety attributes, the safety will be separated for the purpose of the HSIP B/C analysis.

The foregoing will determine both which countermeasure has the best B/C ratio and the likely cost of that countermeasure. At this level of review, PM will bring other departments within the DOT, such as environmental and ROW, into the process. Finally, a ranking of projects and tentative selection is sent to HQ for final approval. Often there are two layers of review, one by HQ T&S to verify that the project is approvable for HSIP funding, and one by HQ managers to verify overall resources and conformance to DOT mission and objectives.

Besides selection of projects, the processed data are used in HSIP annual reports and in planning decisions that rely on HSIP funding. (Appendix D.5 has an overview of nationwide summary crash data drawn from state HSIP reports.) The processed data may be used in before/after evaluations and to update the Safety Performance Functions (SPF; see Appendix D.2) and CMFs for that state or a particular region.

2.2 Summary Tables

In this section we present our findings in three summary tables: a) Data collection and processing, b) QC methods, and c) How ranking is accomplished. After the processing and QC, the reviewed data are accessible to the DOT. Some are loaded in processing programs and other data are stored, waiting to be loaded.

The third table summarizes the step in which DOT staff must query the data—i.e., ask a question. There are two main questions:

1. Safety Ranking. Of the roadways (or intersections or corridors) in the state, which are in most need of safety improvements in the form of accident counter measures?
2. Benefit/Cost Evaluation. For this particular project, what is the benefit in safety from this countermeasure in comparison to the cost?

Likewise, there are several administrative approaches to answering those questions, mainly:

1. Does a special group evaluate the projects on the basis of queries from the project-level staff; or
2. Do project-level staff evaluate the data?

Federal HSIP funding is assigned to the state DOT, but then some of the funding may be distributed to local agencies, usually cities or counties. In many states this leads to two different project selection and evaluation processes, depending on the agency that will administer the project.

All states have a Strategic Highway Safety Plan (SHSP) and a State Transportation Improvement Program (STIP); however, there are differences among states in how prescriptive the safety plan is. But in any case, the ranking must correspond to the plan. All projects must be listed in the STIP for public review. Usually local agencies, cities, and counties, as well as municipal planning organizations (MPOs), are queried regarding projects that affect them. Likewise, states are variable in the relations between their DOT districts and headquarters. Typically, approvals of HSIP funding are made at HQ, but often the B/C and recommendations are made at the region level.

Since most funding is related to the HSIP process, here we relate the use of safety data to the HSIP process and list other funding and processes as exceptions.

Table 2-1. Data collection and processing

State	Initial	Initial Storage	Database	Query Database	For Traffic Safety Design
Alaska OLD (currently converting systems)	Toughbook, Forms 200, 209	DMV, Crash Data Repository (CDR)	TraCE (Traffic Records Program Traffic and Criminal Software), Oracle	HAS (Highway Analysis System)	Access, Excel
Alaska NEW	Ditto	Ditto	Ditto	CARE (Critical Analysis Reporting Environment)	Direct access
WASH State Highways	Statewide Electronic Collision and Ticketing Online Records (SECTOR), supplied by DOT	DOT	Crash Location and Analysis System (CLAS)	SafetyAnalyst (which replaces Collision Analysis Location (CAL) and Collision Analysis Corridor (CAC), as well as Intersection Analysis Locator (IAL))	Direct
WASH Local	DITTO	Ditto	Ditto	Local Programs, Excel	Excel and notice to local agencies
Oregon	Paper forms sent to DMV	Crash Analysis Reporting (CAR) Unit within DOT gets reports from DMV and inputs (by hand) into FARS and DOT	SQL (IBM software, Structured Query Language)		SPIS (Safety Priority Index System)

Idaho	IMPACT (Idaho Mobile Program for Accident CollecTion)	DOT district level.	downloaded daily to CIRCA (Crash Information Retrieval Collection and Analysis)	WebCARS, access	
Colorado	All agencies use common form, half electronic, half scanned paper	Reports go to the Department of Revenue (DoR) database.	DOT imports all data into its own database and removes some unneeded data (i.e., personal ID, VIN, etc)	Vision Zero Suite, developed by DiExSys, to do 1) query and 2) analysis	Vision Zero Suite
Arizona	About 59% of crash data received in 2015 were in electronic form, and the rest were hard copies. All hard-copy crash reports are scanned and stored digitally since July 2012.	Traffic and Criminal Software (TraCS), other electronic tools, and hard-copy reports.	Arizona DOT's central crash database, called Accident Location Identification and Surveillance System (ASLISS). Roadway, traffic, and other safety data elements reside in different databases.	Safety Data Mart (SDM) online tool after QC/QA is completed.	Anyone with access to SDM can query crash data.

Table 2-2. Quality control (QC) methods

	Initial	Initial Storage	QC and Location
Alaska OLD	Toughbook, Forms 200, 209	DMV, CDR (Crash Data Repository)	All records include geo locations; often difficult to determine correct location from citizen reports. Crash staff are responsible for geolocating all records into DOT&PF GIS system.
Alaska NEW	Ditto	Ditto	
WASH State Highways	Statewide Electronic Collision and Ticketing Online Records (SECTOR), supplied by DOT	DOT	85% of reports are generated electronically. For the other 15%, the paper reports are entered into the database by State Patrol, not DOT. SECTOR is the capture system provided free by WA DOT to patrol officers. It's not required, but whatever system they use must abide by the same business rules. Reports come to the Data Office in real time. They validate the data. If accepted, they assign a 7-digit code and send it back to the officer. It is then transmitted as a "collision message" to State Patrol back office, local records management system, and Division of Licensing.
WASH Local	DITTO		
Oregon	Paper forms sent to DMV	Crash Analysis Reporting (CAR) Unit within DOT gets reports from DMV and inputs (by hand) into FARS and DOT	Crash Analysis Reporting (CAR) Unit within DOT gets reports from DMV and inputs (by hand) into FARS and DOT database, performing QC for proper location, etc., in the process.
Idaho	IMPACT Idaho Mobile Program for Accident CollecTion	DOT district level	Within DOT, the IMPACT data are downloaded daily to CIRCA at the district level. There are 6 districts, and each gets about the same number of downloads. At this point, QC is performed on location, weather, cause, etc. If questions, officer is asked to clarify.

Colorado	Electronic and paper	Dept. of Revenue (DoR) database	When DOT imports from DoR, DOT does QC and also removes some unneeded data.
Arizona	Electronic and hard copies	TraCS, other electronic tools, and hard-copy reports.	Statewide crash data goes through a QC/QA process by Arizona DOT's Traffic Records Section. In most cases, location or crash type errors are detected and corrected before accepting them in the ALISS database.

Table 2-3. How ranking is accomplished

	Query Database	**For Traffic Safety Design**	**Notes**
Alaska OLD	HAS (Highway Analysis System)	Access, Excel	
Alaska NEW	CARE	Direct	Regions perform a Benefit/Cost analysis (B/C) based on Crash Reduction Factors.
WASH State Highways	SafetyAnalyst	Direct	Regions analyze the countermeasures and do a B/C using SafetyAnalyst, which is based on CMF.
WASH Local	Local Programs, Excel	Excel and notice to local agencies	Counties and some cities use risk-based selection, generally of lower-cost projects. Local Programs (HQ level department) does analysis to compare alternatives. Larger agencies (usually cities) do an analysis based on CMFs. Either must be approved by Local Programs.
Oregon		SPIS (Safety Priority Index System)	Funding is apportioned to districts by their crash rate/history. Each district then does a B/C for projects. Cut off at funding limit.
Idaho	WebCARS accesses CIRCA	TREDIS	District engineers select candidate projects, do B/C, and then submit to HQ. Funding is competitive between districts.
Colorado	DOT database	Vision Zero Suite	HQ T&S first provides the 5 regions scans of areas of higher-than-expected crashes and also indicates whether a correctable fix is available. Regions submit applications to HQ T&S, where they use Vision Zero Suite to do B/C analyses, including the use of CMFs. Then HQ T&S does a ranking. No committee is involved at this point. There is competition within the regions based on B/C analyses and resulting rankings. About 50% is available for local projects, which are included in regional requests.

Arizona	Safety Data Mart	The Arizona Strategic Highway Safety Plan (SHSP) used a set of criteria to identify emphasis areas by crash characteristics. HSIP identifies primarily lane departure and intersectionrelated hot spots using network screening.	In transition, moving from traditional approach using historic crash data to predictive analysis for network screening, project-level safety evaluation, design alternative selection, Benefit/Cost analysis, and other factors.

2.3 State Details

2.3.1 Alaska

Police reports, using Form 200 (see Appendix E.2), record damage of $2000 or more, or any injury. Citizen reports, limited by insurance companies to $500±, use Form 209 (see Appendix E.1).

Citizen reports are predominantly used to report property-damage-only crashes. An officer may send a driver to the website for the Form 209. The driver can print the pdf, fill it in by hand, sign it, and send it to the DMV, where it is placed in a file and scanned. The DMV collects crash reports but does not parse the data. The DMV sends a copy of any crash report to the Department of Transportation and Public Facilities (DOT&PF) either electronically, as a pdf, or as a hard copy. The DOT&PF staff and contract personnel enter a subset of data from the report into the Oracle database, either electronically or by hand, and they also geolocate the data.

There are four formats for Form 200 crash data:

1. TraCS (electronic), which flows directly into the DMV and DOT&PF's systems
2. Fairbanks police (electronic), which flows directly into the DMV and DOT&PF's systems
3. Anchorage police, which is generated into a pdf uploaded to an FTP site for use by the DMV and DOT&PF
4. Hand entry on paper, which is typically from rural areas; some are scans of paper reports; some are just paper reports. For example: walrus accident on St. Lawrence Island.

Troopers, some local police, airport police, and university police use Toughbook to enter Form 200 data. Then TraCS, a Department of Public Safety system, gathers electronic data from it. Trooper staff review each report; if in order, the TraCS data are sent to the DMV system. From there, the parsable TraCS data are sent to the DOT&PF Oracle database.

Fairbanks police (FPD) use a different data platform for Form 200 data. Once approved, FPD data are also electronically transmitted to the DMV and then to the DOT&PF.

Anchorage police collect data electronically, then generate a pdf of Form 200, which is sent via an FTP electronically to the DMV and DOT&PF.

Alaska Department of Transportation and Public Facilities (ADOT&PF) crash data staff and contract personnel enter pdfs and paper documents into the Oracle database at DOT&PF. All records require crashes to be geo-located, though it is often difficult to determine the correct location from citizen reports. Percent % of crashes were reported on Form 200 (police), 20 percent on Form 209. But beginning in 2013, the FHWA/ National Highway Traffic Safety Administration (NHTSA) recommended that Form 200 be Model Minimum Uniform Crash Criteria (MMUCC) compliant. When the form was updated in 2012, those changes were incorporated, and now Form 200 is more complex and difficult for law enforcement to fill in. Accordingly, some officers will not use Form 200 for accident types they would have reported in the past. Therefore, a greater proportion of reports are now on Form 209. Preliminary numbers from 2013 show about 8000 Form 200s and 4500 Form 209s. This has led to an increased workload for the DOT&PF staff and its contracted data entry personnel.

Data Retrieval and Use

The Highway Analysis System (HAS) was the former database. Developed by Boeing, it had data back to 1977. DOT&PF has decommissioned HAS and has contracted with the University of Alabama to take advantage of its Critical Analysis Reporting Environment (CARE) software for data analysis. During the transition, data are available by request from DOT&PF HQ T&S. They extract data from Oracle to ACCESS and then build an EXCEL or MS Access spreadsheet for the requester. (See Appendix E.3.)

CARE, the interface between the Oracle database and the user, will replace the old methods of querying the HAS. The new system will improve, simplify, and streamline information retrieval. (See Appendix E.4.) CARE's development is expected to be complete by the end of 2016.

CARE's primary users will be regional traffic/safety engineers. Of note for students and researchers, ultimately there may be some public access. It is unlikely that design engineers will go directly to CARE; more likely they will place a request through regional traffic/safety staff. These procedures are still in development.

An important matter is managing expectations about the new system. HQ T&S is working on policies and procedures, but these are not available for HSIP report preparation, nor for public access yet. HQ is working with regional traffic staff to decide what they need and how to access CARE. User groups are working directly with the University of Alabama on CARE's development.

Special features of the Alaska version of CARE: animal strikes, lighting, darkness.

Time to fill a request? For a minor one (e.g., number of crashes on a segment), expect the response to take one day, while more complicated requests take longer. HQ T&S has to extract the data from Oracle via MS Access and EXCEL, as noted above. CARE will improve this process. Currently, design engineer or traffic engineer requests arrive by e-mail. With CARE, the traffic or design engineer will use the interface directly. Training will be required.

The HSIP annual report is prepared by regional and HQ traffic engineers. The 2014 report was based on HAS, while the 2015 report will be from the Oracle database via CARE.

The overall HSIP approval process (see Appendix E.5) is as follows:

1. The Headquarters Traffic and Safety office (HQ T&S) produces parameters and procedures regarding the HSIP ranking process and generates the crash data. These crash data had been available to the districts ("regions" in Alaska) in HAS and now (or soon) will be available in CARE.

2. The region now selects projects based on parameters from HQ T&S, scopes the project, and performs a B/C analysis based on crash reduction factors. The region will consult with local agencies regarding projects, but the DOT does not transfer HSIP funding to local agencies.
3. The region submits to HQ T&S its report, which includes its B/C analysis and discussion of conformance to the SHSP.
4. HQ T&S sends requests from the regions to the FHWA for approval and notifies the regions of approved projects.
5. The region submits funding requests for approved projects, previously approved projects, and phased projects to HQ T&S. (Approved non-HSIP projects may also be submitted for other funding sources.) HQ T&S selects HSIP projects and sends them to the Project Development Department (formerly "Planning") for inclusion in the HSIP funding plan.
6. Regions then develop the project and seek Project Development Authorization (PDA) for each HSIP project. This includes placing the project in the STIP, etc. Then the region designs and constructs the project.
7. HQ T&S reviews changes in the HSIP projects.

Safety Data are used in steps 1, 2, and 3 by HQ T&S and by the regions in planning, traffic, and design. One intention behind the transition to the CARE software is that all can access the data through a dashboard. In Appendix E.5 are excerpts of the Alaska HSIP Manual, listing the steps of data use.

2.3.2 Washington

Washington has two distinct tracks for HSIP projects: "state roads" versus local agencies. (Note the contrast: Alaska has no local HSIP projects, while in Washington 70 percent of the state HSIP funding is for projects managed by local agencies.) For either track, the raw data are available via the Statewide Electronic Collision and Ticketing Online Records (SECTOR) system, which is managed by the Data Office.

<u>Data Office (Transportation Data & GIS Office)</u>

See Appendix F.1.a for a brochure about the Transportation Data & GIS Office.

Washington uses SECTOR, which, among other things, is a system that can scan barcodes on drivers' licenses and registrations. About 85 percent of its reports are generated electronically. The other 15 percent, the paper reports, are entered into the database by the State Patrol, not DOT. SECTOR is the capture system that is provided free by WSDOT to patrol officers (and others). They do not have to use it, but whatever system they use must abide by the same business rules.

Reports come to the Data Office in real time, and it validates the data. If the report is accepted, the office assigns a seven-digit code and sends the report back to the officer. It is then transmitted as a "collision message" to the State Patrol back office, local records management system, and the Division of Licensing. Only the officer's report, not the refined report, goes back to the database. See Appendix F.1.b for data flow.

The Crash Location and Analysis System (CLAS) is used to refine the data for use in engineering, and only for state roads. There is a CLAS working group. SafetyAnalyst takes data directly from CLAS.

Future development plans include connecting CLAS to a work zone database so it can add to the CLAS data. Also, a new link/node system will be developed to assign unique identifiers and give ancillary location information (e.g., a vehicle that leaves a country roadway and crashes on a state roadway).

In terms of discoverable data, officer data are not protected, although refined data (for engineering) are protected and cannot be used in discovery. This split began in 2014. Refined data are in the DOT database. The Crash Data Mart and Crash Data Portal allow engineer and public access. One can call the Data Office and request data or access them directly. A new crash data portal is being developed. The Data Office sends feeds to specific agencies (e.g., a county).

<u>Washington State DOT Projects</u>

The general screening procedures are developed by the Capital Program Development and Management Office (CPDM) and the Highway Safety Issues Group (HSIG), which includes regions. The criteria are approved by the Highway Safety Executive Committee (HSEC). See Appendix F.2.a for a flow chart with acronyms.

1. The CPDM (planning) regional offices then makes lists of candidate projects using SafetyAnalyst, which replaces Collision Analysis Location and Collision Analysis Corridor (CAL/CAC), as well as Intersection Analysis Locator (IAL), methodologies. The lists are submitted to the HSEC, and if they are approved, the regions are notified.
2. The region then analyzes the countermeasures, does a B/C analysis using SafetyAnalyst, and documents the process for each project.
3. Documents are reviewed by CDPM for eligibly for HSIP funding.
4. Approved projects go to a review panel of HQ and regional staff, and if they are approved, a review by the HSEC and HSIG.
5. The region develops and programs the project, then the project is placed on the Programed Project List.
6. Legislative approval, design, and construction.

Safety Data (from SECTOR) are used in steps 2 and 3. Currently the SafetyAnalyst program is utilized (see below).

Washington Local Programs

See https://www.wsdot.wa.gov/LocalPrograms/Traffic/FedSafety.htm for an overview of city and county programs and Appendix F.3.a for a glance at the Washington DOT guidance for local agencies.

1. Local Programs (an office within DOT HQ) gets data from CLAS and sends a request for proposals to local agencies. CLAS data are tied to an agency, not intersections. Agencies can access CLAS data, but Local Programs does a summary and sends it to the agencies.
2. Local agencies then identify projects. The differences between county and city projects include the following:
    a. County projects go to the County Safety Program

    Counties usually have lower-cost projects, and these are reviewed for risk analysis. Counties are given crash data for the county and are encouraged to develop a local road safety plan, so countermeasures proposed should tie into risk factors.

b. City projects go to the City Safety Program

   Cities know that CMF will be used in a competitive benefit/cost analysis, usually favoring a high-cost program.

3. The local agency decides on location and countermeasures.
   a. Counties often identify areas with high crash potential, then Local Programs does the safety analysis and compares it to other alternatives.
   b. Cities do a more competitive analysis using B/C CMF, etc.
   c. Both may be aided by consultants who are familiar with B/C analysis.
4. Crash data from CLAS are used by a variety of agencies. For higher cost projects, generally in cities, the cities or their consultants evaluate crash data to identify an effective project with CRF and CMF for input into a B/C analysis. For lower cost projects, generally in counties, the data must indicate the risk and how it will be reduced.
5. Local Programs then reviews. There is no state money in the process; local agencies must provide the match, so the project does not need legislative approval. It must conform to the state's safety program for either cities or counties, and it must be entered in the STIP. Local Programs reviews all this.

2.3.3 Oregon

The Oregon SQL Database has all roads in the state, but the DOT only analyzes state roads, arterials, and collectors, which have 92 percent of the accidents. Responders submit crash reports on about 40 percent of crashes. Citizens must report to the DMV any property damage only (PDO) crash exceeding $1500. Only one-third to one-half of PDO reports come from citizens. The DMV does not have a fillable form; the user must print the form, fill it in, and send it to the DMV. Fatal and "A" (serious) injury crashes make up 5 to 7 percent of total crashes.

The Crash Analysis Reporting (CAR) Unit within the DOT gets reports from the DMV and inputs them (by hand) into FARS and the DOT database, performing QC for proper location, etc., in the process.

The DMV has the official copy of the crash report and inputs selected portions of the data, some of which are protected. The DOT does not keep a copy of the DMV report. The DOT's database is SQL, an IBM system. There is no distinction among roadways, arterial, collector, or local—all data are in the database and all public roadways are covered in one system.

Oregon uses the Safety Priority Index System (SPIS). See Appendix G.1 for an overview of Oregon's Project Safety Management System with an explanation of SPIS. After the QC is performed on the database, SPIS calculates a score.

1. Each year, HQ uses SPIS to calculate a score for any roadway section that has one or more fatalities and/or three or more of any kind of non-fatal crash during that year.
2. Based on crash data, funding is allocated to the regions for decisions on projects.
3. The regions investigate segments with the top 5 percent scores and considers public input and district O&M input. They develop suggestions for use of the funds. There is no B/C study at this point; rather, more informal suggestions are the result.
4. From these suggestions, they develop a "300 percent list." Included are projects whose estimated costs sum to 300 percent of the funding allocated to the region for safety. A preliminary B/C analysis is performed on all the projects on this 300 percent list.
5. The list is then reduced to a "150 percent list." At this point, they do project scoping, taking a multidisciplinary approach (including environmental, ROW, etc.) Then the B/C analyses are refined.
6. The final proposed list is made by cutting off at 100 percent. Final approval is given by the Transportation Commission.
7. The DOT provides the report on the use of HSIP dollars and must collect data from itself and from local agencies if they have done the work.

2.3.4 Idaho

Idaho's system has been paperless since 2007. An accident of less than $1500 damage and no injury is non-reportable. A uniform crash report is required by law. Officers report to the Idaho Mobile

Program for Accident CollecTion (IMPACT). Within the DOT, the IMPACT data are downloaded daily to CIRCA at the district level. There are six districts, and each gets about the same number of downloads. At this point, QC is performed on location, weather, cause, etc. If questions arise, the officer is asked to clarify.

The FARS report is also handled here. The Idaho State Police, sheriffs, and local police all feed data to IMPACT. (Appendix H.1 has a press release regarding IMPACT 2K and WebCARS Software.)

Data are input via the IMPACT and stored in the Crash Information Retrieval Collection and Analysis (CIRCA) system, which can collect data from high impact locations.

1. Engineers at the district level use the data from CIRCA to determine where improvements are needed.
2. Engineers then use WebCARS to query CIRCA.
3. Engineers at the district level then do an initial B/C analysis using the Transportation Economic Development Impact System (TREDIS).
4. The district results are combined statewide, and an overall B/C priority listing is developed.
5. A committee at headquarters uses the B/C analysis to decide funding, then funding is apportioned to districts.

2.3.5 Colorado

Each law enforcement agency has its own system for data collection, but they all use a common form. Colorado State Patrol and major cities use an electronic form. The other agencies (over half the state) use paper, which is then scanned and entered either by hand or by optical readers or other electronics.

Initial reports go to the Department of Revenue. From the Department of Revenue database, the DOT imports all data into its own database. At that point, the DOT does QC and also removes some unneeded data (personal ID, VIN, etc).

The Colorado DOT uses a program called Vision Zero Suite, developed by DiExSys, to do both query and analysis. The analysis checks crash types and identifies locations with a greater than expected number of crashes. The database includes all public roads in the state.

HSIP projects are approved on a case-by-case basis. HQ T&S first provides the five regions with scans of areas with higher than expected crashes and also indicates whether a correctable fix is available. The regions submit applications, and HQ T&S staff use Vision Zero Suite to do B/C analyses, including the use of CMFs.

Projects are then ranked without committee involvement. Approvals are made by HQ T&S, working closely with the FHWA.

Funds are allocated in advance to regions, based on the historic numbers of crashes. Then there is competition within the region based on B/C analyses and resulting rankings. State and local HSIP funds are available at about a 50/50 rate. Regional requests include both state and local requests.

The HSIP annual report is prepared by the HQ T&S office. Some information comes from Vision Zero Suite, although financial aspects come from finance programs (separate from Vision Zero Suite).

Before and after studies are a new feature. Among other things, these studies can be used to modify CMFs.

2.3.6 Arizona

The Arizona DOT is transitioning from traditional historic crash-data-based hotspot (e.g., segments, intersections) identification to predictive analysis using safety performance functions. The agency started experimentation with the Highway Safety Manual (HSM) Part C and the Interactive Highway Safety Design Model (IHSDM), and has successfully completed a few pilot projects on the state highway system using the IHSDM. It is also working on implementing AASHTOware SafetyAnalyst for network screening using available data from the state highway system.

Crash data come from all levels of public safety agencies (i.e., state, county, city, Gila River Indian Community and a few other tribal law enforcement agencies) in electronic and hard-copy forms. The Arizona Department of Public Safety (DPS) collects and submits 100 percent of the crash data in electronic form using Traffic and Criminal Software (TraCS). Local agencies use TraCS, other electronic tools, or hard-copy reports. All crash data received by the Arizona DOT are entered into the Accident Location Identification and Surveillance System (ALISS) central database. Approximately 59 percent of crash data received by the Traffic Records Section in 2015 were in electronic form and the rest came in hard copies, which were manually entered into the ALISS. All hard-copy crash reports are currently being scanned and digitally stored as records and for necessary use. There are microfilmed crash reports prior to July 2012 going back to 2006 to comply with the record retention policy. There are about 110,000 reported crashes per year in Arizona. Arizona Motor Vehicle Crash Facts are posted online (http://www.azdot.gov/mvd/statistics/arizona-motor-vehicle-crash-facts).

The analysts and other users can query crash data using the Safety Data Mart (SDM) online tool.

However, other safety data (e.g., roadway data, traffic data) are housed in different databases. Currently, GIS is being used to integrate, analyze, and distribute various databases according to needs. The Arizona DOT is looking into developing a comprehensive data system covering, at a minimum, the Model Inventory of Roadway Elements (MIRE) – Fundamental Data Elements (FDE) that makes the most efficient use of existing GIS and enterprise systems as possible.

For the last few years, at least 20 percent of the annual HSIP funding has been allocated for local public agencies. See Appendix J.1 for a sample flow chart on the local public agency funding approval process developed for FY2017. Arizona is transitioning to a more competitive program, and HSIP funding will be 100 percent open to any public agency beginning FY2019.

The Arizona DOT is in the process of completing the following research projects related to HSM implementation:

- SPR-651: Incorporating Safety Performance into Project Design
- SPR-693: Feasibility Study for Arizona's Roadway Management Process using the Highway Safety Manual and SafetyAnalyst
- SPR-704: ADOT State-Specific Crash Prediction Models: an Arizona Needs Study
- SPR-721: Data Needs for Tree Removal CMFs on Arizona State Highways.

The final reports will be published and available on the Research Center website (http://azdot.gov/planning/research-center/library/research-reports).

2.4 Training

2.4.1 Alaska

The Alaska State Crash Data Manager has written a manual for data entry clerical tasks. As CARE is implemented, training will be needed for

- HSIP report preparation
- Traffic/safety engineers (regional and state headquarters)
- Public users
- Police officers
- Maybe the Local Technical Assistance Program (LTAP), depending on locality.

2.4.2 Washington

The Crash Data & Reporting Branch already does some training of engineers in data handling and use. There may be a need to train local jurisdiction personnel, although many use consultants who already know the system. Data can be very complex, and therefore, training may be less important for engineers (and other similar types) because the Crash Data and Reporting office has people who can help with requests, alleviating the need for engineers to query the system directly.

2.4.3 Oregon

Headquarters trains regional staff on the use of tools. Regions do the training of local agency personnel; a small city may use the state DOT or consultants for training

2.4.4 Idaho

The Highway Safety Manager trains police officers in the use of IMPACT. There are three levels of training: Initial (4 hours), Supervisors, and Refresher. Headquarters provides training on WebCARS.

2.4.5 Colorado

Regional traffic engineers must be trained by DiExSys before using the Vision Zero Suite. A DiExSys license requires that they conduct two training sessions per year. Locals and consultants also attend.

2.4.6 Arizona

Most of the training classes are hands-on and are organized as needed. Both in-house experts and outside instructors provide training.

2.5 Conclusions

While the basic notion of the acquisition, flow, storage, and use of data is similar in all the states queried, many of the details are quite different among states and, in some cases, different among the data used for state highway projects and data used for local projects within that same state. Differences are reviewed in some detail above. Here we try to extract some issues that may be important to the education and training of engineers and other transportation professionals.

For traffic, safety, and design engineers, the important knowledge about the data acquisition encompasses its limitations. For example, might some rural roads or jurisdictions be under- or over-reported?

The most common reported difficulty with the data seems to be crash locations, but all states recognize this and work to validate and modify, if appropriate, location data through some sort of QC process. Not only milepost or GPS location, but also ownership of the road—state route vs. local road—is often important. Road construction often complicates these jurisdictional and location data.

Those using the data should also be aware that several agencies may have been involved in the data's gathering and compilation. Inter-agency cooperation and coordination are an important part of assuring accuracy and usability of the data. Despite the automated nature of data transfer from one agency to another, there may be errors in the data or their interpretation. The QC process necessarily involves a degree of judgment and good communications skills.

A key use of crash data would be the adjustment of the SPF and CMF for the state or local situation. Typically, engineers use SPFs and CMFs from the HSM or from an online service of the FHWA, the CMF Clearing House (see Appendix D.4), http://www.cmfclearinghouse.org/ . The CMFs are used to evaluate the benefits of various countermeasures, based on roadway/intersection type, average annual daily traffic (AADT), and other parameters. However, there may be local factors that would suggest adjustment of the standard CMF for use in the B/C analysis. These might be evaluated by using before and after data for modifications of the SPFs and CMFs for local roads.

An important management issue can be the selection of crash data. The traditional method for project selection has been to evaluate the number of crashes at particular segments/locations and compare the findings with SPF, then base the estimation of benefit on these numbers. However, the more modern method is to use an EB analysis that takes into account the crash data from other similar locations. Several of the computer programs the DOTs use do this automatically. This provides a more statistically sound method of estimation. However, it also injects a layer of complexity. Therefore, knowledge of the basis of the computer programs that estimate benefit is required to inform sound engineering judgment.

The SPFs and CMFs should be validated for particular states, regions, or locations on the basis of a before and after analysis. This requires several years of crash data from before the countermeasure was installed and several years of data afterward. The Colorado example in the appendix uses six years of crash data from before the countermeasure and six years afterward—for a total of 12 years. Certainly AADTs will change over 12 years, although these can be accounted for directly, as well as costs. However, other factors may change as well in 12 years, making good comparisons difficult.

2.6 Acknowledgments

We thank the following for useful interviews and data:

Alaska:
Clint Farr, ADOT&PF, Statewide Program Development, Juneau
Pam Golden, PE, ADOT&PF, Regional Traffic Engineer, Fairbanks

Washington:
Nadine Jobe, WSDOT, Crash Data and Reporting Manager, Tumwater
Warren Stanley, WSDOT, Business Project Manager, Tumwater

John Milton, Ph.D., PE, WSDOT, Director, Quality Assurance and Transportation System Safety
Ida van Schalkwyk, Ph.D., WSDOT, Traffic Safety Research Engineer
Matthew Enders, PE, WSDOT Local Programs

Arizona:
Kohinoor Kar, Ph.D., PE, PTOE, Senior Project Manager, Research Center

Idaho:
John Tomlinson, Highway Safety Manager, Idaho DOT, Boise
Debbie Dorius, FARS Analyst, Office of Highway Safety, Idaho DOT, Boise

Oregon:
Kevin J. Haas, PE, Traffic Standards Engineer, Oregon Department of Transportation
Douglas W. Bish, PE, Traffic Engineering Unit Manager, Oregon DOT, Salem

Colorado:
David Swenka, Safety Analysis Engineer, Engineering Branch, Traffic & Safety, Colorado DOT

3. Gathering and Using the Data (Academic Perspective)

The subject of safety is among the most fundamental principles in transportation engineering, but there is minimal understanding as to how safety data are used and defined in both academic and professional settings. This study administered a set of open-ended interviews and surveys to transportation educators, researchers, and practitioners with the objective of better understanding how safety data are currently being used in course curricula, research, and professional practice. Our findings concluded that, while various forms of safety data are collected and analyzed in professional practice and research, the application of safety data within course curricula is confined mainly to highlighting facts such as vehicle collisions. The application of safety data within practice is oriented toward design projects, while in research, safety data have an important role toward the exploration and development of new methods and solutions for transportation-related issues. The use of safety data has always been imperative in engineering. Knowing how transportation practitioners are currently using these data is crucial to determining ways in which safety data can be ideally incorporated in academics to best prepare students in the classroom for the workplace and to provide guidance to faculty who teach these courses.

3.1 Introduction

Safety is one of the most important concepts in civil engineering. In the first canon of the American Society of Civil Engineers' code of ethics, it states that "engineers shall hold paramount the safety, health, and welfare of the public and shall strive to comply with the principles of sustainable development in the performance of their professional duties" (Merit 2016). The importance of safety goes beyond the cost and efficiency of a project and should be treated as a priority in engineering practice, design, research, and education. Despite its increasing application and importance within the field of transportation engineering (Vlahogianni et al. 2015), the specific ways in which data relating to safety are used in three variably different settings (i.e., education, research, and practice) have not been explored or compared. This study examines how engineers within education, research, and practice describe their use and definition of safety data. Understanding these approaches can inform educational efforts to improve student preparation for practice as future transportation professionals.

The topic of safety ranges in complexity in practice and design. For example, a safety solution in transportation can be a simple premise, such as encouraging the public to wear a helmet while riding a bicycle, to a more complex situation such as redesigning curvature on a roadway to increase sight distance. Such a wide range of complexity causes varied definitions and uses of data relating to safety. In order to better understand the relevance of safety data in transportation and their ultimate application or applications, it is imperative to understand how people using the data define this information and apply it within their work.

One area of limited exploration is how safety data, and particularly "big data" related to safety in transportation engineering, are actually defined and used in education, research, and practice (Vlahogianni et al. 2015). Vehicle crash rates are a good example of one type of safety data. Practitioners will review historical crash rates when redesigning a roadway. Their review may trigger the need for additional safety treatments such as a guardrail or concrete jersey barrier. Alternately, educators and researchers may treat these data differently, focusing their efforts on human factor elements or general safety trends. As a result, does safety data play a different role in education, research, and practice? Is safety data perceived differently within these three contexts? Understanding the state of practice is important because it allows for the understanding of current trends within the transportation community and reveals areas of opportunity and limitations that need to be addressed.

On the basis of this line of inquiry, this research examined how safety is being defined and used in university courses, transportation engineering research entities, and practical settings across the United States. This research also sought to understand what transportation engineers mean by "big data" in the context of safety, and how "big data" are used in academic and professional settings. This information was obtained from interviews and survey responses with educators, researchers, and practicing engineers and assessed on the basis of the respondents' professional background, experience with safety data, and use of certain types of safety data.

3.2 Literature Review

Safety data and “big data” are a critical and changing aspect of transportation, but there exists a limited understanding as to the role and representation of “big data” in academia and in design. The usage and meaning of “big data” to educators, researchers and practitioners are important because a specific definition serves as a means of communication within and across these settings (Montfort, 2013). It also helps educators align the ways in which safety is implemented in courses and how “big data” are used by transportation professionals.

The term “big data” has been used in a variety of ways across different fields and is commonly used in business to describe large amounts of complex data that can reach exabytes in size, and require particular means of storage, analysis, and management (Chen et al. 2012). Most literature defining “big data” associate the term with variety, volume, variability, complexity, and value (Katal et al. 2013; McAfee et al. 2012). The “big data” concept is associated with business analytics, along with logistics and supply management (Wang et al. 2016; Vera-Baquero et al. 2015). In health care, “big data” are seen as a means to improve outcome predictions and treatment effectiveness, to disseminate knowledge regarding treatments, and to encourage patients through the delivery of information (Schneeweiss 2014; Murdoch and Detsky 2013). “Big data” are used in a variety of settings, from science (e.g., the Large Hadron Collider) to government applications for security programs to the private sector, such as Amazon or WalMart, and to international development and social media platforms (Katal et al. 2013). Within the transportation arena, there exists increasing expectations of pairing “big data” with the development of intelligent transportation systems (ITS) as a means of increasing infrastructure safety (Shi and Abdel-Aty 2015; Vlahogianni et al. 2015). “Big data” in transportation have been used to estimate travel demand and flow (Toole et al. 2015; Lv et al. 2015), traveler movements and patterns (Kieu et al. 2015; Huang and Levinson 2015; Carrel et al. 2015), and safety (Shi and Abdel-Aty, 2015; St-Aubin et al. 2015). “Big data” within the noted research come from a wide range of sources, including GPS, video, and mobile data. These varied applications exemplify the increasingly diverse potential of these data to revolutionize many fields.

3.2.1 How Is Safety Data Incorporated into Transportation Courses?

Safety has been recognized as an important aspect of transportation engineering. However, some education and research communities believe that safety education has been limited in undergraduate and graduate programs in civil engineering (Gross and Jovanis 2008).

The current state of transportation safety education, and specifically highway safety education, through course offerings in engineering and public health was investigated. Engineering and health programs were contacted and provided with a survey asking department heads to answer questions regarding safety courses that were offered at their respective institutions and to provide faculty-teaching tools such as class syllabi, course outlines, and web page information. The findings supported the hypothesis that highway safety is underrepresented in transportation curricula throughout the United States in comparison to design, operations, and planning (Gross and Jovanis 2008) and provided a baseline contention that highway safety is not adequately addressed in engineering schools within the United States. A 2012 study researched the most important topics for the first course in transportation engineering from two different perspectives, namely educators and practitioners (Turochy 2012). From the collected survey responses, traffic safety was the only safety topic that was included out of 34 topics, and it ranked 8th.

The use of safety data is a critical component of transportation safety practice, and therefore requires an equivalent degree of preparation in students who wish to pursue a career in the field. However, there has not been any targeted research investigating the use of safety data and “big data” in transportation programs. This could potentially mean that there is a lack of safety data topics being incorporated in civil engineering curricula.

3.2.2 How Are Transportation Researchers Using Safety Data?

In transportation research, one contemporary and growing field is connected vehicles. Huge amounts of data are required to develop and implement connected vehicles (Lu et al. 2014). Two main objectives of connected vehicles are the urgent need to improve efficiency and safety on the roadway and to increase the demand for mobile data of users on the road (Lu et al. 2014).

Similarly, there is substantial research on autonomous vehicles and the large amounts of data required for vehicle functionality. One of the main issues that has been extensively researched focuses on determining the trajectories of vehicles that must avoid colliding with fixed and moving obstacles (Frazzoli et al. 2002). Multiple solutions to this problem have emerged, but the potential answer still cannot guarantee that the solution will be collision-free (Dolgov et al. 2010).

Researchers use data from collision sets, such as the domestic National Highway Traffic Administration or the international Australian National Crash In-depth Study (ANCIS), with the intent of finding out how collisions occur (Beanland et al. 2013). Causes, such as driver inattention, can be drawn from collision data, and this type of conclusion can then guide engineers to determine the appropriate engineering, enforcement, or education-related outcomes.

There is often a disconnect between the learning and research activities in an academic environment and the skill set and application techniques required in the workplace. Higher education programs do not always address aspects of practice that are frequently seen as critical by practitioners (Felder and Brent 2003). Additionally, disconnects between research and practice are inhibitors of technology transfer efforts (Hood et al. 2014). Understanding the linkages between and applications from these diverse contexts can begin to inform communication between them and improve outcomes for each.

3.2.3 How Are Transportation Professionals Using Safety Data?

Professional transportation engineers depend on data to perform certain tasks. Safety data and “big data” in transportation come from many sources, including traffic surveillance systems such as roadway detectors (Shi and Abdel-Aty 2015). Practitioners use safety data collected from roadway detectors to design infrastructure and mitigate for congestion and vehicle collisions. Highway engineers rely on safety manuals, such as the American Association of State Highway and Transportation Officials’ (AASHTO) “A Policy on Geometric Design of Highways and Streets,” to develop roadway geometries that comply with required specifications. Databases populated with crash data are used to predict the number of expected crashes for implemented countermeasures. The Crash Modification Factors Clearinghouse provides data for practitioners and researchers to quantify safety applications on roadways.

Traffic control devices enhance safety on streets by providing guidance and direction. Roadway engineers rely on the Manual on Uniform Traffic Control Devices (MUTCD) resource, which are the standards developed by the Federal Highway Administration (FHWA), to foster safe and efficient travel for all users.

These examples represent just a sampling of the many references and resources that professional transportation engineers utilize in practice, but no explicit literature explains the ways and types of safety data they use. This research study sought to fill this knowledge gap by recognizing the types of data and their uses in practice settings.

3.3 Materials and Methods

3.3.1 Data Collection

In order to determine how safety data are used in education, research, and practice and to compare the similarities and differences in these uses, interviews were conducted in an open-ended fashion using a pre-defined protocol. Researchers were afforded the flexibility of asking follow-up questions, which gave participants the opportunity to provide in-depth responses. A separate, open-ended survey was also developed to capture core understandings from a broader population.

Interviews and surveys were conducted to gather information about the definition of “big data” and to determine how educators, researchers, and practitioners use safety data in transportation engineering. In general, interviews are best used to gain detailed accounts of a topic or phenomena, while surveys are used to learn general patterns from a large sample of people and are particularly useful when determining expert opinions on issues (Driscoll 2011). Additionally, interviews allow the interviewer to gather a depth of information from an interviewee, including self-disclosure (Rothe 1991). Open-ended questions in qualitative research serve to gain detailed and comprehensive information on specific topics (Rapley 2001). The benefits of an open-ended question allow for a schedule of topics in which the order does not matter and the questions are answered in a follow-up question mode to collect the most information possible from participants (Potter and Hepburn 2005). Gathering detailed information was necessary to identify the potentially varied differences in use within the three discipline contexts. For such a complex research topic heavily related to perspectives from many experts, the combination of both methods was employed to capture both the in-depth details from the personal interviews and the significant data from the surveys for comparison.

3.3.2 Interview Development and Implementation

Interviews were conducted with the objective of gathering information to answer the research question regarding the definition of “big data” from education, research, and practice perspectives and to identify similarities or differences in the use of “big data” in these settings. Similarly, this method allowed us to gather information about safety data and how individuals apply them in different settings. By conducting interviews, the interviewees were encouraged to disclose their understandings of a phenomenon. Since the targeted participants were transportation faculty and professionals with a safety background and expertise in safety data or “big data,” interviewing these participants revealed detailed information related to the research goal.

An interview protocol was developed with 26 total questions. The protocol was divided into three sections: professional background, teaching background, and research background. The interviews lasted between 15 and 30 minutes, depending on the interviewee’s background. The interview questions are shown in figure 3-1.

3.3.3 Participant Selection

The participants chosen for this study were limited to instructors teaching transportation courses, faculty conducting research using safety data, and transportation professionals with expertise in safety data. In order to identify participants with transportation engineering expertise, a list of transportation engineering professors and their e-mail addresses from all civil engineering schools in the United States was developed; this list consisted of 404 individuals. By recruiting from online sources, it was possible to obtain a large sample size. When one can obtain samples that are heterogeneous with respect to a specific expertise, it is possible to reach people with special or rare characteristics (Birnbaum 2004).

Interview requests were e-mailed to professionals with diverse occupational responsibilities and experiences and who currently work in industry and use safety data as part of their job. These contacts were identified from the professional network of the authors. A total of 404 individuals were contacted on three different occasions, of which 22 participated in interviews. Table 3-1 differentiates the expertise of the interview participants; it should be noted that some participants shared duties that overlapped in the three categories (i.e., education, research, or practice), so this total number exceeded the actual number of interviewees.

Interview Questions

1. Tell me a little bit about yourself, how did you end up in your current job?
2. What does "big data" mean in terms of safety data?
3. What types of big data are there related to safety?

Section 1: Professional Background

4. Do you or have you ever worked in the transportation industry?
5. Do you or did you use safety data in your job?
6. What type of data do/did you use?
7. How do/did you apply it in your job?
8. Where do/did you obtain the safety data?
9. Is there a challenge with big data, if so what is the challenge?
10. How important is safety data?
11. Do you think incorporating safety data in transportation courses can improve students learning about safety?
12. In terms of software, what type of software do you use?

Section 2: Teaching Background

13. Do you or have you ever teach transportation courses?
14. Have you ever teach any transportation safety courses?
15. Which courses?
16. Do you use safety data in the transportation courses that you teach?
17. What type of safety data do you use?
18. Do you use any resources related to safety data?
19. What topics are covered for the courses where you use safety data?
20. What are the challenges with using safety data in a class?
21. How necessary do you think safety data is within course curricula?

Section 3: Research Background

22. Do you or have you ever done transportation research?
23. Do you use safety data in the transportation research that you do?
24. Where do you obtain the safety data?
25. How do you use this data?
26. How necessary do you think safety data is within transportation research?

Figure 3-1: Interview question protocol

Table 3-1: Interview participant expertise areas

Professional Setting	Practice	Education	Research
Number of Participants	16	14	12

3.3.4 Survey Instrument Development and Implementation

In addition to the participant interviews, an online survey was developed to collect and stratify information as to how “big data” and safety data are used in the three settings. Previous transportation research had utilized surveys to collect information related to undergraduate-level transportation courses in order to improve the quality of the instruction materials. That survey tool allowed a variety of questions to be asked so that the most meaningful and representative information could be obtained from respondents (Turochy et al. 2013). The nature of the survey questions was similar to that of the interview and consisted of three different expertise areas: teaching, research, and professional industry. The survey consisted of 21 questions related to the definition, application, challenges, and resources related to safety data and “big data.” Similar to the interview questions, the survey was utilized as a tool to ask questions to targeted participants. For this study, the number of survey participants was not equally proportional in terms of educators, researchers, and practitioners since the survey request was sent predominantly to participants in academia. Individuals with practical expertise working with the management and application of safety data were comparably fewer than those with educational expertise and who taught courses or conducted research using safety data. Although there were limitations with this sample because of the disproportionately small number of practitioners, their collective responses were very similar in nature, so making generalizations was possible, given this level of consistency.

3.3.5 Participant Selection

Transportation faculty, practicing engineers, and department of transportation (DOT) staff were contacted by e-mail and asked to complete the survey. The desired survey participants included instructors who taught transportation courses, individuals who conducted research utilizing safety data, and transportation engineers with expertise in safety data. The survey request was sent to the same transportation engineering professors and transportation professionals across the United States who were contacted for the interview. In order to target desired participants whose jobs consisted of analyzing and managing "big data," the first survey question inquired about the participant's use of safety data in his or her job. If the participant's daily employment duties were not related to safety data, the survey ended for this participant. Table 3-2 shows the number of survey participants for each of the three settings. Note that the total number exceeded the actual number of participants (N=44) as some individuals were associated with more than one category.

3.3.6 Data Analysis

On the basis of the interview and survey responses, a qualitative analysis was performed to assess the different views and perspectives of safety data and "big data" by researchers, practitioners, and educators. A general inductive coding approach (Thomas 2006) was utilized to analyze the data from both the interviews and survey data by closely reading the interview and survey transcriptions and looking for meanings inherent to the research question. Inductive coding was suitable for the interviews and surveys performed in this research study because the questions asked were primarily open-ended. This type of question required the data to be read multiple times to derive codes (Hsieh and Shannon 2005; Thomas 2006). The extensive open-ended answers were processed in order to quantify the approximate percentages of the number of participants whose answers fell under the developed code.

Table 3-2: Survey participant expertise areas

Professional setting	Practice	Education	Research
Number of Participants	6	38	25

3.4 Results

The analysis of the data indicated that (1) the concept of "big data" relating to safety appears to have shared meaning across academic, research, and practicing cohorts, (2) "big data" relating to safety were used in a targeted manner, and (3) safety is rarely explicitly manifested in existing transportation curriculum in terms of design. The following results were derived from both the survey and interview results, with the percentages shown reflecting all of the participants.

3.4.1 How Is the Term "Big Data" Defined?

When asked to define the term "big data," approximately 90 percent of the participants (N=61) associated the term with large amounts of data stored for multiple uses in the field of engineering design or being complex in nature. Some individuals provided quantitative dimensions to describe the size of "big data."

"The biggest data set we work with is our collision data set which … has about 150,000 records" (Participant 17). Large numbers were used as a description to cite examples. "There are dozens of attributes for each record, and generally in our database there are millions of individual attributes" (Participant 17). Approximately 6 percent of the participants (N=4) were not familiar with the term. "Don't have such a definition" (Participant 45). "I have heard of the term, but am not familiar with it" (Participant 34). Definitions varied in context and meaning based on a participant's perspective. For example, practitioners who worked in industry shared a different meaning of "big data" than educators and researchers who teach at universities. From a practice perspective, the term "big data" was defined by using technical examples and applications, and these participants generally linked "big data" with applications in engineering design. On the other hand, participants with exclusively teaching and research expertise defined "big data" theoretically as large data points collected and stored in databases for multiple engineering applications. "When it comes to traffic safety I think 'big data' in this country is defined as data that is of a national scope, looking at large data sets that include not just crash data but a more complete picture of what's going on" (Participant 11). "Big data is data that is so large in quantity that you have difficulty analyzing it by traditional means" (Participant 10). Researchers did not link "big data" with technical applications as much as professionals. In figure 3-2, the meaning of "big data" from the survey and interview results across academia and practice are summarized. Every educator who also performed research provided a single definition for both professional settings; for this reason, educators and researchers were grouped under the academia category for this analysis.

Setting	**Practice**	**Academia**
Generalized definition	Term was defined as a large collection of crash data and various job applications.	Term was defined as a large dataset of information pertinent to autonomous vehicles and real-time information.
Variety (sources, formats, etc.)	Mostly explained "big data" in terms of crash data Some mentioned "real time" data to communicate to travelers Commonly mentioned in terms of databases	Most mentioned crash data Some autonomous vehicles, connecting vehicles Some mentioned driver behavior data. Some mentioned traffic data.
Volume (quantitative descriptions, means of storage, storage required, etc.)	Descriptions range from hundredths of points to millions of data points Mostly described as "large datasets"	Descriptions range from thousands to millions to data points Also described qualitatively in terms of "large sizes"
Velocity (frequency of data collection, frequency of data delivery)	In relation to crash data, most considered historical information Some mention of realtime/passive collection of data	Automatic/passive collection of data Real time data One related data collection in terms of observations per second
Complexity (complications associated with "big data")	Some mention how big data can be really complex where a normal citizen user would not have the ability to do the analysis. Some mention that big data is complex in volume and speed.	Due to large size of datasets, participants note that the data requires extensive effort for processing and analysis and storing it.
Value (benefits and costs associated with "big data"	Most mentioned how big data was really important in their line of work.	Most mentioned how big data was really important within research. Not many mentioned how big data was essential for teaching.

Veracity (accuracy of data collected)	Some mentioned that accuracy was always a challenge in collecting this type of data.	Some were concerned with not obtaining the amount of data needed for research.
Data types used to explain "big data"	GPS Collision data Inventory data Naturalistic driving studies	GPS Video and sensor data Real time data Outputs from video Collision data

Figure 3-2: Defining "big data" in practice and academia

There were some similarities and differences between the way the term "big data" was defined and described from the practice and academic perspectives. In general terms, the first word associated with "big data" was the term "large." Practitioners specifically expressed it as large amounts of collision data, while academicians used large amounts of real-time data to define "big data." In terms of variety, both perspectives quantified "big data" in many forms and applications. In terms of volume, the descriptions were similar, as "big data" were explained in large numbers ranging from hundreds to millions of data points. In terms of velocity, academicians described "big data" as occurring in many observations per second. While practitioners did not provide a specific velocity for the collection of "big data," they explained that real-time data are a fast way to collect "big data." In terms of complexity, practitioners mostly described how only trained individuals could perform analysis because of its complexity.

Academicians acknowledged the challenges of storing "big data" because of their size and complexity. In terms of value, academicians and practitioners explained that the use of "big data" was crucial in their line of work but less so from a teaching standpoint. Lastly, in terms of veracity, practitioners explained how obtaining a desired level of accuracy was always challenging. Academicians reflected on the challenge of obtaining a sufficient amount of data for research purposes.

3.4.2 How Are Safety Data Incorporated into Transportation Courses?

Approximately 92 percent of participants (N=48) who taught transportation courses reported incorporating safety data, albeit in different forms. The most common type used was collision data. "I tried getting students introduced to the general estimate system (GES)…. as well as [to] huge states that provide relatively detailed crash data basis online" (Participant 10). Many participants stated that they incorporated safety data in their classes by utilizing data sets for homework assignments on class projects. A few participants indicated that they briefly discussed safety data in their courses. "On a limited basis, you know, I would say just to highlight the fact that in the United States there are certain traffic crashes" (Participant 1). "We do a little bit of that, but I would say in most cases I would do the queries and provide that data directly to the students" (Participant 6). "We mainly use the reported crash data; (it) is about all we have time to really get into" (Participant 8). Other participants did not provide detail on the type of safety data they incorporated but highlighted it in general terms such as "data analysis and calculations" (Participant 4).

Despite the fact that 92 percent of participants (N=48) indicated they incorporated safety in the transportation courses that they taught, approximately 35 percent (N=17) rarely talked about safety data-related topics in their courses. Of the remaining 65 percent of participants (N=31) who incorporated safety into transportation courses that they taught, approximately 66 percent (N=21) included collision data as the most common type of safety data. "We mainly use crash data; it is about all we have time to really get in to" (Participant 8). "Really, just crash data, small crash data so they can calculate some rates, or do some comparisons before and after" (Participant 7).

Approximately 6 percent (N=2) mentioned that roadway type represented the safety data that they incorporated in their courses. Another 6 percent of participants (N=2) identified safety standards, while approximately 9 percent of interview participants (N=3) answered that driver characteristics were the type of safety data incorporated into their classes. "We cover the characteristics for drivers and safety factors a little bit" (Participant 4). The remaining 13 percent of respondents (N=4) answered that traffic volume data were the main type of safety data in their courses.

Academicians highlighted potential challenges with using safety data in their work, as well as incorporating it into their courses. The challenges faced by participants in the academic environment involved privacy issues, as well as a student's ability to work with data sets. One of the participants pointed out that, in order to utilize real-world safety data for a class, a process to obtain permission must be completed. Another challenge was that some students simply were not comfortable using data sets and the large quantity of numbers and were simply unprepared to work with them.

3.4.3 How Are Transportation Researchers Using Safety Data?

All of the participants (N=37) used safety data for research purposes, with approximately 70 percent (N=26) utilizing crash data for their research. "Well, it's usually preparing data sets of crash counts or crash disparity for different crashes by the different road entities" (Participant 2). "Typically crash data that would be housed by state [departments of transportation] or state departments of public safety" (Participant 9). "Crash data are used to understand weather impact on crashes" (Participant 21). "I use traffic crash data to conduct effectiveness evaluations of countermeasures using the naive before and after and the empirical bayes (EB) methodologies" (Participant 5; Hauer et al. 2002). Approximately 11 percent of participants (N=4) stated that they used naturalistic driving data as safety data in their research; this type of data assesses human behavior. "We recently had a few projects that focused on naturalistic driving study data" (Participant 10). Additional types of safety data used for research included bicycle data at 5 percent of the participants (N=2), roadway inventory data at 5 percent (N=2), and traffic volume data at 9 percent (N=3). It was also noted that real-time data were utilized by researchers and educators to a limited extent in the context of this study. Applications that require real-time data include connected and autonomous vehicles, which remain novel research areas.

3.4.4 How Are Professional Engineers Using Safety Data?

The most frequently mentioned type of safety data by transportation industry professionals was collision data, with approximately 87 percent of participants (N=20) stating that they utilized collision data for their jobs. In addition, every participant cited collision data in their response when asked to identify the primary application of safety data in the workplace, and a specific application mentioned was the collision study. “We use crash data to know where crashes have occurred. What happened in the crash? Who was involved? What are the primary contributing circumstances? What was the reason for the crash? Was there any impairment involved?” (Participant 15). Another application involved roadway improvements to mitigate collisions. “We do a lot of analysis related to collisions, attributes of collisions, and looking for patterns in collisions” (Participant 17). “I have to analyze the data so it can be as small as looking at one intersection and looking at that data to fill out applications for highway improvement programs and funding” (Participant 16). Practitioners used collision data to analyze fatality records. “We use the fatality analysis reporting system. We can get into more detail because it’s a fatal crash and so there’s thousands of elements in there, and quite often there’s toxicology [or] prior driving record; there’s a lot more depth to that level of data” (Participant 18). Participants also pointed out that they utilized collision data to study roadway segments. “We most frequently use crash reports that allow us to summarize safety performance and crash patterns within a study area for a multi-year timeframe” (Participant 38). “Safety data is useful for identifying hot spots or high priority corridors for improvements” (Participant 12). As a logical next step following the study and analysis of these data, participants mentioned that collision data were then utilized for countermeasures. “Crash counts are used for crash analyses and identification of counter measures. Traffic counts likewise (are) used for computing the effectiveness of countermeasures” (Participant 35).

Another type of data currently being used by transportation professionals is traffic volume data, since approximately 27 percent of the participants (N=6) answered that they considered traffic volume data as a type of safety data used in their line of work. “Traffic volume data is a big one” (Participant 20). “We use traffic volumes to figure out how much traffic is utilizing each facility and we use traffic data to kind of put a rate to the crashes to really normalize from a statistical standpoint of what is a high crash location” (Participant 15). “We use traffic count information” (Participant 24).

Roadway inventory data were a third type considered to be safety data by participants. Approximately 18 percent of the participants (N=4) considered roadway inventory as part of safety data in their line of work. “During a crash, I can link the roadway inventory data to look at what type of conditions existed on the roadway, the number of lanes, the speed of the roadway…" (Participant 20). Roadway conditions and physical roadway components such as curvatures, signals, and pavement data are also considered roadway inventory data. "We have our entire database which gives us data on curves, signs, signals, and lane widths" (Participant 7).

According to participants, there were challenges when utilizing safety data. One of the biggest challenges was that safety data were not always accurate or precise. "There are some accuracy issues because there's so many hands that touch it" (Participant 14). "The challenges are that the table of safety data doesn't give you the whole picture of what happened in the crash, so if you need to know more details then you need to request the actual crash report and read those" (Participant 16).

The availability of safety data was be another challenge. "As an engineer, and I am sure you experienced this, you want more data than you are going to get" (Participant 20). Participants also mentioned cost as a challenge related to data. They stated that it was often expensive to collect data, which can be a challenge because of budget limitations.

3.5 Discussion

3.5.1 Similarities and Differences among the Three Different Professional Settings

This research sought to determine the similarities and differences of safety data usage by educators, researchers, and industry practitioners. Based on the feedback provided from the interview and survey participants, the use of collision data as safety data represented the most common similarity. However, discrepancies were present in the ways that collision data were applied. In practice, collision data are mostly used for technical transportation designs that serve as countermeasures to reduce vehicle collisions. Conversely, collision data usage in academics primarily enhances safety learning by serving as an analysis tool. Collision data in research are being used to study collisions and develop methods that could potentially mitigate crashes. The results suggest that usage of collision data in research and practice are most similar to one another.

The similarities of data usage by academicians and practitioners are significant because there needs to be a connection. Recognizing that similarities exist in these settings suggests that there is value to developing methods that could better transfer knowledge and skills regarding safety data from an academic setting to a practice setting in the transportation engineering field. Collision data have been identified as the most commonly used type of safety data for these three settings, so continued usage of collision data within each of these settings should provide students who transition into researchers or practitioners with some level of familiarity.

On the other hand, there is a slight difference in the coverage of traffic volume data within academia and practice. The use of these data in teaching and research is generally less than the usage of traffic volume data in transportation practice. In order to guide student learning to practice, the inclusion of this type of data should be embedded to a larger scale in academia. Another type of safety data that is common in transportation practice is roadway inventory data. These data include items such as lane width, traffic signs, and sign conditions (Khattak et al. 2000). The use of these items relates to the safety of transportation infrastructure, but the use of roadway inventory data does not seem to be well-represented in teaching and research. Incorporating roadway inventory data activities as part of in-class assignments or projects would provide the student with additional insight into industry practices.

The next section discusses some of the potential methods to incorporate safety data into course curricula based on the results from this research and the literature review.

3.5.2 Methods to Include Safety in Course Curricula

One method of teaching safety to engineering students in a technical manner would be to incorporate quantitative safety data, such as historical crash data, into the course curricula. This type of data could be used to design infrastructure and educate engineering students about the importance of this information and how it can be analyzed. Safety data are collected in the field, but they can be costly (Bertini et al. 2005). However, in many circumstances this type of data may already be available to the public from local, regional, or state departments of transportation.

Because of the costs associated with the collection of safety data, the utilization of existing technology and software applications and the development of new methods to cost effectively collect safety data should be considered. For instance, a mobile device has been previously utilized to collect data by using software to collect transportation information such as pedestrian counts, vehicle counts, and GPS coordinates. Researchers from Portland State University "incorporated the data collection 'process' into the classroom and laboratory by creating tools using equipment readily available at low cost and custom open-source software. [They] developed software for handheld computers which, when coupled with global positioning system (GPS) devices, allowed a great deal of transportation-related data to be collected rather easily" (Bertini et al. 2005). While this device and approach was a potential low-cost alternative for collecting data at the time, adopting software in today's smartphones to collect significant amounts of transportation data has become increasingly common.

The inclusion of civil engineering projects or assignments that require safety data collection has the potential to provide students with an understanding of the uncertainty and complexity of safety data. In previous research, historical crash data from Los Angeles, California, were used to determine whether safety improvements would mitigate secondary crashes, or whether it would not be cost efficient to apply these mitigations because of the infrequent nature of these events (Moore et al. 2004). With the management of this safety data, researchers were able to provide a solution based on the crashes occurring in the area. In a similar manner, civil engineering students could learn about the importance of safety by taking on course projects or assignments that required similar applications of safety data. This type of project would encourage students to rely on real-world safety information, such as crash data, to solve engineering problems. One potential project could be the identification of existing roadways with multiple crashes. Students would be provided with supporting data and asked to design a plan to mitigate crashes. Another project focus area could be roadway inventory analysis. Student groups would be assigned a section of urban roadway and asked to perform a roadway inventory count. Using those data, locations for improvement could be identified and followed by a discussion that examined the costs and benefits resulting from any action taken.

The use of safety data to complete tasks and assignments in civil engineering courses serves as a positive way to incorporate safety material into existing course curricula. This approach allows transportation students to develop greater awareness of how safety data might be utilized to make design decisions. On the basis of the results of this study, the challenge with using safety data in civil engineering courses is not always the data but the ways they should be used in the course to highlight the importance of safety to students. Further research is needed to determine the methods required to include this type of data in engineering education and the best approach to increase the understanding and importance of safety data to transportation students.

3.6 Conclusions

This research closely examined and identified the current uses of safety data within transportation research, practice, and education. Safety data are well utilized in transportation practice and research but to a lesser degree in the classroom environment, so an opportunity exists to incorporate safety data into transportation curricula.

The most common type of safety data used within these three professional settings is collision data. This type of data is accessible, relatable, and applicable within each of the education, research, and practice settings. With regard to roadway inventory and traffic volume data, these two types of safety data are used on a regular basis in transportation practice, but less so in the teaching and research settings.

Practitioners and academicians provided similar definitions of the term “big data” as a large collection of data compiled for use in transportation applications. However, practitioners explained the term describing technical applications related to transportation and traffic engineering whereas academicians explained the term by associating it with novel research topics such as connected vehicles and the use of real-time data. This study addressed “big data” and their inclusion and utilization among the three professional settings, but a more in-depth study regarding “big data” and their use in academics would be beneficial to learn more about current trends and allow educators to develop better curricula offerings that reflect transportation industry practices. In addition, the continual evolution of technology promises new methods and technologies to collect, assess, and use safety data in the future. A follow-up study is recommended to assess potential improvements and deficiencies regarding the inclusion of safety data in course curricula, not only in terms of design but also in terms of operations, maintenance, planning, and construction. Moreover, with “big data” nearly considered safety data in transportation engineering, there are no studies regarding the use of “big data” in both academic and practice settings. A more in-depth study should be conducted to compare the findings with this study and learn more about how the availability and application of real-world data sources impact the learning and studies of transportation engineering students.

4. Safety Data Analysis and Methodology

4.1 Introduction

Analyzing crash data is an effective way to identify and address traffic safety problems. However, for a meaningful analysis, transportation analysts need to find out what types of data are collected that can be used for safety analysis, be familiar with different data features and what can be done with them, and use appropriate statistical and data mining techniques.

This chapter focuses on explaining different data sources that are gathered by public agencies, showing how these data sets can be fused and employed to predict the number and severity of crashes, what types of statistical approaches can be used to analyze the data, and which data types are missing for further safety analysis. As a case study, Washington state was examined to determine where research protocols could take place, and this chapter focuses on data collected in Washington by various public agencies, including the Washington State Department of Transportation, the police departments of various cities and counties, and weather stations.

4.2 Transportation Safety Analysis

Several common analyses that are undertaken by transportation jurisdictions are as follows (Persaud 2001):

1. Before and after evaluations
2. Identification of hazardous locations
3. Cost-benefit analysis in development of countermeasures
4. Cross-sectional evaluations
5. Risk estimation/analyses/evaluations.

In the following subsections, each of the mentioned analysis will be discussed briefly.

4.2.1 Before and After Evaluations

Before and after evaluations examine the effectiveness of a safety improvement by using data collected before and after the implementation of countermeasures. As an example, Griffith (1999) analyzed the effects of continuous shoulder rumble strips in reducing the number of single-vehicle run-off crashes on freeways. In this study, crash data from California and Illinois were analyzed before and after the implementation of rumble strips. The results of the before and after evaluation showed that using rumble strips reduces such crashes on freeway facilities.

In another study, Renski et al. (1999) evaluated the effects of increasing the speed limit on crash injury severity. In their analysis, they considered changes in single-vehicle crashes on North Carolina interstate highways. They showed that an increase in speed limit from 55 to 60 and from 55 to 65 increased the probability of sustaining minor injuries. However, increasing the speed limit from 65 to 70 did not change the crash severity significantly.

4.2.2 Cost-Benefit Analysis in the Development of Countermeasures

Developing countermeasures for cost-benefit analysis consists of estimating and comparing the costs and benefits of alternative safety improvements (Persaud 2001). In this process, the analyst can determine and rank strategies that will increase transportation safety with respect to the required budget and resources. While cost-benefit analysis is widely used in other principles, its application in safety analysis requires assigning the value of human life, which makes it quite controversial (Elvik 2001).

4.2.3 Cross-Sectional Evaluations

This type of analysis is a highly researched area in transportation safety analysis. The main application of cross-sectional evaluations is to understand the causality of relationships between roadway characteristics (traffic volume, weather condition, rumble strips, etc.) and crash rates by estimating econometrics models. By estimating such models, one can find which roadway attributes contribute to crash frequency. Furthermore, different roadway improvement scenarios can be evaluated and corresponding changes in crash rates can be found.

For example, Abdel-Aty and Radwan (2000) modeled crash frequency based on roadway characteristics. They found that accident likelihood increased with an increase in traffic volume and speed, among other contributing factors.

4.2.4 Risk Estimation/Analyses/Evaluations

In risk estimation, risk levels are measured, monitored, and evaluated by determining the probability of accidents and their consequences. The process computes a number of travel risk performance measures for roadways.

4.3 Data Sources

Before performing safety analysis, the analyst needs to know what the available data sources are and what can be done with them. This research identified the following data sources in Washington state:

- Crash records
- Roadway characteristics
- Traffic data
- Weather data.

We expect similar data types to be available to safety analysts in other states in the Northwest region. In the following subsections, each data source is detailed.

4.3.1 Crash Records

Crash records are among the fundamental sources of data for safety analyses. This information is typically collected by police departments, and recorded data are entered in police report sheets corresponding to a single crash. Table 4-1 summarizes different data types that are recorded for each crash in Washington state.

Table 4-1: Crash record data

No.	Name	Data Type	Description
1	Road type	Category	Mainline, ramp, couplet, frontage
2	Region name	Category	Eastern, North Central, Northwest
3	Mile post	Number	Mile post of crash
4	Quarter	Category	A three-month period of the year: Q1, Q2, Q3, Q4
5	Month	Category	Months of year
6	Day of week	Number	Monday to Sunday: 1 to 7
7	Time	Time	Time of crash occurrence
8	Injury type	Category	No injury, injury, serious injury, fatal
9	Collision severity	Category	Property Damage Only (PDO), injury, fatal
10	Number of injuries	Number	Number of injured people in the reported crash
11	Number of fatalities	Number	Number of fatalities in the reported crash
12	Number of vehicles	Number	Number of vehicles involved in the reported crash
13	Junction relationship	Category	Role of intersection in the accident
14	Weather	Category	Blowing sand, clear or partly cloudy, fog or smoke, overcast, raining, severe crosswind, snowing, sleet or hail or freezing rain, other
15	Roadway surface conditions	Category	Dry, ice, snow/slush, wet, standing water, other
16	Lighting conditions	Category	Dark-no street lights, dark-street lights off, dark-street lights on, dawn, daylight, dusk, other
17	Roadway characteristics	Category	Curve and grade, curve and level, curve at hillcrest, curve in sag, straight and grade, straight and level, straight at hillcrest, straight in sag
18	First impact location	Category	Median, intersection, off ramp, on ramp, right shoulder, left shoulder, median shoulder
19	Urban/rural	Category	Urban, rural
20	First vehicle movement	Category	Backing, change lanes, illegally parked or stopped, make U-turn, merge, move straight, passing, stopped in traffic, turning, vehicle position in previous accident, vehicle/trailer in tow

No.	Name	Data Type	Description
21	Vehicle type	Category	Bus or motor stage, motorcycle, passenger car, pickup and panel truck, school bus, scooter bike, truck double trailer, truck trailer, truck, truck tractor and semi-trailer
22	Vehicle action	Category	Backing, changing lane, go straight, go wrong way, make turn, merge, passing, slowing, starting from parked position, stopped
23	Driver age	Number	Age of driver
24	Gender	Category	Male or female
25	Driver injury type	Category	Dead at scene, died at hospital, evident injury, no injury, non-traffic fatality, possible injury, serious injury
26	Involvement cause	Description	Type of event that may have led to the involvement of a vehicle in a collision.
27	Driver restraint	Category	Lap and shoulder used, lap belt used, no restraints, shoulder belt used, unknown
28	Event for a vehicle involved in a collision	Category	Cargo loss or shift, collision involving: fixed object, motorcycle, other object, parked vehicle, pedal cyclist, and pedestrian. Explosion or fire, jackknife, rollover, ran off the road

4.3.2 Roadway Characteristics

Roadway characteristics are often available in GIS format, which includes many attributes as well as graphical representations available at the WSDOT website. For each type of data, some of the main attributes among different available attributes are summarized in table 4-2.

Table 4-2: Road characteristics data

No.	Name	Data Type	Description
1	Acceleration lanes	GIS	Location, length and width
2	Access control	GIS	A code depicting the type of access control present
3	Alignment, horizontal	GIS	Horizontal curve data
4	Alignment, vertical	GIS	Vertical curve data
5	Bridges	GIS	Location and characteristics

6	Coincident routes	GIS	Routes where two (or more) state routes share the same physical alignment
7	Design speed horizontal curve, greater than or equal to 20	GIS	The speed for which a horizontal curve is designed and is greater than or equal to 20 mph
8	Design speed horizontal curve, less than 20	GIS	The speed for which a horizontal curve is designed and is less than or equal to 20 mph
9	Design speed vertical curve	GIS	Design speed of vertical curves
10	Divided highway	GIS	Information of divided highways
11	Highways of statewide significance	GIS	Location along the state routes
12	Intersections	GIS	Location, configuration, illumination, traffic control type, legal speed limit
13	Lanes	GIS	Increasing or decreasing number of lanes, widths, lane surface type
14	Legal speed limits	GIS	The maximum speed as is set
15	Maintenance areas	GIS	Location of maintenance areas
16	Medians	GIS	Median widths and barrier and surface type
17	At grade railroad crossings	GIS	Location and speed limit of railroad crossings that are at the same grade as the state route
18	Shoulders	GIS	Shoulder widths, and surface type
19	Special use lanes	GIS	Lane types specified for a particular mode of travel
20	Terrain type	GIS	The contour of the roadway showing the train type along the route
21	Tunnels	GIS	Location, length, structure, and number of lanes that increases or decreases
22	Turn lanes	GIS	Lanes designated for turn movements at intersection locations
23	Undercrossing	GIS	Undercrossing along the State Routes
24	Urban/Rural	GIS	Urban or rural indicators along the State Routes
25	Weigh stations and rest areas	GIS	Location and type of rest areas

It is noted that roadway characteristics are tied to geospatial coordinates. At the same time, crash records include information on the roadway and milepost of the crash. Therefore, by fusing these two data sets, crashes of different types can be placed on roadways with different characteristics. This data set can be used to predict the impacts of different roadway characteristics on the number of crashes and their severity.

4.3.3 Traffic Data

Traffic volume, speed, and heavy vehicle percentage are among contributing factors to the number and severity of crashes. While at a certain location, volume and speed data are available over time, this is not the case throughout the transportation network. Table 4-3 summarizes the available traffic data throughout Washington state's transportation network.

Table 4-3: Travel data

No.	Name	Data Type	Description
1	Daily vehicles miles traveled (DVMT) information	Number	Aggregate of DVMT of the state, cities, counties
2	Annual average daily traffic (AADT)	Number	Data received from data collection sites
3	Truck percentage	Number	Showing percentages of SNLG, DBL, TRPL
4	Speed Limit	Category	Number of vehicles within 5 mph increment categories ranging from 0 to 95

Traffic data, when integrated with crash records and roadway characteristics data, can improve the accuracy of crash prediction models.

4.3.4 Weather Data

Weather conditions play an integral role in the number and severity of crashes. Weather conditions not only change the friction and drivability of the road but also influence visibility. The Washington State DOT does not collect weather data for its roadways throughout the year. As a result, weather data needs to be collected from other resources, such as the ASOS website. ASOS's automated data collection system provides 1-minute, 5- minute, and hourly weather information. ASOS collects and archives data from weather stations at airports. Different types of available weather information are summarized in table 4-4.

Table 4-4: Weather information

No.	Name	Data Type	Description
1	Air temperature	Number	Air temperature at the station
2	Dew point	Number	Dew point temperature
No.	Name	Data Type	Description
3	Relative humidity	Number	Relative humidity in percentage
4	Wind direction	Number	Wind direction in degrees from north
5	Wind speed	Number	Wind speed in knots
6	One hour precipitation	Number	Precipitation that is measured in one-hour intervals
7	Visibility	Number	Visibility in miles
8	Wind gust	Number	Wind gust in knots
9	Cloud coverage	Number	Coverage of clouds in different sky levels

4.4 Safety Applications Using the Available Data

This section talks about some of the applications that are possible with the available data. However, as different sources of data are available, the analyst needs to integrate them to have one data set that includes the necessary attributes. Therefore, the following subsections discuss how different data sources should be integrated to allow accurate crash prediction.

4.4.1 Data Preparation

Several data preparation steps, such as data cleaning and integration, should be performed before the collected data are processed. Data cleaning is necessary for almost every data analysis application and not just for safety studies. In this step, the analyst needs to identify the missing information and filter out unnecessary data records. The second step is data fusion, in which the analyst integrates different attributes from different data sources to have a unique data set.

One common technique for developing crash prediction models is to divide road sections into several segments. Then, crashes on any road segment are aggregated and analyzed on the basis of roadway geometric characteristics, traffic data, and weather conditions on each segment. Generally, segmentation of the roadways can be done in different ways (Cafiso et al. 2013):

1. Completely homogeneous, i.e., a new segment will start if any of its attributes changes
2. Based on curvature, i.e., a new segment starts when a tangent section transitions into curve
3. Homogeneous based on AADT, i.e., a new segment starts when AADT changes
4. Fixed length segment, i.e., all the segments are of fixed length (typically 1- or 2-mile)
5. One curve and one tangent, i.e., each segment consists of a curve and a tangent.

After defining a segmentation technique, the analyst should assign all available attributes to each segment separately so that each row of data contains all necessary attributes for each segment. Among different safety analysis types, two main safety applications (i.e., crash rate prediction and crash severity prediction) are discussed in the following sections.

4.4.2 Crash Rate Prediction

The goal in crash rate prediction is to estimate the number of crashes on a roadway segment based on a number of independent variables, as discussed above. Classic linear regression models are a viable option to depict the relationship between the expected crash frequency (or rate) and several explanatory variables. The model specification can be expressed as follows (McCullagh and Nelder 1989):

$$E(y_i) = \mu_i = L_i \sum_{j=1}^{J} \beta_j x_{ij} \tag{1}$$

where:

y_i is crash rate on road segment i,

$E(y_i)$ is the expected crash rates on road segment i,

L_i is the length of road segment i,

x_{ij} is the jth explanatory variable on road segment i, and

β_j is the corresponding coefficient to x_{ij} that will be predicted.

While linear regression models are simple and straightforward for predicting crash rate, more elaborate models, such as Poisson or negative binomial distributions, can be used for predicting the number of crashes, and logit or probit models for crash severity prediction. One can generalize these models by utilizing a smooth and invertible linearizing link function. The function transforms the expectation of the response variable, μ_i, to its linear predictor:

$$g(\mu_i) = \sum_{j=1}^{J} \beta_j x_{ij} \tag{2}$$

where g(.) is a monotonic, differentiable link function that is used to connect the linear predictor of the explanatory variables to the expected crash frequency. While this model is linear, it can be enhanced by using nonlinear parameters to define broader model specifications called generalized nonlinear regression models (GNM).

The GNM-based models utilize a customized relationship to represent the relationship, U(x), among traffic accidents and their contributing factors. Polynomial, exponential, parabolic, logarithmic, and interacting functions (to name a few) may be used to depict proper data features. Statistical analysis of crash data will be used to determine the format of U(x) for each explanatory variable. Aggregating the nonlinear predictors for all the independent variables, equation (1) can be rearranged as:

$$E(y_i) = \mu_i = L_i \sum_{j=1}^{J} w_j U_j(x_{ij}) \tag{3}$$

In the above equation, w_j is the corresponding weight for $U_j(x_{ij})$. Consequently, the GNM link functions become:

$$g(\mu_i) = \sum_{j=1}^{J} w_j U_j(x_{ij}) \tag{4}$$

Although generalized nonlinear models allow both dependent and independent variables to be nonlinear functions, they are still linear in parameters. Such specifications can all be analyzed in the context of Ordinary Least Square (OLS) regression.

As an example, the estimated model for predicting crash rates on Washington state routes is shown in tables 4-5 and 4-6 (Hajbabaie and Wang 2016).

Table 4-5: Definition of variables

Variable	Description	Notation
Roadway geometry		
Number of Lane (NOL)	≤ 2	NOL1
	2 < NOL < 4	NOL2
	NOL ≥ 4	NOL3
Lane Surface Type (LST)	Portland Cement Concrete	LSTPD
	Others (Asphalt)	LSTOD
Outer Shoulder Type (OST)	Asphalt	OSTAD
	Others	OSTOD
Median Surface Type (MST)	Soil	MSTSD
	Others	MSTOD
Speed Limit (SPL)	≤ 60 mph	SPL1
	> 60 mph	SPL2
Weather Characteristics		
Road Surface Condition (RSC)	Dry	RSCD
	Non-Dry	RSCND

Table 4-6: Estimated coefficients of the GNM crash rate prediction model

Variables	Estimate	Std. Error	t-value	Pr (>|t|)	
Intercept	2.96600	0.10130	29.29100	0.00000	***
Numerical variables					
AADT/Lane	0.00005	0.00000	12.41300	0.00000	***
Variables	Estimate	Std. Error	t-value	Pr (>|t|)	
Inner shoulder width (ft)	-0.03034	0.01318	-2.30200	0.02147	*
Median width (ft)	0.00031	0.00015	2.08400	0.03730	*
Absolute value of grade	-0.08305	0.03341	-2.48600	0.01303	*
Categorical Variables					
No. of lanes (equals to 1 if more than 4; 0 otherwise)	0.62940	0.08706	7.23000	0.00000	***
No. of lanes (equals to 1 if more than 2 and less than 4; 0 otherwise)	0.17080	0.05540	3.08300	0.00208	**
Outer shoulder type (equals to 1 for asphalt; 0 otherwise)	-0.15220	0.08736	-1.74200	0.08168	.
Inner shoulder type (equals to 1 for asphalt; 0 otherwise)	-0.13680	0.08732	-1.56600	0.11750	
Lane surface type (equals to 1 for Portland cement; 0 otherwise)	-0.17220	0.04492	-3.83400	0.00013	***
Road surface condition (equals to 1 for dry, 0 otherwise)	-2.98900	0.07004	-42.68200	0.00000	***
Interaction variables					
Length of the curve in a segment × outer shoulder width (equals to 1 if less than 4 ft)	0.18640	0.13020	1.43200	0.15242	
Road surface condition (equals to 1 for dry, 0 otherwise) × AADT/Lane	0.00007	0.00000	16.03800	0.00000	***
R^2		0.75			

4.4.3 Crash Severity Prediction

Crash severity is often available in police reports. For predicting such categorical data, the logistics regression approach is generally used (Bham et al. 2011). Such models can handle bivariate response variables, i.e., variables with two possible values, and can be extended and applied to variables that can take a discrete set of values reflecting multiple categories. Since the response variable is nominal (unordered), a generalized logit model is suitable.

In the available data set, we can define three types of crash severity: fatal, injury, and property damage only (PDO) crashes. Therefore, the crash severity type, denoted by Y (response variable), can be stated as follows:

$$Y = \begin{cases} 1 & \text{if crash is fatal} \\ 2 & \text{if crash is injury} \\ 3 & \text{if crash is PDO} \end{cases} \tag{5}$$

By using the above categories for crash severity, we can use other variables, such as road conditions, traffic characteristics, and weather conditions, as explanatory variables. Consequently, we can write the GNMbased logit probability function as follows:

$$\Pr(Y_i = k) = \frac{1}{1 + \sum_{k=1}^{K-1} e^{U_{ki} w_k}} \tag{6}$$

The above equation shows the probability of having a specific severity type (k) based on the explanatory variables. It should be noted that in such estimations, one of the categories is designated as the base condition, and the probability of occurrence for other categories will be relative to the base case. As an example of this type of model, the crash severity prediction model that was developed for the Washington state routes, is shown in tables 4-7 and 4-8 (Hajbabaie and Wang 2016).

Table 4-7: Estimated coefficients for predicting the probability of PDO crashes

Variables	Estimate	Std. Error	t-value	Pr(>|t|)	
Intercept	-1.77400	0.31960	-5.551	0.000000	***
Numerical variable					
AADT/Lane	0.00027	0.00002	13.92	0.000000	***
Inner shoulder width	0.14890	0.04641	3.209	0.001334	**
Curved portion in a segment	-2.73000	0.28600	-9.545	0.000000	***
No. of left directed curve in a segment	-0.61500	0.09104	-6.755	0.000000	***
Absolute value of grade	0.47370	0.07644	6.197	0.000000	***
Categorical Variable					
No. of lane (equals to1 if more than 4; 0 otherwise)	1.22800	0.33790	3.635	0.000278	***
No. of lane (equals to 1 if more than 2 and less than 4; 0 otherwise)	1.30800	0.17370	7.532	0.000000	***
Inner shoulder type (equals to 1 for asphalt, 0 otherwise)	-1.87700	0.25940	-7.234	0.000000	***
Outer shoulder type (equals to 1 for asphalt, 0 otherwise)	1.60700	0.26820	5.992	0.000000	***
Road surface condition (equals to 1 for dry, 0 otherwise)	0.49920	0.23080	2.163	0.030552	*
Interaction variables					
Road surface condition (equals to 1 for dry, 0 otherwise) × AADT/Lane	0.00009	0.00004	2.373	0.017650	*
R^2		0.41			

Table 4-8: Estimated coefficients for predicting the probability of injury and fatal crashes

Variables	Estimate	Std. Error	t-value	Pr(>|t|)	
Intercept	-0.69680	0.26510	-2.628	0.00859	**
Numerical variable					
AADT/Lane	35.07000	3.27000	1.07E+01	0.00000	***
AADT/Lane^2	-8.45500	2.42500	-3.486	0.00049	***
Inner shoulder width (ft)	-0.12810	0.02534	-5.057	0.00000	***
Logarithm of median width (ft)	0.30570	0.07035	4.346	0.00001	***
Curved portion in a segment	-1.63400	0.26060	-6.269	0.00000	***
Absolute value of grade	0.34090	0.07063	4.827	0.00000	***
Categorical Variable					
No. of lane (equals to 1 if more than 4; 0 otherwise)	1.27600	0.22220	5.74	0.00000	***
No. of lane (equals to 1 if more than 2 and less than 4; 0 otherwise)	0.89420	0.11500	7.774	0.00000	***
Outer shoulder width (1 if greater than or equal to 10ft; 0 otherwise)	-0.21100	0.10030	-2.104	0.03534	*
Inner shoulder type (equals to 1 for Portland cement; 0 otherwise)	0.63340	0.30290	2.091	0.03651	*
Road surface condition (equals to 1 for dry; 0 otherwise)	0.00003	0.00002	1.876	0.06067	.
Interaction variables					
Road surface condition (equals to 1 for dry; 0 otherwise) × AADT/Lane	0.00007	0.00001	14.113	0.00000	***
			0.2		

4.5 Data and Methodology Deficiencies

This section focuses on data attributes that are missing and, if collected, can significantly contribute to more accurate estimation of crashes. For instance, detailed data on vehicle acceleration/deceleration rates, braking, and steering information could improve crash severity prediction. In the remainder of this section, data and methodology deficiencies are discussed.

4.5.1 Segmentation Issues

As mentioned in Section 4.4.1, for performing crash prediction analysis, the analyst needs to assign different available data to the segments of roads. Therefore, we need to divide each road into several segments and study the causality relationships between these segments' characteristics (e.g., curvature, surface type, precipitation) and the crash rates or crash severity.

Crash locations are based on police reports. Oftentimes, the location is estimated on the basis of the mile post and as such is not accurate. Furthermore, the number of accidents is proportional to the segment length. Therefore, if segments are short (e.g., less than 100 feet), the probability of an accident is very low as the crash occurrence can be considered a rare event. On the other hand, using long sections to avoid these issues will sacrifice homogeneity.

The literature does not show any consensus on a threshold for short segments. Miaou and Lum (1993) found that segments shorter than 80 meters may introduce bias into the estimation of linear models; however, the problem wasn't found when Poisson distribution was used. Ogle et al. (2011) showed that segments shorter than 60 meters may yield uncertain results in crash analyses.

The relationship between crashes and road geometry has also been used by some studies to address the segmentation problem. For instance, Cenek et al. (1997) used a fixed segment length of 200 meters. Cafiso et al. (2008) used homogeneous sections (based on curvature and roadside hazards) with different lengths. They concluded that models that contain geometry and design consistency variables are more reliable than those that do not. Other methods for aggregating segment data have also been suggested with the objective of avoiding segments that are too short. For example, Koorey (2009) proposed combining curves and tangents when the radius of curves exceeds a predetermined threshold value. Cafiso et al. (2013) compared five different segmentation techniques with three different model forms. The best results were obtained for the segmentation based on two curves and two tangents.

As discussed, segmentation has a great influence on the quality of safety analysis; however, the best segmentation depends on the available data.

4.5.2 Time-Varying Explanatory Variables

Another issue that an analyst may face in using the available data is an inability to capture changes in the value of dynamic explanatory variables over the aggregation period. In other words, for performing safety analysis, data should be aggregated over certain study periods (e.g., a year) as incidents are rare events. However, this aggregation results in the loss of important information on the variation of explanatory variables within the study period. For example, traffic flow rate varies continuously over time; however, AADT is often used as an explanatory variable to describe the effects of traffic volume on crash rate or severity.

Similarly, weather data need to be aggregated. Average precipitation over a segment is found and used as a general indication of the amount of precipitation. This can introduce error in model estimation because of the unobserved heterogeneity (Washington et al. 2010). Furthermore, weather data are collected at certain locations, and for each roadway segment, the information from the closest station is used. Given the distance between weather stations (especially in Washington state), weather data need to be used cautiously.

4.5.3 Temporal and Spatial Correlation

When segments of a road are analyzed over different periods of time, a correlation between different observations is likely. In other words, similar unobserved factors may exist in different segments of a specific road that are not considered in the analysis. The correlation may exist over space (spatial correlation) or over time (temporal correlation). Such correlations will result in parameter-estimation problems (Mountain et al. 1998; Ulfarsson and Shankar 2003; Sittikariya et al. 2009).

4.5.4 Under-Reporting

Crashes are rare events, but less severe crashes or crashes with no injuries are less likely to be reported. This may lead to the under-reporting of less severe crashes. Recent research studies have revealed that under-reporting of crashes introduces bias into the estimation of count data models (Kumara and Chin 2005; Ma 2009). Although the magnitude of the under-reporting rate for each severity level may be unknown, this issue should be considered in the model estimation process (Lord and Mannering 2010).

4.6 Summary

The main objective of this chapter was to better understand the available data sources that can be integrated for a more accurate estimation of crash rate and severity. The focus of this chapter was on a case study of data collected in Washington state. Among different data sources, crash rates, roadway geometric characteristics, traffic data, and weather data can be fused to form a single dataset for safety analysis. The data set can be used to estimate the number and severity of crashes on various roadway segments. Roadway segmentation plays an important role in the accuracy of crash prediction models, yet there is no consensus on the best segmentation method, nor is there agreement on a threshold for short segments.

Since crashes are rare events, data need to be aggregated over an extended duration of time. Therefore, dynamic variables, such as weather conditions and traffic flow rate, are converted to point estimations (e.g., expected values). This aggregation may introduce some error into crash prediction models. Furthermore, crash prediction models are prone to under-reporting of less severe crashes.

Additional detailed data on vehicle speed, acceleration/deceleration rate, heading, and braking can be used to improve the accuracy of crash prediction models; however, such data attributes are still very expensive to collect. Connected vehicle technology is expected to greatly facilitate access to such data attributes.

5. Assessment and Training (Practitioner Perspective)

According to the USDOT's Bureau of Transportation Statistics, approximately 76 percent of all roads in the United States, based on total mileage, are owned and maintained by local agencies. Local roadway networks vary from a network of urban city blocks to thousands of miles of paved or gravel roads. Unfortunately, the local roadway network also experiences the highest overall crash rates. Local agencies responsible for these roadways often have limited resources, staffing, or knowledge of safety tools, though advancements in data collection capabilities have allowed these agencies to collect significant amounts of safety data. There is an immediate need to find out what types of safety data are being collected, what types of safety analyses can be conducted with the collected data, and the engineering or educational approaches that could be implemented to meet the safety objectives. A survey was developed and distributed to local transportation practitioners with the objective of identifying agency challenges and resources available for data collection and the analysis needed to address roadway safety. With the information collected from the survey, a three-part training tool was developed for broad dissemination to practitioners. Learning how transportation practitioners are currently collecting, analyzing, and applying safety data is crucial to providing efficient and effective tools that can help them to improve the ways that road safety can be addressed.

5.1 Introduction

According to the USDOT's Bureau of Transportation Statistics, approximately 76 percent of all road miles in the United States are owned and maintained by local agencies such as towns, counties, highway districts, municipal planning organization (MPOs), and other municipalities (USDOT 2014). Based on the Fatality Analysis Reporting System (FARS) report for 2013, rural roads contributed to approximately 54 percent of all fatal crashes in the United States (NHTSA 2015). In recent years, the transportation industry has encouraged state departments of transportation (DOTs) to increase their level of interaction with local agencies regarding road safety planning and programming.

Historically, most states have focused their safety planning efforts and funding on state highways. This targeted analysis and investment on state highways rather than local roadways is likely due to a number of institutional factors. Limited data in the form of crash information, roadway characteristics, and traffic control devices on local roads, or the lack of safety planning expertise at local agencies or of a champion in state DOTs for expanding opportunities for local agencies are some of the reasons that state DOTs often focus on their own road system. Procedural requirements may discourage participation by local agencies in the competitive state-level highway safety program, since many local agencies simply lack the administrative and financial resources to proactively address road safety problems and therefore must rely on the assistance of state DOTs to address local road safety issues.

Local and rural road owners often have limited financial resources available to implement highway safety improvements. For this reason, it is important that safety improvements return the highest level of benefit. A primary benefit of a safety improvement is to reduce the number of crashes and fatalities, so it is useful for local and rural road owners to understand how a particular safety improvement, or set of safety improvements, can reduce crashes.

Recent advancements in data collection capabilities have allowed transportation agencies to collect significant amounts of safety data. An immediate need is to find out what types of safety data are being collected, what types of safety analyses can be done with the collected data, and what other types of safety data and analysis approaches are required to meet safety objectives. Extensive collection efforts exist with regard to roadway, traffic, licensing, and vehicle data. For example, there are more than five million traffic crashes reported annually in the United States, and vehicle crashes on public highways alone result in more than 840,000 injuries and 1,700 fatalities annually, according to the National Highway Traffic Safety Administration. The documentation process for every single one of these crashes must begin at the scene of the incident with information gathered by a member of the law enforcement community or by the private citizen(s) involved in the crash. This information is subsequently transmitted to a local or state agency for data entry, processing, and aggregation for the purpose of future analysis.

With the increased complexity of various safety data management and analysis activities, and with most transportation agencies faced with limited staff and financial resources, there is opportunity to provide the transportation workforce with the resources needed to effectively understand, manage, and analyze safety data. Safety data collection, management, integration, improvement, and analysis activities are integral to developing a robust data program that leads to more informed decision-making, better targeted safety investments, and overall improved safety outcomes.

5.2 Literature Review

What can local agencies do to aid in reducing fatalities and injury crashes? To achieve the stated domestic goal of 100 percent reduction in fatalities, local agencies must be active participants and take the lead. Local roadway networks vary from a network of city blocks to thousands of miles of paved, dirt, or gravel roads. Unfortunately, local roadway networks experience the highest overall crash rates. Local agencies responsible for these roadways often have limited resources, staffing, or knowledge of safety tools. This situation is compounded when local agencies do not have a safety program.

Even under ideal conditions, addressing safety issues on these extensive rural road networks is difficult, and the lack of resources further complicates the problem. Local agencies often manage their roadways by considering just road maintenance. Safety issues are often ignored or not identified because these networks carry very low traffic volumes. Local roads account for approximately 76 percent of the nationwide road and street network, or about 2.93 million miles. Counties manage about 1.74 million miles of roadway, while cities and townships manage the remaining portion. Safety remains a problem for all local road and street agencies, and safety improvements are needed because fatal crash rates are highest on local roadways.

States are using a variety of approaches to engage local agencies, since they frequently lack the resources to plan and implement road safety projects and programs. State DOTs coordinate through their Local Technical Assistance Program (LTAP) centers to address issues on local road safety or facilitate the distribution of limited funds for local road safety. Many states have developed low-cost treatment options that improve safety on local roads, and local agencies rely on crash databases to determine safety improvement focus areas. Park et al. (2016) identified that local road programs or projects were implemented by the state DOT through both central offices and district office staff, and state DOTs most frequently provided technical assistance and support to local agencies at all project stages. Federal funding was identified as the major source of support in most states for local safety programs, while crash data and risk analysis were identified as the most commonly applied criteria used to determine the funding allocation for local safety programs. Most states included an element in their Strategic Highway Safety Plan (SHSP) that identified and addressed goals and initiatives to improve safety on local roads.

Previous studies have explained how state DOTs engage local agencies in the safety process, as well as determine what organizational characteristics influence how well they accomplish this goal. Indistinguishable between high- and low-performing states have been characteristics such as established partnerships with MPOs, LTAPs, and various coalitions of local agencies, designated staff working as liaisons to local agencies with outreach programs, projects that benefitted local road systems, and DOT leadership support for engaging local agencies in statewide safety planning efforts. Many of these characteristics are foundational elements to engaging local agencies. Characteristics that have been distinguishable among the higher-performing states include professional staff dedicated to supporting local agencies, adopting zero traffic fatalities as a long-term goal, directing HSIP funds to improvements on local road systems, HSIP commitment proportional to the number of serious crashes on local roads, commitment to increase engagement with local agencies as part of the state's SHSP, and adding a systemic component to their HSIPs, including technical assistance to prepare local safety plans and encouragement for multiagency projects (Preston and Storm 2014).

It is the goal of local road safety programs to save lives by reducing fatalities and serious injuries on local roads. States are focused on improving safety on a systemwide level and consider local roadways as another opportunity to achieve the goals and objectives identified in SHSPs. Addressing local road safety issues requires knowledge of various funding mechanisms, access to essential traffic safety data, traffic engineering and safety expertise, and partnerships among and between a wide array of local elected officials, planners, engineers, and other decision-makers. Realizing the complexities of local road safety, many state DOTs offer support in the form of information, training, technical assistance, and project implementation to agencies to assist with local road safety projects.

Previous assessments have revealed a variety of efforts that state DOTs have implemented to improve local road safety. The noteworthy practices serve as a menu of options for DOTs and local agencies to consider when enhancing local road safety. The level of support needed varies across states and depends on the extent of the local road safety problem, the expertise of local agencies within the state, and the resources available for a DOT to provide this support (Gaines et al. 2013).

Safety tools for local agencies that are practical and easy to implement are important resources to have available. However, these tools alone cannot help to reduce crashes if they are not correctly applied. Meeting the safety needs of local agencies is a considerable challenge, given that these agencies operate in an environment of limited resources. The safety practices should be specifically tailored to address the problems and match up with the resources of each agency; there is no one-size-fits-all safety solution, and large financial commitments or complex analyses are not always conducive to implementing a successful local safety program. A documented local roadway safety program can be a proven safety tool, and recognizing the need to implement even a rudimentary safety program is a necessary first step (Wilson 2003).

Past studies have tried to answer the question of how roadway crash data are acquired, stored, and utilized in engineering and management decisions regarding highway projects. Most safety projects are funded by the federal FHWA through the HSIP, and the general flow of the safety data is similar. But there are many differences in the details, especially with computer hardware and software. The methods of data movement between the responder and the DOT often involve an intermediate agency. One of the most pronounced differences is the program used to extract these data for the DOT. Another difference is the transfer of HSIP funding to local agencies. The use of historical crash data in the SPFs also varies from state to state; the older method of only looking at the crash data from a location in question is not uncommon, while the more modern method of using data from similar locations via an Empirical Bayes analysis is becoming more common and is the currently recommended method (see Chapter 2).

5.3 Research Objectives

In order to identify the challenges of addressing road safety and the resources available to local agencies and, in particular, to those in the state of Idaho, it was essential to develop a comprehensive understanding of the needs and priorities with regard to safety data management and analysis. Furthermore, practitioners needed to be provided with a fundamental set of core skills and knowledge required for safety data management and analysis to support local transportation safety decision-making. The assessment of local practitioners and the tools developed to fill the identified knowledge gaps are described in the next section.

5.4 Data Collection Methodology

To identify the resources currently available to local agencies in Idaho, alternative methods were evaluated to determine the ideal approach to take. The use of a survey was selected as the best alternative for this research, since the use of a survey has high relevance to transportation in several specific areas. The focus of this research was on road safety resources, so it was desired to ask agencies and their staff about the resources that they used for data collection and analysis. Regardless of the subject matter to be covered within a survey, transportation surveys serve several purposes (Richardson et al. 1995). The main purpose for this survey was to identify the resources for data collection and analysis, and to identify the challenges that could affect these resources.

The use of an online survey tool provided many benefits. The greatest benefit was that the survey could describe the characteristics of a large population to ensure a more accurate sample from which to draw conclusions and make important decisions. However, even though surveys are a great tool to use, mistakes are a risk and should be avoided. Confusing or misleading questions is one of the most common mistakes. Survey questions must clearly ask a specific and pointed question. When an individual is confused by the question, he or she will typically not answer the question in a way that is useful. Another common mistake is having a question that is too long. It is easy to lose the appropriate meaning when questions are long and the respondent feels overwhelmed or loses focus while reading the question. There are multiple challenges that can be faced when using an online survey, but they are nonetheless beneficial and important.

At the preliminary planning stage of the survey process, several basic issues were addressed prior to proceeding with the design and administration of the survey. A survey target audience was identified, and this group consisted of managers, city engineers, superintendents and directors of public works, street division staff, and traffic division staff from local agencies in Idaho. Their contact information was obtained from various city and county websites or by directly contacting a specific person or agency. Contact information from a total of 250 cities and counties was collected.

The ten-question survey, administered through SurveyMonkey, was designed with multiple types of questions related to road safety. The questions were reviewed to avoid any of the wording mistakes previously mentioned, and the survey was designed to take around 5 to 10 minutes to complete. Some follow-up questions were asked in order to gather more in-depth information from the participant. The questions are shown in figure 5-1.

1. How would you best describe your agency´s familiarity level with the following programs?
2. For your local agency, traffic safety data such as traffic volumes and speeds are collected by: (check all that apply)
3. Does your local agency perform the following analysis to evaluate sites for safety improvements?
4. What software do you use to analyze traffic safety data? (check all that apply)
5. Does your agency regularly use any of the following safety documents? (Check all that apply)
6. In the past 3 years, did your agency receive training in any of the following topics? (Check all that apply)
7. Rank the topic areas of greatest need to your local agency with 1 being the highest need and 6 being the lowest need.
8. If additional resources or funding were made available to address traffic safety, how would you prioritize the following needs?
9. The University of Idaho is developing a series of online "training modules" focused on roadway safety for local agencies in the State of Idaho. What are three essential topic areas / subjects that we must include?
We believe that these training modules will benefit from your future input. Please provide your contact information if you would kindly allow us to reach out to you in the future. Thank you in advance.

Figure 5-1: Survey questions

The survey was sent via e-mail to each of the local agencies. While administering the survey, we encountered the challenge that not all local agencies had a practitioner in charge of road safety. In this scenario, the city or county clerk was asked to provide information related to how their organization manages road safety.

5.5 Survey Results

All responses were collected from local agencies in the state of Idaho. From the 250 cities and counties contacted, 24 agencies responded to the survey, constituting a 9.6 percent response rate. One of the major challenges encountered during the data collection was the actual level of participation. Predicting the level of survey participation is difficult; survey response rates vary widely, and a variety of factors can affect them. The challenge was to work within the existing constraints of the e-mail system, and the first step was to find a way to get e-mail recipients to open and read the e-mail rather than immediately deleting it. Exploring different type of recruitment methods, such as postal mail or telephone contacts, may provide a way to enhance response rates (Watt 1999).

While the number of studies that use e-mail to collect data has been increasing over the past 15 years, the average response rate to surveys appears to be decreasing. A review of 31 studies that used an email survey to collect data reported a mean response rate of 36.83 percent. Making a comparison between different periods of time, the 1995/6 period had seven studies that used e-mail surveys, with an average response rate of about 46 percent, while the 1998/9 period, in contrast, had 13 studies that used e-mail surveys, with an average response rate of about 31 percent (Sheehan 2001). Another investigation identified that the average response rate for studies utilizing data collected from organizations was 35.7 percent with a standard deviation of 18.8 percent (Baruch and Holtom 2008). To overcome the low response rates, some approaches were considered, including personalized email invitations, follow-up reminders, pre-notification of the intent to survey, and simpler formats (Cook et al. 2000; Solomon 2001). Even with the use of these approaches and in comparison with the average response rates of the previous studies, the response rate obtained for this transportation survey was relatively low.

From the 24 agencies that responded to the survey, 41.7% represented a city agency (N=10), 29.2 percent represented a county (N=7), 8.3 percent represented a highway district (N=2), 8.3 percent represented a metropolitan planning organization (N=2), and 12.5 percent were from an undefined agency (N=3). (An undefined agency implied that the agency or responder did not provide any contact information). The participants, based on their geographic location within the state, are shown in figure 5-2.

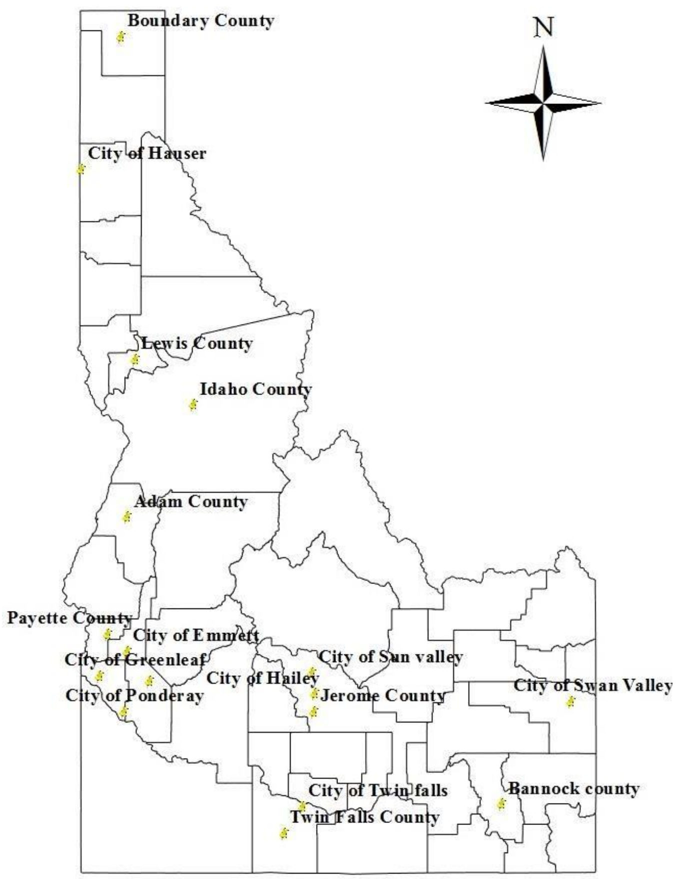

Figure 5-2: Survey participants

Many smaller local agencies noted that a separate entity provided services to manage their roads, and this arrangement was more prevalent when an agency simply did not have a street or traffic division. The most common agency tasked with managing another agency's roads was the local highway district.

5.6 Survey Question Responses

When asked to describe their agency´s familiarity level with specific road safety programs, respondents were provided with the following options and asked to rank their level of familiarity: Toward Zero Deaths (TZD), Highway Safety Improvement Program (HSIP), Highway Risk Rural Road Program (HRRRP), and the Traffic Enforcement Mobilization Agreement (TEMA). Of the four options, the most familiar program was the HSIP and the least familiar program was the TEMA. TZD and the HRRRP were in the midrange of familiarity (see figure 5-3).

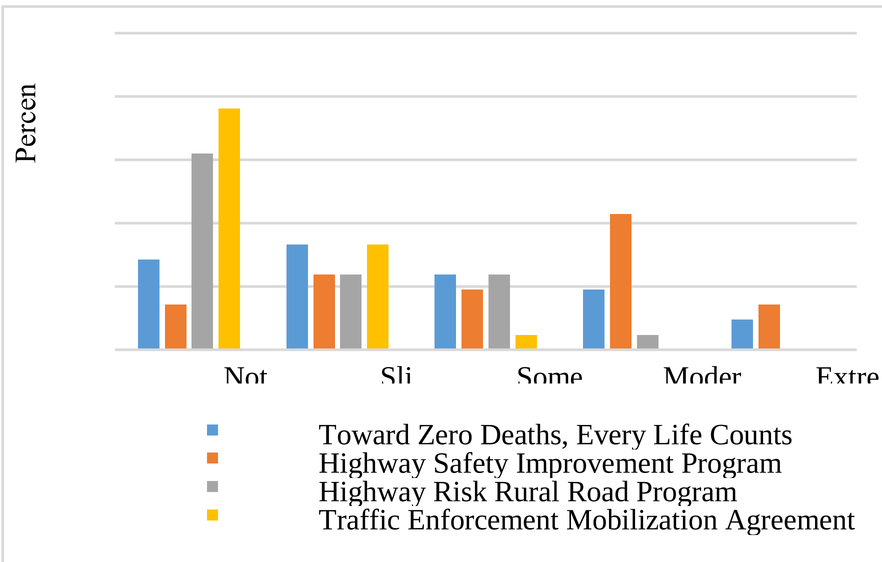

Figure 5-3: Local agency familiarity with safety programs

The collection of traffic safety data, such as traffic volumes and speeds, at their local level varied among internal staff, a consulting or data collection agency, the Idaho Transportation Department, or another local agency. Sixteen local agencies responded that their own internal staff collected traffic safety data, while 15 agencies occasionally requested and received technical assistance from the Idaho Transportation Department. Some of the agencies relied on consultants from other local agencies. The agencies that specified “other” agencies stated that they collected data with help from county and highway district staff.

A local agency was asked whether they perform specific types of analyses to evaluate sites for safety improvements. This question evaluated whether an agency had conducted analyses for a traffic stop sign, traffic control signal warrant, or pedestrian crossing, or had completed a crash analysis. The most common type of analysis used to evaluate sites for safety improvements was a stop sign analysis, followed by a traffic control signal warrant analysis. The analysis least frequently conducted was a crash analysis. Pedestrian crossing analyses were in the middle of the range, with some agencies conducting them while others did not have the capability to do so.

Software applications play an increasingly large role when traffic safety data are analyzed. Responding agencies had the option of identifying the software applications currently in use. Specific options included SafetyAnalyst, WebCARS, Interactive Highway Safety Design Model (IHSDM), and Geographic Information System (GIS) applications. The results showed that the software most used by local agencies to analyze traffic safety data was GIS (N=8). SafetyAnalyst was not identified by any agency. Four agencies noted other options, including the Travel Demand Model - QRS II and various Highway Safety Manual spreadsheets.

Each agency was asked how regularly it used or referenced specific safety documents or resources. The agencies had the option of selecting between the Highway Safety Manual (2010), HSIP Manual (Herbel et al. 2010), High Risk Rural Road Manual, Model Inventory of Roadway Elements (MIRE) (Lefler 2010), Manual on Uniform Traffic Control Devices (2009), ITD Traffic Manual, A Policy on Geometric Design of Highway and Streets (AASHTO 2011), or any other resource. The most commonly referenced resource identified by the agencies was the MUTCD (N=18), followed by the ITD Traffic Manual (N=12). The HRRR Manual and MIRE were the least used. Based on these results, it appears that there is an opportunity to further expose both the HSM and HSIP to local agencies.

Each agency was asked whether it had received training in the last three years in any of the following topics: data collection methods, data analysis procedures, countermeasure identification, cost/benefit analysis, Highway Safety Manual usage, road safety problem identification, or pedestrian and bicycle safety or complete streets. Several agencies (N=11) had received training in pedestrian and bicycle safety or complete streets. For the other training topics, some had received training in data collection (N=4), data analysis (N=3), and cost/benefit analysis (N=3). Training with regard to the Highway Safety Manual was comparably lower than the other suggested topics.

Each responder was asked to prioritize and rank a selected topic area of greatest need to his/her agency. The choices included data collection methods, data analysis procedures, countermeasure identification, road safety problem identification, project prioritization, and project funding. Specific examples were provided for each category to add clarity for each option. The agencies prioritized these topics in the following order: project funding (sources and opportunities), project prioritization (methods and procedures), data analysis (methods and procedures), and road safety problem identification.

In the event that additional resources or funding were made available to address traffic safety, the responders were asked to prioritize the following needs: hire additional staff or expertise, increase training or technical assistance opportunities, increase data collection opportunities or frequency, or enhance data analysis capabilities. The agencies prioritized the needs as follows: hire additional staff or expertise, increase training or technical assistance opportunities, and increase data collection opportunities or frequency.

The final question asked responders to identify three essential topic areas or subjects that would serve as ideal and needed future training topics. The agencies provided many topic alternatives, and the responses were grouped into the following categories: problem identification, countermeasure options, project prioritization, highway and street design, traffic signals and management, road safety audits, and transportation planning, and funding management.

Responders were also given the opportunity to provide their contact information and indicate their willingness to participate in future studies.

5.7 Training Tools

Using the information and insight collected from the survey, a three-part training tool was developed for broad dissemination to transportation practitioners. This tool, which could be used as part of a continuing education training program, was thoughtfully designed to be both dynamic and self-sustainable so that there would be value for those teaching this subject matter related to safety data and safety data management. The development of this tool in the form of a set of transferable PowerPoint presentations was implemented because of its ease of use, accessibility, and distribution. The presentations included components such as definition of learning objectives, identification of reading material, road safety terminology, resources currently available, survey results, local agencies challenges, and recommendations. The three presentations were developed with the intent of being presented as part of a series, but each could also be delivered as standalone presentations.

The presentations themselves were developed in three phases. The first phase consisted of the development of an extensive outline. This helped identify the primary topics for each presentation as well as selected subtopics that would also be included. The subtopics were categorized in terms of relevance and anticipated audience interest. The topics were organized in a logical sequence so that material could be shared in an orderly manner. As each slide deck could be independent from one another, the topics identified for each presentation were selected on the basis of this concept. After these considerations were taken into account, a final outline was developed. The main topics are highlighted in figure 4.

The second phase consisted of assembling the content for each presentation. The slide decks were developed with the intent that anyone could serve as the presenter with minimal material familiarity. With this concept in mind, a detailed script was written in the notes section of each individual slide. This script was designed to provide the necessary background for the speaker while simultaneously allowing the presenter to provide additional insight as needed. Another benefit to developing a script is that the speaker will be encouraged to remain on point for each slide. To maintain audience interest, two to three bullet points were typically provided on each slide, and additional graphs and images were used to illustrate or highlight particular concepts.

The third phase consisted of revising and editing the presentations. The primary objective during the revision process was to make certain that the technical content met the needs of the intended audience. The slides were reviewed so that the information provided was described effectively and succinctly, the content on each slide made sense, the graphics on each slide were suitable, and the key messages and takeaways for each presentation were retained. In the following section, additional details are provided related to the content of each presentation.

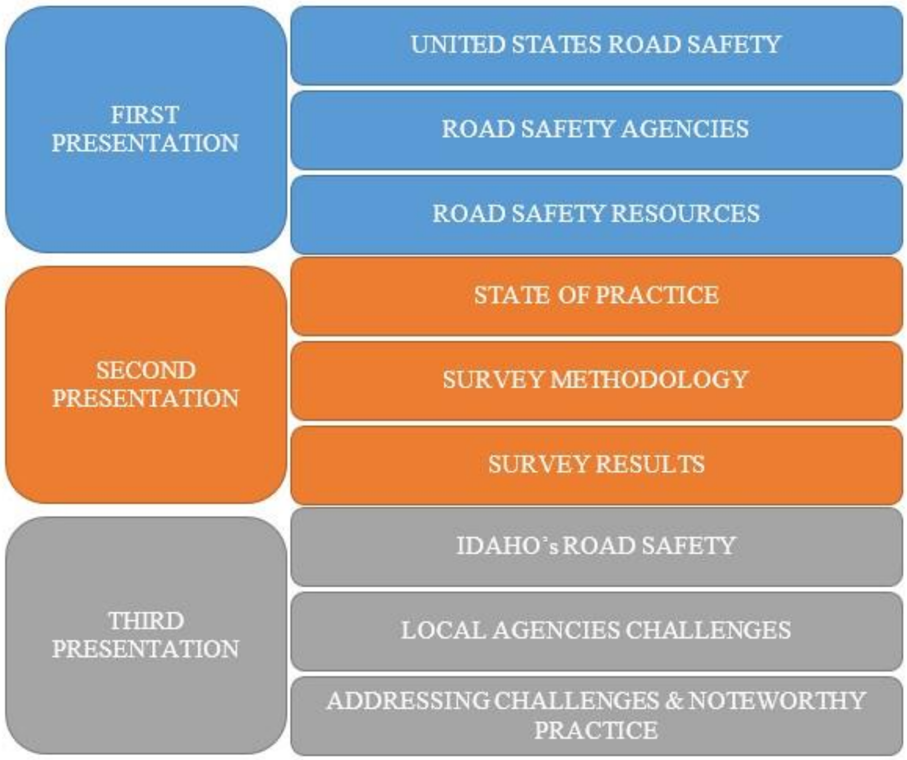

Figure 5-4: Presentation organization.

The slide deck for the first presentation introduces the training by starting with the history of road safety in the United States from the Interstate Act of 1956 through the 2015 FAST Act. Statistical facts of road safety around the United States are presented and include, but are not limited to, the number of fatalities and crashes, roadway ownership in the Pacific Northwest states, and the economic impacts of crashes. The next section describes general concepts of road safety such as defining road safety, describing the road users, and performance measures. Current safety legislation and its importance are then discussed. A significant part of this presentation describes the resources available to local agencies, beginning with the governmental agencies that focus on improving road safety and concluding with a discussion on specific manuals, courses, and available software.

The second presentation highlights ongoing research efforts, along with the methodology and analysis or the local practitioner survey and the responses collected. A description of the purpose, objectives, and methodology used for the research is provided. Each step of the methodology is listed and described, and key elements, such as preliminary assessment, past studies, and data collection, are provided. The target population and survey objectives are described, and each survey question is explained. The survey responses are presented and include information on the response rates and the geographic location of each responder. Specific survey results are discussed, with an explanation identifying the resources that local agencies in Idaho have available and the challenges that agencies face while addressing road safety.

The third and final presentation focuses on the local agencies in the state of Idaho. It identifies the challenges they face and provides recommendations to address these challenges. State-specific road safety statistics from 2010 to 2015 are introduced, along with details as to how the transportation system is organized in the state. The challenges faced with gathering road safety data from local agencies are explained and include causes, consequences, and the importance of addressing them. A discussion follows encouraging the use of the Strategic Highway Safety Plan (SHSP) for local agencies. The advantages and implementation methods are explained, and additional resources are provided so that the practitioner can obtain more information. In the final section of the presentation, noteworthy practices throughout the United States are highlighted to showcase how some states are addressing their challenges. The presentation ends with a short conclusion section that presents the key points of the presentation.

For each presentation, a short questionnaire was developed for the audience to identify the key points of each slide deck and to rate the effectiveness of various aspects of the presentation. The slide deck is intended to serve as a general framework; it is a living document and changes and updates are expected. The feedback provided by users will be reviewed and addressed appropriately during the update process.

5.8 Conclusions

Safety planning efforts and funding often start at the state level and trickle down to local and regional agencies. Analysis of and investment in state highways rather than local roads is often the result of several factors, including statewide priorities on higher volume and higher speed roadways, which typically represent the roadway network of the state DOT itself, data limitations on local roads, the lack of safety planning expertise at local agencies, the lack of a champion in state DOTs for expanding opportunities for local agencies, and other requirements that discourage or restrict participation by local agencies in a highway safety program. In summary, this research sought to examine both the challenges and current resources available to local agencies—and specific to the state of Idaho—to address road safety.

Responses from a local agency survey identified that local agencies in Idaho were familiar with the HSIP and relied on staff from the state DOT to collect traffic data. The most common type of safety analyses used to evaluate sites for safety improvements were the analysis process for stop signs, followed by traffic control signal warrants. The results showed that the software most commonly used to analyze traffic safety data was the GIS and the most common manual used was the MUTCD. In the past three years, most of the agencies received training in pedestrian and bicycle safety or complete streets. Project funding, project prioritization, and data analysis procedures were the three highest needs identified by local agencies in the state of Idaho. If funding was not an issue, other needs that agencies focused on included hiring additional staff or expertise and increasing training or technical assistance opportunities.

All the data collected were also used to determine what challenges local agencies face to address road safety. Some agencies lack the funds needed to implement safety improvements on their road systems and look to the state HSIP or other funding sources for assistance. The increase of competition and the need to do more with less makes choosing the right projects more important than ever. Local agencies often lack the safety data or analytical skills necessary to meet crash data analysis requirements. To stretch limited highway safety funding, local transportation agencies are encouraged to identify and implement the optimal combination of countermeasures to achieve the greatest benefits. However, the lack of local staff requires more multitasking and results in increased inefficiencies. One common challenge related to training was a lack of and uneven commitment to training.

The three presentations that were created will serve the purpose of addressing some of these challenges and providing alternatives to assist local agencies. The primary challenge to address remains the ability of a local agency, often with limited staff or resources, to be properly engaged and informed in order to best address existing roadway safety needs.

6. Assessment and Training (Academic Perspective)

In spite of significantly reducing the number of collision-involved injuries and fatalities over the past decades, the United States still has unacceptably high collision rates in comparison to other developed countries (Dixon 2016). In addition, there are many new challenges such as safety issues related to texting and driving (Issar et al. 2013). As an ever-evolving field, road safety requires continual attention to new research.

To address ongoing and new road safety challenges, preparing students with expertise in road safety research is one important mission that cannot be underestimated by university educators. Yet most civil engineering departments at universities in the United States do not offer an independent road safety course. Instead, students' road safety knowledge is built upon scattered sessions provided in other transportation courses such as transportation engineering, planning, and freight and supply chains. It is also important to note that road safety is interdisciplinary in nature; it intersects civil engineering, psychology, mechanical engineering, urban planning, public health, and other disciplines.

6.1 Introduction

There is a great need for a road safety course as part of the university curriculum. Providing systematic training on road safety is the cornerstone for preparing students to become future safety experts. Beyond the long-term goal, on an individual level, it is important to inform students, most of whom are at the prime age for exhibiting risky driving behaviors, to become more aware of the risks for collision and injury. This chapter presents a path for developing a road safety educational module by identifying learning objectives and reading materials, suggesting teaching methods, and evaluating learning outcomes.

6.2 Learning Objectives

The target audience for this educational module is university students. It is also designed for transportation faculty who need to teach a course focused on or related to road safety. The first task is to identify the essential content for a road safety class. Bloom's Taxonomy, the well-known educational philosophy, is useful for identifying learning objectives. Bloom's Taxonomy offers a framework for characterizing learning outcomes within six levels: knowledge, comprehension, application, analysis, synthesis, and evaluation/creation (Tavakol and Dennick 2011). To ensure that the learning objectives match the multi-level educational purposes, several criteria are established, including

- Cover primary definitions, theories, and measurements of road safety.
- Create a path from easy knowledge to challenging content.
- Plan the class from a global perspective and offer localized geo-spatial scales.
- Apply methods to analyze realistic road safety problems.
- Tutor both qualitative and quantitative skills.
- Integrate knowledge from multiple disciplines.
- Oversee future challenges in road safety.
- Ensure the use of diverse demonstration tools beyond traditional classroom instruction (e.g., lectures, videos, presentations) to foster varied forms of interactions between educators and students.

Research-based road safety education should afford a wide coverage of the essential elements in the field, including definitions, measurements, analytical methods, theories, data, real-world examples, and new challenges—all aimed at helping students shape a comprehensive framework with a set of problem-solving tools. In sum, learning objectives are defined according to the learning outcomes of gained knowledge, skills, and interests. In conformity with the above-mentioned selection criteria, three lectures are proposed:

- Road safety: basics and issues;
- Road safety: history, measures, and analytical methods; and
- Road safety: new challenges.

Within each lecture, multiple learning objectives are accordingly defined. In the first lecture, "Road safety: basics and issues," the learning objectives include the following:

- Define road safety terms such as injury, exposure, risk factor, protective factor, crash frequency, and crash rate.
- Calculate the quantifiable measurements and interpret their meaning.
- Distinguish different sets of road safety measures.
- Describe road safety issues at different geo-spatial scales, such as developed countries versus developing countries, the United States, Washington state, and the city of Seattle; and compare the consistency and controversy across different locales.
- Analyze road safety status and identify the most significant contributors.
- Summarize road safety trends by age group, vehicle type, and human error.

In the second lecture, “Road safety: history, measures, and analytical methods,” the learning objectives include the following:

- Recall the history of road safety research in the last century.
- Apply road safety research methods, including the 5E’s, the public health approach, Haddon’s Matrix, and the system approach, to analyze safety outcomes.
- Describe the road safety research focus in the past, present, and future.
- Classify road safety measures and evaluate the effects of different sets of road safety measures such as speed management and traffic control.
- Interpret safety outcomes with related safety theories.
- Assess the costs and benefits of road safety financial budgets.

For the third lecture, “Road safety: new challenges,” the learning objectives include the following:

- Describe the working mechanism of self-driving cars and, given their increasing popularity, assess the possible road safety risks.
- Discuss the pros and cons of promoting the use of self-driving cars.
- Identify risk factors related to visibility issues and road safety.
- Propose solutions to improve road visibility and assess the strategies’ reliability.

- Summarize the risk between texting and driving; and conceptualize a framework to evaluate the effectiveness of observed strategies in preventing texting and driving.

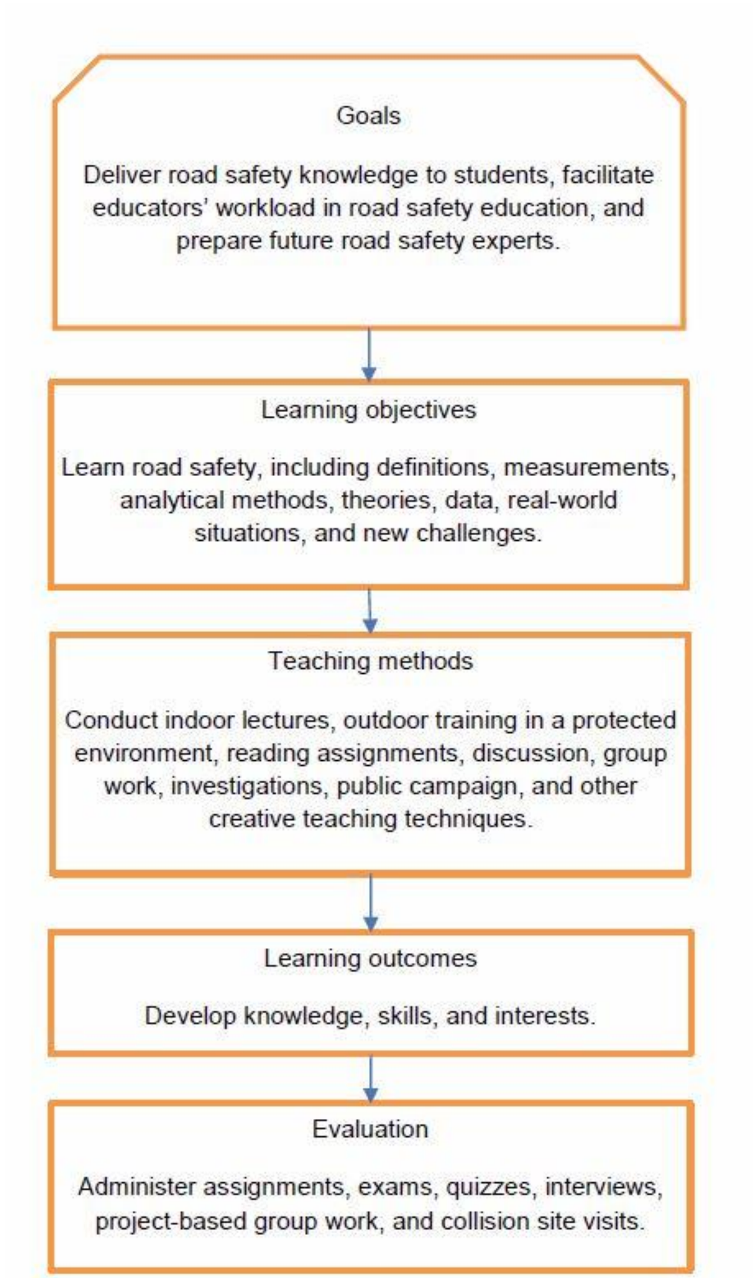

Figure 6-1: Framework for designing educational modules

6.3 Reading Materials

Reading materials should both incite in-depth thinking and provide a thorough coverage of subject knowledge. The objective of assigned readings is to encourage students to think logically, deeply, creatively, comprehensively, and independently. Reading material that is too challenging risks frustrating students. However, readings that are too simple risk wasting students' time if new knowledge or insights cannot be gained. With these guiding principles in mind, the following points are proposed as selection criteria for reading material:

- Present multi-level road safety reports to deliver real-world examples, from the international level to the localized state and city levels.
- Require students to read classical selections to establish a good sense of knowledge in the field.
- Assign textbook readings to standardize definitions, measurements, and simple analytical methods.
- Expose students to new challenges through videos, news, events, ongoing projects, and cutting-edge research articles.
- Select articles that best represent their field.

6.4 Course Content

Though there is no agreement upon the course content for a road safety class, there is consensus about the increasing importance of road safety worldwide. To decide course content, it is necessary to clarify the nature of road safety knowledge.

First, over a long period, the success of road safety improvement is mostly based on the reduction of crash frequency, risk, and injury severity. These conventional measures help students to establish the terms and to evaluate the performance of safety improvements. However, such conventional measures depend on long-term data collection. Additionally, with regard to crash data, missing values are common, and under-reporting is widely observed, especially for property damage and accidents with slight injuries. Much content within the field needs further validation through improved data collection.

Second, many factors underlie a road crash such as roadway network, external circumstances, built environments, and human factors. Most underlying factors, such as weather, are either hard to control or are completely uncontrollable. Therefore, many educational efforts may not directly serve to solve real-world problems. Given the uncertainties and limitations of knowledge in the field, the relatively solid conventional knowledge is given priority in determining course content.

Third, a safety-focused class helps correct misunderstandings and incorrect knowledge. Road safety issues are closely attached to everyone's daily life. Many students have already developed some preconceptions about safety, some of which may be incorrect. This class provides a perfect opportunity to correct misconceptions and poor understanding. Yet, if this class were completely designed to be analysis-based, involving a significant amount of modeling and simulation, it would risk exceeding the capability of undergraduate students in comprehending new knowledge. Choosing content from easy pieces and gradually promoting in-depth thinking are crucial for designing a successful educational module.

6.5 Teaching Methods

This educational module adopts conventional teaching in that the majority of its content is delivered through indoor lectures, but this conventional approach is enhanced with discussion, debates, group projects, and guest lectures by scholars from different disciplines as well as practitioners. Within the format of the indoor lectures, various activities are also designed to promote active learning.

This module also suggests using innovative teaching methods to improve students' learning performance such as experiencing a road crash simulator (if the university has such access), visiting representative crash sites and proposing improvement strategies, observing traffic-violating behaviors in the real world (e.g., pedestrians improperly crossing streets, vehicles failing to yield, or bicyclists riding on arterial routes), and organizing public campaigns to gain insight into public awareness regarding road safety. In general, after students establish notions and understand the frameworks of safety research, diverse teaching methods create different ways to approach road safety issues.

6.6 Learning Outcomes

Whether the selected learning objectives and suggested teaching methods contribute to the success of this teaching module needs evaluation. On this point, one difficulty is that there is no general agreement about how to evaluate the effectiveness of various road safety class learning objectives. However, one bonafide criterion for evaluating this teaching module's quality is assessing the amount of knowledge that students gain. Factors that encourage knowledge gain include being consistent in the design of learning objectives, working to make the assessment of student performance multidimensional, and ensuring that the module results in skills learned and interests engaged.

Student learning outcome assessments can be formal or informal, high- or low-level, and individual or collective. Formal evaluation would be exams, assignments, group projects, and quizzes. Informal evaluation largely depends on in-class participation and real-world problem-solving skills. High- or low-level performance is based on task ease. The conventional approach to assess student learning, in terms of knowledge and skills, is by using the following four categories:

- Initial status and endpoint of student learning: it is necessary and relatively easy to gather evidence of knowledge and skills from two stages, the beginning and the end.
- Student learning after a period of instruction: it is important to determine whether students have obtained knowledge along the planned track—that is, not just at the end but during the process of instruction. The educational module design should be continually adapted to the needs of students based on their feedback, which is also a way of improving the cost effectiveness of teaching. Unfortunately, there is no agreement on the best time intervals for measuring learning; however, it would be beneficial to continually interact with students to assess their growth in knowledge.
- Costs of learning: cost-ineffective teaching can mean educators devoting too much time to delivering a set amount of knowledge or students spending too much time learning a set amount of skills. In such circumstances, educators should search for alternative ways to deliver knowledge and to train skills.
- Feedback: educators should be open to all types of reactions, opinions, and views from students. Feedback from diverse perspectives helps improve the module design.

To evaluate learning outcomes, this module adopts the conventional method of assessment, including assignments, debates, quizzes, and group projects. These measures would be enough to assess student learning. Additionally, evaluating the quality of this teaching module depends on evaluating the above-mentioned four criteria.

REFERENCES

Chapter 1

Driving Safety. National Highway Traffic Safety Administration, 2014. Web. 20 Feb. 2015.

Chapter 3

AASHTO. (2011). A Policy on Geometric Design of Highway and Streets, 6th Ed.

Beanland, V., Fitzharris, M., Young, K. L., and Lenné, M. G. (2013). “Driver inattention and driver distraction in serious casualty crashes: Data from the Australian National Crash In-depth Study.” *Accident Analysis & Prevention,*54, 99-107.

Bertini, R., Monsere, C., Byrd, A., Rose, M., and El-Seoud, T. (2005). “Using custom transportation data collection software with handheld computers for education, research, and practice.” *Transportation Research Record: Journal of the Transportation Research Board*, (1924), 37-45.

Birnbaum, M. H. (2004). “Human research and data collection via the Internet.” *Annual Review of Psychology*, 55, 803-832.

Carrel, A., Lau, P. S., Mishalani, R. G., Sengupta, R., and Walker, J. L. (2015). “Quantifying transit travel experiences from the users’ perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses.” *Transportation Research Part C: Emerging Technologies*, 58, 224-239.

Chen, H., Chiang, R. H., and Storey, V. C. (2012). “Business Intelligence and Analytics: From Big Data to Big Impact.” *MIS Quarterly*, 36(4), 1165-1188.

Crash Modification Factors Clearinghouse. (n.d.). <http://www.cmfclearinghouse.org/> (Sep. 05, 2016).

Dolgov, D., Thrun, S., Montemerlo, M., and Diebel, J. (2010). “Path planning for autonomous vehicles in unknown semi-structured environments.” *The International Journal of Robotics Research*, 29(5), 485-501.

Driscoll, D. L. (2011). “Introduction to primary research: Observations, surveys, and interviews.” *Writing Spaces: Readings on Writing*, 2, 153-174.

Federal Highway Administration. Manual on Uniform Traffic Control Devices. (n.d.). <http://mutcd.fhwa.dot.gov/> (Sep. 05, 2016).

Felder, R. M., and Brent, R. (2003). “Designing and teaching courses to satisfy the ABET engineering criteria.” *Journal of Engineering Education*, 92(1), 7-25.

Frazzoli, E., Dahleh, M. A., and Feron, E. (2002). “Real-time motion planning for agile autonomous vehicles.” *Journal of Guidance, Control, and Dynamics*, 25(1), 116-129.

Gross, F., and Jovanis, P. P. (2008). “Current state of highway safety education: Safety course offerings in engineering and public health.” *Journal of Professional Issues in Engineering Education and Practice*, 134(1), 4958.

Hauer, E., Harwood, D., Council, F., and Griffith, M. (2002). “Estimating Safety by the Empirical Bayes Method: A Tutorial.” *Transportation Research Record: Journal of the Transportation Research Board 1784*.

Hood, M. M., Thompson, S. R., Vance, R. J., Renz, M. S., Harder, B. T., Toole, J., and Hunter, S. T. (2014). *Guide to Accelerating New Technology Adoption through Directed Technology Transfer* (No. Project 20-93).

Hsieh, H. F., and Shannon, S. E. (2005). “Three approaches to qualitative content analysis.” *Qualitative Health Research*, 15(9), 1277-1288.

Huang, A., and Levinson, D. (2015). “Axis of travel: Modeling non-work destination choice with GPS data.” *Transportation Research Part C: Emerging Technologies*, 58, 208-223.

Katal, A., Wazid, M., and Goudar, R. H. (2013, August). “Big data: issues, challenges, tools and good practices.” *Contemporary Computing (IC3)*, IEEE, 2013 Sixth International Conference, 404-409.

Khattak, A. J., Hummer, J. E., and Karimi, H. A. (2000). “New and existing roadway inventory data acquisition methods.” *Journal of Transportation and Statistics*, 3(3), 33-46.

Kieu, L. M., Bhaskar, A., and Chung, E. (2015). “A modified density-based scanning algorithm with noise for spatial travel pattern analysis from Smart Card AFC data.” *Transportation Research Part C: Emerging Technologies*, 58, 193-207.

Lu, N., Cheng, N., Zhang, N., Shen, X., and Mark, J. W. (2014). “Connected vehicles: Solutions and challenges.” *IEEE Internet of Things Journal*, 1(4), 289-299.

Lv, Y., Duan, Y., Kang, W., Li, Z., and Wang, F. Y. (2015). “Traffic flow prediction with big data: a deep learning approach.” *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 865-873.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., and Barton, D. (2012). “Big data: the management revolution.” *Harvard Business Review*, 90(10), 61-67.

Merit, B. (n.d.). *Code of Ethics*. <http://www.asce.org/code-of-ethics/> (Sep. 05, 2016).

Montfort, D. B., and Brown, S. (2013). “What do we mean by cyberlearning: Characterizing a socially constructed definition with experts and practitioners.” *Journal of Science Education and Technology*, 22(1), 90-102.

Moore, J. E., Giuliano, G., and Cho, S. (2004). “Secondary accident rates on Los Angeles freeways.” *Journal of Transportation Engineering*, 130(3), 280-285.

Murdoch, T. B., and Detsky, A. S. (2013). “The inevitable application of big data to health care.” *JAMA*, 309(13), 1351-1352.

Potter, J., and Hepburn, A. (2005). “Qualitative interviews in psychology: Problems and possibilities.” *Qualitative Research in Psychology*, 2(4), 281-307.

Rapley, T. J. (2001). “The art(fulness) of open-ended interviewing: some considerations on analysing interviews.” *Qualitative Research*, 1(3), 303-323.

Rothe, J. P. (1991). Educating for safety. *Society*, 28(3), 5-9.

Schneeweiss, S. (2014). “Learning from big health care data.” *New England Journal of Medicine*, 370(23), 21612163.

Shi, Q., and Abdel-Aty, M. (2015). “Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways.” *Transportation Research Part C: Emerging Technologies*, 58, 380-394.

St-Aubin, P., Saunier, N., and Miranda-Moreno, L. (2015). “Large-scale automated proactive road safety analysis using video data.” *Transportation Research Part C: Emerging Technologies*, 58, 363-379.

Thomas, D. R. (2006). “A general inductive approach for analyzing qualitative evaluation data.” *American Journal of Evaluation*, 27(2), 237-246.

Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., and González, M. C. (2015). “The path most traveled: Travel demand estimation using big data resources.” *Transportation Research Part C: Emerging Technologies*, 58, 162-177.

Turochy, R., Fricker, J., Hawkins, H., Hurwitz, D., Ivey, S., Knodler, M., and Young, R. (2013). “Assessment of introductory transportation engineering course and general transportation engineering curriculum.” *Transportation Research Record: Journal of the Transportation Research Board*, (2328), 9-15.

Turochy, R. E. (2012). “Structuring the content of the first course in transportation engineering: Perspectives of engineers and educators.” *Journal of Professional Issues in Engineering Education and Practice*, 139(3), 206-210.

Vera-Baquero, A., Colomo Palacios, R., Stantchev, V., and Molloy, O. (2015). “Leveraging big-data for business process analytics.” *The Learning Organization*, 22(4), 215-228.

Vlahogianni, E. I., Park, B. B., and van Lint, J. W. C. (2015). “Big data in transportation and traffic engineering.” *Transportation Research Part C: Emerging Technologies*, 58, 161.

Wang, G., Gunasekaran, A., Ngai, E. W., and Papadopoulos, T. (2016). “Big data analytics in logistics and supply chain management: Certain investigations for research and applications.” *International Journal of Production Economics*, 176, 98-110.

Chapter 4

Abdel-Aty, M. A, and Radwan, A. E. (2000). “Modeling Traffic Accident Occurrence and Involvement.” *Accident Analysis & Prevention*, 32 (5), 633–42.

Bham, G. H., Javvadi, B. S., and Manepalli, U. R. R. (2011). “Multinomial Logistic Regression Model for SingleVehicle and Multivehicle Collisions on Urban US Highways in Arkansas.” *Journal of Transportation Engineering*, 138 (6), 786–797.

Cafiso, S., D’Agostino, C., and Persaud, B. (2013). “Investigating the Influence of Segmentation in Estimating Safety Performance Functions for Roadway Sections.” *Proceedings, 92nd Annual Meeting of the Transportation Research Board*, TRB, Washington, DC.

Cafiso, S., Graziano, A. D., Silvestro, G. D., and La Cava, G. (2008). “Safety Performance Indicators for Local Rural Roads: Comprehensive Procedure from Low-Cost Data Survey to Accident Prediction Model.” *Proceedings, Transportation Research Board 87th Annual Meeting*.

Cenek, P. D., Davies, R. B., McLarin, M. W., Griffith-Jones, G., and Locke, N. J. (1997). “Road Environment and Traffic Crashes.” *Transfund New Zealand Research Report*, (79).

Elvik, R. (2001). “Cost-Benefit Analysis of Road Safety Measures: Applicability and Controversies.” *Accident Analysis & Prevention*, 33(1), 9–17.

Griffith, M. (1999). “Safety Evaluation of Rolled-in Continuous Shoulder Rumble Strips Installed on Freeways.” *Transportation Research Record*, (1665), 28–34.

Hajbabaie, A., and Wang, Y. (2016). “Regional Map Based Analytical Platform for State-Wide Highway Safety Performance Assessment.”

Koorey, G. (2009). “Road Data Aggregation and Sectioning Considerations for Crash Analysis.” *Transportation Research Record*, 2103, 61–68.

Kumara, S, and Chin, H. (2005). “Application of Poisson Underreporting Model to Examine Crash Frequencies at Signalized Three-Legged Intersections.” *Transportation Research Record*, (1908), 46–50.

Lord, D., and Mannering, F. (2010). “The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives.” *Transportation Research Part A: Policy and Practice,* 44 (5), 291–305.

Ma, P. E. (2009). “Bayesian Analysis of Underreporting Poisson Regression Model with an Application to Traffic Crashes on Two-Lane Highways.” *Proceedings, Transportation Research Board 88th Annual Meeting*.

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*, 37. CRC press.

Miaou, S.-P., and Lum, H. (1993). “Modeling Vehicle Accidents and Highway Geometric Design Relationships.” *Accident Analysis & Prevention*, 25 (6), 689–709.

Mountain, L., Maher, M. and Fawaz, B. (1998). “The Influence of Trend on Estimates of Accidents at Junctions.” *Accident Analysis & Prevention,* 30 (5), 641–49.

Ogle, J. H., Alluri, P. and Sarasua, W. (2011). “MMUCC and MIRE: The Role of Segmentation in Safety Analysis.” *Proceedings, 90th Annual Meeting of the Transportation Research Board*.

Persaud, B. (2001). *Statistical Methods in Highway Safety Analysis*. National Academy Press, Washington, DC.

Renski, H., Khattak, A., and Council, F. (1999). “Effect of Speed Limit Increases on Crash Injury Severity: Analysis of Single-Vehicle Crashes on North Carolina Interstate Highways.” *Transportation Research Record*, (1665), 100– 108.

Sittikariya, S., Shankar, V., and Venkataraman, N. (2009). *Modeling Heterogeneity: Traffic Accidents*.

Ulfarsson, G., and Shankar, V. (2003). "Accident Count Model Based on Multiyear Cross-Sectional Roadway Data with Serial Correlation." *Transportation Research Record*, (1840), 193–97.

Washington, S. P., Karlaftis, M. G., and Mannering, F. (2010). *Statistical and Econometric Methods for Transportation Data Analysis*. CRC Press.

Chapter 5

American Association of State Highway and Transportation Officials (AASHTO). (2010). *Highway Safety Manual*, First Edition, Washington, D.C.

AASHTO. (2011). A Policy on Geometric Design of Highway and Streets, 6th Edition, Atlanta, GA.

Baruch, Y. and Holtom, B. C. (2008). "Survey response rate levels and trends in organizational research." *Human Relations*, 61(8), 1139-1160.

Cook, C., Heath, F., Thompson, R., (2000). A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys, *Educational and Psychological Measurement*, 60(6), 821-836

Federal Highway Administration. (2009). *Manual on Uniform Traffic Control Devices*. 7th Edition, Washington, D.C.

Gaines, D., Waldheim, N. and Herbel, S. (2013) "Assessment of Local Road Safety Funding, Training, and Technical Assistance." Federal Highway Administration Office of Safety, Washington, DC.

Herbel, S., Laing, L., McGovern, C. (2010). *Highway Safety Improvement Program (HSIP) Manual*. First Edition, Washington, DC.

Lefler, N., Council, F., Harkey, D., Carter, D. McGee, H., and Daul, M. (2010). *Model Inventory of Roadway Elements* (MIRE). FHWA, Washington, DC.

National Highway Traffic Safety Administration. (2015). *Traffic Safety Facts 2013 Data*, <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812169> (July 2015)

Park, S., McTish, P., Holman, J., Giancola, A. R., and Davenport, J. S. G. (2016). *National Cooperative Highway Research Program (NCHRP)* Synthesis 486: State Practices for Local Road Safety. Transportation Research Board, Washington, DC.

Preston, H., and Storm, R. (2014). *NCHRP Project 20-24, Task 87: State DOT Administration of Local Road Safety Aid*. Transportation Research Board, Washington, DC.

Richardson, A. Ampt, E. and Meyburg, A. (1995). "Survey Methods for Transport Planning" *Melbourne: Eucalyptus Press*. 3-4.

Sheehan, K. B. (2001). "E-mail Survey Response Rates: A Review" *Journal of Computer-Mediated Communication*, 6(2).

Solomon, David J (2001) Conducting Web-Based Surveys. *Practical Assessment, Research & Evaluation*. 7(19) 1-4

U.S. Department of Transportation Bureau of Transportation Statistics. (2014). *State Transportation Statistics 2014*, 12th Edition, New Jersey, SE

Watt, J. H. (1999) “Internet systems for evaluation research.” *New Directions for Evaluation*, 1999(84), 23-43.

Wilson, E. M. (2003). *NCHRP Synthesis 321: Roadway safety tools for local agencies*. Transportation Research Board, Washington, DC.

Chapter 6

Dixon, A. (2016). “US road safety improves, but still lags behind many other countries.” *The Nation's Health,* 46(7), 8.

Issar, N. M., Kadakia, R. J., Tsahakis, J. M., Yoneda, Z. T., Sethi, M. K., Mir, H. R., Archer, K., Obremskey, W. T., and Jahangir, A. A. (2013). “The link between texting and motor vehicle collision frequency in the orthopaedic trauma population.” *Journal of Injury and Violence Research,* 5(2), 95-100.

Tavakol, M., and Dennick, R. (2011). “Post-examination analysis of objective tests.” *Medical Teacher,* 33(6), 447-458.

APPENDICES

A. [intentionally blank]

B. Acronyms, Abbreviations, and Brand Names

C. Questionnaire for Interview

NOTE: APPENDICES D THROUGH J ARE IN A SEPARATE PDF FILE AVAILABLE FROM HTTP://WWW.RAPERKINS.NET/PRESENTATIONS/APPENDICIES%20D%20TO%20J.PDF

D. General

    1. An Introduction to the Highway Safety Manual

    2. SPF

    3. CMF

    4. CMF Clearing House Brochure

    5. NCHRP Crash Data Snapshot

E. Alaska

    1. Form 209, operators report of accident

    2. Form 200, police report of accident

    3. Crash Data Flow, old and new

    4. CARE Dashboard

    5. HSIP Flowchart

F. Washington

    1. Data Office

        a. Transportation Data & GIS Office Brochure

        b. Washington State Crash Analysis Flow

        c. WSDOT Collision Data Systems Overview

    2. State Roads

        a. Safety Flow Chart – state roads

    3. Local Programs

a. Local Guide (cover and index only)

G. Oregon

1. PowerPoint of Safety Program with explanation of SPIS.

H. Idaho

1. IMPACT press release

I. Colorado

1. Example of Before/After Analysis

J. Arizona

1. Flow chart of local program HSIP approval

K. Content Table of the Three Learning Objectives

APPENDIX B - Acronyms, Abbreviations, Brand Names

A

AADT, Annual Average Daily Traffic

ALISS, Accident Location Identification and Surveillance System

B

B/C, Benefit/Cost analysis or ratio

C

CAC, Collision Analysis Corridor, Washington

CAL, Collision Analysis Location, Washington

CARE, Critical Analysis Reporting Environment, Alaska (and other states, from University of Alabama) <http://care.cs.ua.edu/care.aspx>

CDR, Crash Data Repository, Alaska Term

CIRCA, Crash Information Retrieval Collection and Analysis, Idaho

CLAS, Crash Location and Analysis System, Washington

CMF, Crash Modification Factor, typically supplied by the HSM

CPDM, Capital Program Development and Management Office, Washington

CRF

D

DiExSys, see Vision Zero Suite

DMV, Department of motor vehicles

DPS , Department of Public Safety, Arizona

E

EB Empirical Bayes, statistical method, see "Estimating Safety by the Empirical Bayes Method: A Tutorial." Ezra Hauer, E., Harwood, D., Council, F., and Griffith, M. (2001). *Transportation Research Record: Journal of the Transportation Research Board,* 1784.

ETL, Extract, Translate, Load program

F

FARS, Fatal Accident Reporting System, <http://www.nhtsa.gov/FARS>

FHWA, Federal Highway Administration

G

GIS, Geographic Information System generic

GPS, Global Positioning System <http://www.gps.gov/>

H

HAS, Highway Analysis System, Alaska

HQ, Headquarters, generic term

HSEC, Highway Safety Executive Committee, Washington

HSIG, Highway Safety Issues Group, Washington

HRRRP, Highway Risk Rural Road Program

HSIP, Highway Safety Improvement Program <http://safety.fhwa.dot.gov/hsip/>

HSM, Highway Safety Manual, <http://www.highwaysafetymanual.org/Pages/default.aspx>

I

IAL, Intersection Analysis Locator, Washington

IHSDM, Interactive Highway Safety Design Model <www.ihsdm.org>

IMPACT, Idaho Mobile Program for Accident CollecTion, Idaho

L

LRS, Linear Referencing System

LTAP, Local Technical Assistance Program

M

MMUCC, Model Minimum Uniform Crash Criteria

MPO, Municipal Planning Organizations

N

NHTSB, National Highway Transportation Safety Board <http://www.nhtsa.gov/>

O

O&M, Operations and Maintenance generic term

P

PDA, Project Development Authorization pdf, Portable Document Format, an Adobe program

PDO, Property Damage Only

PM, project management

Q

QC, Quality Control, generic

R

RFP, Request for Proposals

ROM, Rough Order of Magnitude (generic estimating term)

ROW, Right of Way

S

SA, SafetyAnalyst, Washington (and other states)

SDM, Safety Data Mart, Arizona

SECTOR, Statewide Electronic Collision and Ticketing Online Records, Washington

SHSP, Strategic Highway Safety Plan generic <http://safety.fhwa.dot.gov/hsip/shsp/>

SPF, Safety performance function

SPIS, Safety Priority Index System, Oregon

SQL, IBM software, Structured Query Language

STIP, State Transportation Improvement Program

T

T&S, Traffic and Safety generic term

TEMA, Traffic Enforcement Mobilization Agreement

TraCS, Traffic and Criminal Software, Traffic Records Program in Alaska Oracle. <http://www.dot.state.ak.us/highwaysafety/trafficrecords_tracs.shtml>

TREDIS, Estimating Program <http://www.tredis.com/products/product-overview>

TTAP, Tribal Technical Assistance Program

Toughbook, Data Entry Computer, Alaska (and other states)

TZD, Toward Zero Deaths

V

VIN, Vehicle Identification Number

Vision Zero Suite, ELT program developed by DiExSys for Colorado DOT

VPSO, Village Public Safety Officer, Alaska

W

WebCARS, Online Crash Analysis and Retrieval System, Idaho

APPENDIX C - Questionnaire for Interview

This questionnaire is designed as an outline for an in-person interview, either face-to-face or by telephone. It features leading questions followed by notes and suggestions to guide the interviewer. Often, only the leading questions needs be asked and the notes given here are only a checklist for the interviewer.

1. Usual formalities and introductions
    a. Expected time of interview
2. Description of research project
    a. PacTrans
    b. Outline of project

The Highway Safety Improvement Program (HSIP), which provides federal funding for state and local highway projects, requires a data-driven, strategic approach to improving highway safety on all public roads …. (ref: http://safety.fhwa.dot.gov/hsip/). Our research focuses on data. We review: 1. How Highway safety data are gathered, and 2. How data are used in project selection for highway projects. Our research was based on interviews with traffic and engineering professionals in PacTrans and other states. We hope this research will inform discussion of improvements in transportation academic and practical transportation engineering education.

3. We'd like your help with our research project: Safety Data Management and Analysis: Addressing the Continuing Education Needs for the Pacific Northwest. The intent of this survey/interview is to examine the flow of safety data within the DOT, especially the details of the process including: agency report development, safety data software applications, use of GIS, and other techniques to improve the quality of safety data – all with eye towards educational needs identified. Describe your job as it relates to safety data
    a. Note, is this the "data office" or the safety office? Is there a difference?
4. Next we get into some details of the data from crash>responder>report>QC>storage
    a. Programs used

b. Problems with data

c. Location

d. Incident

i. What types of incident are reported?

ii. Dollar limit, Citations, other

e. Who are responders?

i. Troopers, Municipal police, VPSO, citizens, others,

ii. Jurisdiction?

f. Do all do it the same way?

g. Forms, electronic, paper

h. Self-report (different form)

i. Where does report go and how?

i. First to DMV? ii. Direct to DOT?

iii. Report to Municipality?

j. Is it raw? Does DMV process/filter/edit?

i. How long does this take?

ii. Is there a single data base?

1. DMV “collects crash reports” not “collect data”

k. Who does QC on the data, especially location?

l. Storage of data

i. Enter into DOT Program directly?

ii. ii. ETL, extract translate load program

iii. From DMV to DOT for filter/edit/process

m. Who requests the data?

5. Since HSIP funding is often most important, how is that funding allocated between Statewide/HQ, Regions,

Local Governments

a. Fixed percentage

b. Competition

c. Fixed algorithm

6. [Start with state projects, local below] How are candidate projects ranked?

   a. Who does this ranking, levels?

   b. Iteration - Loop?

   c. What program extracts the data and changes it to information?

   d. Who does this?

   e. Does data remain at the primary storage site (DMV) ?

      i. And extracted by DOT as needed

      ii. Or transferred to DOT control

7. [Presumably] some preliminary ranking is made based on ROM estimates and CRF

   a. Is this so?

   b. Who does this?

   c. Programs used?

      i. B/C

      ii. CMF

      iii. ROM Estimates

   d. Is this preliminary ranking reviewed?

8. [Presumably] after the list is pared down, more detailed estimates and B/C are needed

   a. Is this so?

   b. Who does this?

   c. Programs used?

   d. Is this ranking reviewed?

9. At what point in the process are other factors besides safety B/C entered? Such as:

a. Project Management

i. Programs, staffing

b. Environmental

c. Right of Way

d. Other

e. Citizen and other input

f. O&M

10. How is final list approved?

a. When is it entered in STIP?

b. Legislative approvals?

11. How this process is different for funding transferred to local governments?

a. Is this done by regions?

12. How is the data used in HSIP Annual Reports?

a. Statewide Traffic and Safety?

b. Who does Evaluations?

Note: Appendices D through J are in a separate pdf file available from

http://www.raperkins.net/Presentations/Appendicies%20D%20to%20J.pdf

APPENDIX K - Content Table of the Three Learning Objectives

Learning objective 1: Road safety basics and issues
Background
Reading Materials
  Part A: Reports
  Part B: Textbooks
  Part C: Journal Articles
Safety terminologies
  Accident/collision/crash
  Injury
  Exposure
  Risk
  Risk factor
  Safety measure
Crash statistics (safety facts)
  Road safety problems worldwide
  Road safety problems in developing countries
    Brazil
    Russia
    China
    India
  Road safety problems in the US
  Road safety problems in Washington State
Crash characteristics (safety issues)
  Characteristics of crashes by age groups
    Children
    Young drivers
    Graduated Driver Licensing
  Crashes varied by traffic participation
    Truck-involved crashes
    Motorcycle crashes
  Characteristics of human error
    Speeding
    Drunk driving
    Drowsy driving
    Human factors
    Risk compensation

Learning objective 2: Road safety research history, analytical methods, and measures
Reading Materials
  Part A: Reports
  Part B: Textbooks
  Part C: Journal Articles
Important webpages

Road safety research history
Road safety analysis approaches
  3E's: Engineering, Education/Encouragement, and Enforcement

Extended to 5Es: 3Es plus Environment and Emergency Aid/Evaluation
VALT 2003
Public Health Approach
Haddon’s Matrix
The system approach
Safety measures
User-related measures
Vehicle-related measures
Infrastructure-related measures (road environment)
Rescue services (not CEE students’ focus)
Planning-based accident prevention
Traffic control

Learning objective 3: Road safety, new challenges
Reading Materials
Part A: Reports
Part B: Textbooks
Part C: Journal Articles
Risk of new technologies: the introduction of self-driving cars
Safe driving in reduced visibility: Influence of fog/smog on driving behavior
Driver distraction: texting