Potential Uses of Reduced Datasets from the Roadway Information Database: A White Paper

PUBLICATION NO. FHWA-HRT-20-009

DECEMBER 2019

U.S. Department of Transportation
Federal Highway Administration

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

FOREWORD

The second Strategic Highway Research Program (SHRP2) Roadway Information Database (RID) is linked to Naturalistic Driving Study data. The level of coverage and accuracy of its mobile data, combined with the inclusion of supplemental data from existing sources, make the RID a powerful stand-alone database. Using the RID requires basic geographic information system (GIS) expertise, since the roadway data elements are conflated to a GIS-based network. To make data accessible to novice GIS users, the Federal Highway Administration developed reduced datasets that can be used off the shelf with minor programming skills. Making data more accessible will increase their use and application, thereby increasing opportunities for research to influence policies and practices that ultimately reduce the number of traffic fatalities and serious injuries on our Nation's highways.

This report describes the potential highway safety research questions that can be addressed using reduced datasets. Potential research topics span seven topic areas: (1) crash predictions, (2) safety performance impacts of horizontal curve features, (3) safety performance impacts of intersection features, (4) safety performance impacts of access management, (5) risk factors for systemic safety analysis, (6) crash assignments, and (7) driver awareness of signalized intersections when entering urban areas. This report will be of interest to individuals involved in highway safety, safety training, crash and injury reduction, roadway design and policymaking, and traffic operations and management.

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

TECHNICAL REPORT DOCUMENTATION PAGE

<table>
<tr><td colspan="3">1. Report No.
FHWA-HRT-20-009</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient's Catalog No.</td></tr>
<tr><td colspan="4" rowspan="2">4. Title and Subtitle
Potential Uses of Reduced Datasets from the Roadway Information Database: A White Paper</td><td colspan="2">5. Report Date
December 2019</td></tr>
<tr><td colspan="2">6. Performing Organization Code</td></tr>
<tr><td colspan="4">7. Author(s)
R.J. Porter (ORCID: 0000-0001-8535-3451), Omar Smadi (ORCID: 0000-0002-3147-9232), Skylar Knickerbocker (ORCID: 0000-0002-0202-5872), Zachary Hans (ORCID: 0000-0003-0649-9124), and Ian Hamilton (ORCID: 0000-0003-0949-5495)</td><td colspan="2">8. Performing Organization</td></tr>
<tr><td colspan="4" rowspan="2">9. Performing Organization Name and Address
VHB
8300 Boone Blvd., Ste. 700
Vienna, VA 22182-2626
Institute for Transportation
Iowa State University
Ames, IA</td><td colspan="2">10. Work Unit No.</td></tr>
<tr><td colspan="2">11. Contract or Grant No.</td></tr>
<tr><td colspan="4" rowspan="2">12. Sponsoring Agency Name and Address
U.S. Department of Transportation
Federal Highway Administration
1200 New Jersey Avenue, SE
Washington, DC 20590</td><td colspan="2">13. Type of Report and Period
White paper; June 2018–June 2019</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code
HRDS-20</td></tr>
<tr><td colspan="6">15. Supplementary Notes.
The Task Order Contracting Officer's Representative is Yusuf Mohamedshah (HRDS-20; ORCID: 0000-0003-0105-5559), Federal Highway Administration (FHWA).</td></tr>
<tr><td colspan="6">16. Abstract
While the second Strategic Highway Research Program (SHRP2) Roadway Information Database (RID) was designed to link to SHRP2 Naturalistic Driving Study data, the level of coverage and accuracy of its mobile data, combined with the inclusion of supplemental data from existing sources, makes the RID a powerful stand-alone database as well. The RID as a stand-alone database is particularly useful for research efforts that seek to relate the safety performance of road segments and/or intersections, measured by the frequency, type, and severity of crashes, to the segment and intersection traffic, traffic control, and geometric characteristics. This white paper describes potential highway safety research questions that can be addressed using five new RID reduced datasets created by the FHWA Office of Safety Research and Development. These five new datasets are (1) average annual daily traffic, (2) intersection widths, (3) intersection crashes, (4) curve crashes, and (5) homogenous segments. Chapter 1 of the white paper provides an overview of the RID, a discussion of the value of the RID as a stand-alone database, and a definition for a RID reduced dataset. Chapter 2 provides overviews of the five new reduced datasets along with previously developed reduced datasets that are complementary to the new ones. Chapter 3 contains potential highway safety research questions that can be addressed using the five new datasets independently or in combination with previously developed reduced datasets. Potential research topics span seven topic areas: (1) crash predictions, (2) safety performance impacts of horizontal curve features, (3) safety performance impacts of intersection features, (4) safety performance impacts of access management, (5) risk factors for systemic safety analysis, (6) crash assignments, and (7) driver awareness of signalized intersections when entering urban areas (following rural travel). Chapter 4 provides a summary and conclusions.</td></tr>
<tr><td colspan="2">17. Key Words
Strategic Highway Research Program, Roadway Information Database, Naturalistic Driving Study, reduced datasets, highway safety research</td><td colspan="4">18. Distribution Statement
No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161.
http://www.ntis.gov</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td colspan="3">20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages:
37</td><td>22. Price
N/A</td></tr>
</table>

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

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

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AADT	average annual daily traffic
CCRD	curve crash reduced dataset
CMF	crash modification factor
DOT	department of transportation
FHWA	Federal Highway Administration
GIS	geographic information system
HPMS	Highway Performance Monitoring System
HSRD	homogenous segments reduced dataset
ICRD	intersection crash reduced dataset
IWRD	intersection width reduced dataset
LRS	linear referencing system
NDS	Naturalistic Driving Study
QA	quality assurance
R&D	research and development
RID	Roadway Information Database
SDF	severity distribution function
SHRP2	second Strategic Highway Research Program
SPF	safety performance function

EXECUTIVE SUMMARY

The second Strategic Highway Research Program (SHRP2) Safety Study, completed in 2015, collected an unprecedented amount of data on driver behavior and driving context.[1] The SHRP2 Naturalistic Driving Study (NDS) captured detailed data on more than 3,500 volunteer drivers, including continuous data and video of the road and driver for over 35 million vehicle miles of driving.[2] The data include more than 4,200 crashes and near-crashes. A complementary effort to create the SHRP2 Roadway Information Database (RID) collected detailed roadway characteristics for 12,500 centerline-miles, compiled existing driving-context information for another 200,000 centerline-miles, and made it possible to link the resulting RID data to NDS data. Together, the NDS and the RID are the SHRP2 Safety Data.

While the RID was designed to link to SHRP2 NDS data, the level of coverage and accuracy of its mobile data, combined with the inclusion of supplemental data from existing sources, makes the RID a powerful stand-alone database as well. The RID is particularly useful for research efforts that seek to relate the safety performance of road segments and/or intersections, measured by the frequency, type, and severity of crashes, to the segment and/or intersection traffic, traffic control, and geometric characteristics. Within these research areas, the RID provides quality data for horizontal curves, vertical grades, and intersections that have not historically been available to researchers on a regular and widespread basis. The ability to improve the reliability of research results with highly accurate data at horizontal curves and intersections could be particularly impactful given that approximately 50 percent of traffic fatalities in the United States are associated with a horizontal curve or intersection.[3,4] This white paper describes potential highway safety research questions that can be addressed using five new RID reduced datasets created by the Federal Highway Administration Office of Safety Research and Development. These five new datasets are (1) average annual daily traffic, (2) intersection widths, (3) intersection crashes, (4) curve crashes, and (5) homogenous segments. The research questions are organized under the following seven general topic areas:

1. Crash prediction.
2. Safety performance impacts of horizontal curve features.
3. Safety performance impacts of intersection features.
4. Safety performance impacts of access management.
5. Risk factors for systemic safety analysis.
6. Crash assignments.
7. Driver awareness of signalized intersections when entering urban areas (following rural travel).

Within each topic area, there is a list of specific research topics and an indication of the applicable reduced datasets to support each topic.

By integrating data within a uniform schema over a multi-State context, the RID allows researchers to develop robust, directly comparable datasets that are often not achievable through traditional open data platforms or data requests. This uniformity and consistency, both temporally and spatially, helps address some of the challenges associated with obtaining data from multiple sources.

Furthermore, the reduced datasets within the RID are highly safety-relevant resources that allow users to streamline the data collection and integration process. By providing a robust, geographically diverse dataset in a readily available format, the RID allows the research community to spend its efforts expanding the breadth and detail of individual research questions. This foundation of data provided by the RID enables researchers to iterate and test multiple approaches within the context of the dataset, rather than spend significant resources on collecting and assembling data in an ad hoc and piecemeal fashion.

CHAPTER 1. INTRODUCTION

This white paper describes potential highway safety research questions that can be addressed using five new Roadway Information Database (RID) reduced datasets created by the Federal Highway Administration (FHWA) Office of Safety Research and Development (R&D). These five new datasets are (1) average annual daily traffic (AADT), (2) intersection widths, (3) intersection crashes, (4) curve crashes, and (5) homogenous segments. The intended audience includes local, State, and Federal agencies interested in potential data sources for addressing their safety research needs; organizations that develop safety research needs and priorities; researchers at universities and private industries; and instructors and graduate students interested in highway safety research topics and data sources.

The remainder of this introductory chapter provides an overview of the RID, describes the value of the RID as a stand-alone database, and defines what is meant by a reduced dataset. Chapter 2 provides overviews of the five new reduced datasets along with previously developed reduced datasets that are complementary to the new ones. Chapter 3 discusses the white paper with potential highway safety research questions that can be addressed using the five new datasets independently or in combination with previously developed reduced datasets. Chapter 4 provides a summary and conclusions.

WHAT IS THE RID?

The second Strategic Highway Research Program (SHRP2) Safety Study, completed in 2015, collected an unprecedented amount of data on driver behavior and driving context.(1) The SHRP2 Naturalistic Driving Study (NDS) captured detailed data on more than 3,500 volunteer drivers, including continuous data and video of the road and driver for over 35 million vehicle miles of driving.(2) The data include more than 4,200 crashes and near-crashes. A complementary effort to create the SHRP2 RID collected detailed roadway characteristics for 12,500 centerline-miles, compiled existing driving context information for another 200,000 centerline-miles, and made it possible to link the resulting RID data to NDS data. Together, the NDS and the RID are the SHRP2 Safety Data.

The RID was developed to provide quality data that are linkable to the SHRP2 NDS database. The RID is stored in a secure, flexible environment and is accessible using geographic information system (GIS) tools. The RID provides the road element for safety research on more than 5 million trips taken by NDS participants. The RID enables researchers to look at road characteristics matched to NDS trips to explore relationships between driver behaviors, vehicle characteristics, and roadway environments. This capability of the RID makes it a very useful resource for NDS users interested in the influence of specific site types and roadway characteristics because it allows users to focus on only those NDS trips that traversed segments containing the items of interest. The RID has distinct advantages for R&D purposes over comparable open transportation data sources, particularly regarding the construction of multi-State databases. These advantages generally fall within three performance measures of data quality: accuracy, uniformity, and integration.(5)

Accuracy

RID data, particularly the detailed roadway features and intersection attributes desired by researchers, are collected according to a standardized methodology. Furthermore, these standards apply equally to six different State contexts. Typically, researchers would struggle to develop comparable datasets across jurisdictional lines; data collection standards and accuracy vary by State and may even vary within a State as interagency agreements may limit the ability for State transportation agencies to develop data across municipal or county boundaries. Analysis of separately aggregated multi-State databases may yield results confounded by varying data collection and data quality procedures.

Uniformity

Similar to inconsistencies in measures of data accuracy between States, State transportation agencies collect and store data in different formats according to different specifications. The Model Inventory of Roadway Elements provides a guideline for useful safety-related data elements, but it is not prescriptive as to the form these data need to take. These differences can lead to a dramatic increase in the time required to process data, if not make direct comparisons impossible. For instance, a State may elect to indicate a divided roadway by bifurcating the road centerline and storing relevant data (i.e., traffic volumes and number of lanes, etc.) in a direction of travel format. By contrast, a neighboring State may elect to store all roadway data within a single centerline and indicate median presence, traffic volumes, and lanes within a single feature's attributes. Other States may not store these data in a GIS environment at all, relying instead on a relational database; these differences in State agency data management practices may remove any ability to spatially analyze and compare data. The RID datasets significantly reduce processing time by providing data in a consistent, readily digestible format for detailed and reliable analysis.

Integration

As a result of uniform data, the RID is able to provide a highly integrated dataset that significantly reduces individual user processing time. Users can integrate various complex datasets within the RID according to a set of common spatial and tabular relationships, rather than applying research hours toward developing ad hoc spatial and attribute connections. This process is further enhanced through the dynamic segmentation tool provided with the RID dataset. This tool allows even moderately skilled GIS users to quickly query RID data within an intuitive graphical user interface.

The RID serves as a template on how transportation agencies can integrate data from disparate sources to improve decision making beyond safety. The RID has the potential to serve as a template for a national integrated database to support decision making in a performance measurement environment.

RID AS A STAND-ALONE DATABASE FOR HIGHWAY SAFETY R&D

The RID includes data from four sources: (1) a mobile data-collection effort conducted under SHRP2 project S04B to obtain detailed, high-quality data on 12,500 centerline-miles of roads most frequently driven by NDS participants and of greatest interest to safety researchers;

(2) acquired data from each of the six participating sites; (3) existing, supplementary datasets created and managed by public and private entities; and (4) Highway Performance Monitoring System (HPMS) data.

The mobile data collection for the RID included a quality assurance (QA) plan with accuracy requirements and tolerances for what was considered acceptable data.(6) The QA process involved random site visits to make "ground truth" measurements for comparison to the RID as well as equipment testing at control sites with additional ground truth information. Table 1 summarizes the accuracy requirements for the mobile-collected RID data.

Table 1. RID accuracy requirements.(6)

Data Element	Minimum Accuracy Requirement (+/−)
Curvature radius	100 ft (curves less than 1,500-ft radius) 250 ft (curves between 1,500-ft and 6,000-ft radius) Within 13% (curves over 6,000-ft radius)
Curvature length	100 ft (curves less than 1,500-ft radius) 250 ft (curves above 1,500-ft radius)
PC	50 ft
PT	50 ft
Grade (+ or –)	1.0%
Cross slope/ Superelevation	1.0%
Lane width	1 ft
Paved shoulder width	1 ft
Inventory features (signs) location	7 ft

PC = point of curvature; PT = point of tangency.

Data elements collected from already-existing sources and incorporated into the RID as supplemental data include historical crash data (spanning the time period from five years before the start of the SHRP2 data collection through its conclusion and two years after), traffic information in the form of AADT, aerial imagery, speed limit data, traffic laws in effect within each State during the data collection period (e.g., cell phone and text messaging, alcohol- and drug-impaired driving, graduated driver licensing, seat-belt use), and weather data.(6) While the RID was designed to link to NDS data, the level of coverage and accuracy of the newly collected mobile data, combined with the inclusion of supplemental data from existing sources, makes the RID a powerful stand-alone database.

The RID as a stand-alone database is particularly useful for R&D efforts that seek to relate the safety performance of road segments and/or intersections, measured by the frequency, type, and severity of crashes, to the segment and/or intersection traffic, traffic control, and geometric characteristics. Within these research areas, the RID provides quality data for horizontal curves, vertical grades, and intersections that historically have not often been available to researchers on a regular and widespread basis. The ability to improve the reliability of research results with highly accurate data at horizontal curves and intersections could be particularly impactful given that approximately 50 percent of traffic fatalities in the United States are associated with a horizontal curve or intersection.(3,4) A recent report from the FHWA Office of Safety titled

Reducing Roadway Departure Crashes at Horizontal Curve Sections on Two-Lane Rural Highways used the RID as a stand-alone database to create roadway departure crash modification factors (CMFs) for the following geometric design and traffic control features:[7]

- Combinations of horizontal curve radius, superelevation (including superelevation rate and presence of normal crown), and speed limit combinations representing side friction demand.
- Measures of horizontal alignment design consistency, including modified change radius rate and ratio of tangent length to radius.

The studies resulted in quantitative safety performance information that can be used in support of design policy and design criteria applications, such as those in the American Association of State Highway and Transportation Officials' *Policy on Geometric Design of Highways and Streets*.[8] In presenting and concluding the research, the report authors identified the following RID strengths:

- Served as an integrated data source with consistent data collection protocols and QA across multiple States.
- Provided additional variables for model specification (e.g., intersection presence, vertical grade, lane width, shoulder width) that likely improved estimates of design consistency and horizontal curvature effects by reducing omitted variable bias.
- Allowed the research team to build an adequate sample size of high-quality data in shorter periods of time.

FHWA continues to support the development of reduced datasets to further support these and other areas of highway safety research using the RID.

RID REDUCED DATASETS

Reduced datasets are intended to enhance the data within the RID and make the data more accessible to users with a range of GIS and data management skills. Making data more accessible will increase their use and application, thereby increasing opportunities for research to influence policies and practices and ultimately reduce the number of traffic fatalities and serious injuries.

The reduced datasets are unique, discrete outputs for data analysis. All reduced datasets have corresponding documentation describing how they were developed and their intended uses. Once developed, the reduced dataset is integrated within the data warehouse for each respective State. The reduced datasets can be used like all other features in the RID, including using the linear referencing system (LRS) and dynamic segmentation. The next chapter provides overviews of reduced datasets relevant to this white paper.

CHAPTER 2. EXISTING REDUCED DATASETS

This chapter provides an overview of the five new reduced datasets as well as previously developed reduced datasets that are complementary to the new ones. The chapter covers the following reduced datasets:

- AADT (new).
- Intersection widths (new).
- Intersection crashes (new).
- Curve crashes (new).
- Homogenous segments (new).
- Intersections.
- Curves.
- Speed limits.
- Mobile-HPMS presence.
- Lanes
- Divided/undivided.

AADT

The AADT reduced dataset was created to provide easy access to traffic data for users. AADT is a highly requested attribute used in various types of research projects, including traffic safety. While AADT is present in multiple features within the RID, inclusion is neither systematic nor consistent. Before the AADT reduced dataset, the following limitations existed in the AADT information within the RID:

- Features that only contain traffic data were inconsistently named between States.
- AADT was often included in other features that also contain other roadway data, such as lanes and shoulders, among others.
- Multiple years of data were often present in different data layers, requiring the user to decide which layer(s) to use.
- AADT coverage varied between features (some contain only State roadways while others contain all roadways).
- AADT data are not available in both travel directions of divided roadways.

The AADT reduced dataset addresses these issues by integrating all the AADT available in the RID (2010 to 2013). The final output provides an easier-to-use layer that contains a single record per roadway segment and eliminates the need for users to determine an appropriate data source. The AADT for each year that data were available in the RID is present for each roadway segment, eliminating the need for users to decide which year of AADT data to use or exclude roadways that do not have AADT for that given year.

There are some inconsistencies in how the original data were conflated to the RID's LRS, which can cause multiple AADT values to be associated to the same roadway segment. The reduced dataset has identified where ambiguity in the AADT exists and provides the minimum and maximum AADT reported. The AADT reduced dataset creates a unified AADT feature from multiple sources, which will make accessing AADT data easier for all types of users.

INTERSECTION WIDTHS

The intersection width reduced dataset (IWRD) provides a longitudinal representation of width measured in the direction of travel for both intersections and access points along routes where RID mobile data were collected. Access points represent locations, such as median openings and driveways. While the RID previously included intersection point locations, approximate longitudinal extents were not provided. Furthermore, the RID did not provide locations of possible access points along routes. Figure 1, figure 2, and figure 3 show examples of the intersection width dimensions provided by this reduced dataset.

IWRD attributes include the longitudinal extent (in feet) of intersections and the corresponding unique intersection identifier(s), if applicable, based on spatial coincidence and proximity. Unique intersection identifier(s) do not apply to access-related locations.

Intersection widths may be valuable to intersection performance- and safety-related research. For example, traffic engineers consider intersection width when establishing yellow and red times in a traffic signal timing plan. Intersection widths may therefore impact intersection performance and safety due to corresponding differences in yellow and red times. Intersection width also increases the exposure of vehicles traversing the intersection and vulnerable road users crossing the intersecting street. Additionally, performance and safety along segments of roadways may be impacted by the frequency and extent of intersections and access. As with intersections, access points increase possible vehicular conflict (e.g., median openings allowing left-turn movements across traffic) and may decrease operational efficiency.

The longitudinal extents of intersections and access locations within the IWRD are based primarily on two previously developed reduced datasets: (1) the divided/undivided reduced dataset and (2) the lanes reduced dataset. These reduced datasets were derived from the original RID median strip and lanes feature classes. Some inconsistencies exist within these feature classes, impacting final IWRD results.

Figure 1. Image. Example width of a four-leg intersection.

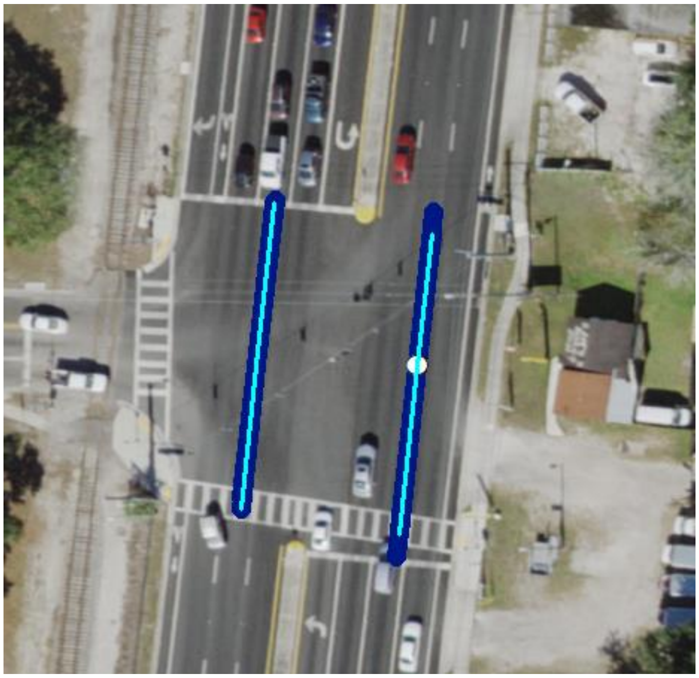


Figure 2. Image. Example widths for each travel direction along a divided road.

© Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community (see Acknowledgements).

Figure 3. Image. Example widths of an intersection and access point along an undivided road.

INTERSECTION CRASHES

The intersection crash reduced dataset (ICRD) provides users with easy access to intersection and intersection-related crash information captured from the State department of transportations (DOTs) as part of the RID supplemental data effort. The dataset was created to merge the historical crash data from each State to the intersection data collected during the RID mobile data collection. This type of information is valuable for any safety research. However, the following characteristics of the RID made systemically identifying the total number of intersection-related crashes difficult:

- Users had to make assumptions and decisions about what spatial threshold to employ to relate crashes to intersections.
- The same crashes may be located near multiple intersections.
- Multiple years of data are available, which requires a significant amount of time to process.
- Some States have data on and off the system, and these data are maintained independently of each other, which doubles the number of layers to process.
- The structure, information, field names, and coded values within the crash datasets are not consistent across all States.

The ICRD was developed using existing layers to address these issues as well as to provide new information that will reduce the time users need to invest to obtain information. The ICRD transforms inconsistent information among States and provides consistent outputs.

The ICRD differs from most other reduced datasets because no new layers are created. The ICRD instead enhances existing layers by joining intersection attributes to existing crash layers and crash information to the intersection reduced dataset. In total, three outputs were developed as part of the ICRD. The first output is a summary of crash information that has been joined to the previously created intersection reduced dataset. This is the primary output of the ICRD and provides the total number of crashes, crashes by year, and crashes by severity for each intersection along with other related information summarized for each intersection in the RID. The second output joins the nearest intersection information to each of the crash datasets. This provides easier access for users to identify to which potential intersection a crash is proximate. The third output is a near table[1]that relates all crashes and intersections that are within 500 ft of each other. When merging crash and intersection data for developing crash counts, a fixed threshold of 250 ft was used on divided roadways and 150 ft on undivided roadways. However, the near table also provides the ability for users to define their own threshold values (500 ft or less) and establish their own counts by intersection. These three outputs provide easier access for users when performing any safety analysis related to intersections.

The ICRD is a valuable tool for researchers when performing any safety-related research at intersections. Crash data vary significantly across States and, since data are separated by year, require significant effort when processing. The ICRD has simplified this work of processing intersection crash data for users.

CURVE CRASHES

The curve crash reduced dataset (CCRD) provides easy access for RID users to curve-related crash information captured from State DOTs as part of the RID supplemental data effort. The dataset was created to merge the historical crash data from each State to the horizontal curve data collected through the mobile effort. While this type of information is valuable for any safety research, the following characteristics of the data made it difficult to systemically identify the total number of curve-related crashes prior to the creation of the CCRD:

- Users typically do not have access to quality curve location information, including the points of curvature and tangency (begin and end of curves).
- The same crashes may be located near multiple curves, and the RID includes two curves on undivided routes representing travel directions, which results in the crash being associated with both curves.
- Multiple years of data are available, which requires a significant amount of time to process.
- Some States have crash data maintained independently on and off the system, which doubles the number of layers to process.

---

[1]"Near table" is an ArcGIS term referring to the process of calculating distances, and other proximity information, between GIS features, in this case, between features in two feature classes (intersection reduced dataset and crash data layers).[(9)]

- The structure, information, field names, and coded values within the crash datasets are not consistent across all States.

The CCRD was developed to address these issues and provide a layer that reduces the time needed to obtain the information. The curve reduced dataset, which uses data from the mobile data collection process, was used to identify the beginning and ending of each curve. A 250-ft buffer before and after the curve was also used to capture any crashes that may have occurred while approaching or departing the curve. The CCRD also transforms inconsistent information among States and provides consistent outputs.

As with the ICRD, the CCRD does not create any new layers. The CCRD instead enhances existing layers by joining curve attributes to existing crash layers and crash information to the curve reduced dataset. In total, three outputs were developed as part of the CCRD. The primary output of the CCRD is a summary of crash information that provides the total number of crashes, crashes by year, and crashes by severity for each curve along with other related information summarized for each curve in the RID. A second output joins the nearest curve information to each of the crash datasets. This provides easier access for users to identify to which potential curve a crash is proximate. The third output is a table that relates each curve to all crashes that fall within the curve or 250 ft before or after the start and end of the curve. These relational data were created through a dynamic segmentation process between the crash and curve data. The relational table provides the ability for users to define their own threshold values (i.e., only within the curve, within 100 ft of the curve) and summarize information by curve.

The CCRD will be a valuable tool for researchers when performing any safety-related horizontal curve research. Crash data are not standard among States, and since data are separated by year, use of the crash data requires significant processing. This CCRD reduces this work of processing crash data for users.

HOMOGENOUS SEGMENTS

The homogenous segments reduced dataset (HSRD) is a set of homogenous roadway segments based on roadway features collected through the mobile effort. These features are the most accurate within the RID. Homogenous segments represent continuous roadway extents and a least common denominator among the roadway features of interest. In other words, a new segment is created when any feature of interest changes. Directions of travel are also combined for undivided centerlines within the RID.

There are four different sets of the HSRD. Following are the features taken into consideration in segment definition:

- Alignment: tangent, curve, and not reported.
- Lanes: total number of through lanes, and not reported.
- Shoulder: paved, curb, unpaved (on either side of road), and not reported.
- Median strip: presence (any type), absence, and not reported.
- Lighting: presence, absence, and not reported.
- Rumble strip: presence (any type), absence, and not reported.

All features listed were generalized for homogenous segment definition. In addition, because several features are discontinuous at intersections, an attempt was made to eliminate such discontinuities less than 200 ft in length. Both feature generalization and discontinuity elimination were conducted, in part, to limit the number of potentially short segments and yield more continuous segments over longer extents.

Following are the four sets of the HSRD created and the corresponding features used for segment definition:

- HomogenousSegments_ReducedDataset: alignment, lanes, shoulder, median strip, lighting, and rumble strip.

- HomogenousSegmentsNoLighting_ReducedDataset: alignment, lanes, shoulder, median strip, and rumble strip.

- HomogenousSegmentsNoMedianStrip_ReducedDataset: alignment, lanes, shoulder, lighting, and rumble strip.

- HomogenousSegmentsNoLightingMedianStrip_ReducedDataset: alignment, lanes, shoulder, and rumble strip.

For HSRDs not using the lighting and/or median-strip features for segment definition, information about these features was integrated into the corresponding homogeneous segments. Specifically, the percentages of the homogeneous segment with lighting and/or percentage of the homogeneous segment with a median strip are provided. Similarly, two additional features, barriers and route intersections, collected during the mobile effort were associated with the homogeneous segments. The total length of barriers along each segment is provided as well as the total number of intersections.

In general, the HSRD creates a resource that may be used in evaluating roadway characteristics within the RID. The HRSD will save researchers time and effort in independently developing homogenous segmentation. Researchers may use the HRSD in identifying locations of interest, performing cross-sectional analysis, comparing performance among similar sites, and summarizing roadway characteristics. Researchers may also integrate the original features as well as any other RID datasets with the HSRD.

OTHER RELEVANT REDUCED DATASETS

The five reduced datasets described in the previous sections of this chapter can be used independently but will more often than not be merged with each other (e.g., AADT with homogenous segments), any existing layers within the RID, or with other previously created reduced datasets, including the following:

- Intersection.
- Curve.
- Speed limit.
- Mobile HPMS presence.

- Lane.
- Divided/undivided.

The following sections provide brief descriptions of these six previously developed reduced datasets. These descriptions, along with those for the five new reduced datasets, inform the corresponding research topics in Chapter 3.

Intersection Reduced Dataset

The intersection reduced dataset was developed to eliminate duplicate intersections that are present in the RID for the LRS to function properly. This dataset contains a single point representing each intersection, which simplifies use of the data, including integration with multiple other datasets, and helps limit confusion regarding the actual number of intersections present within the RID. In addition to intersection locations, the reduced dataset also contains distances to the nearest intersection and control type, which provides researchers additional capabilities to query the intersection based on distance to nearby intersections.

Curve Reduced Dataset

The curve reduced dataset was developed to summarize the roadway attributes present along horizontal curves. Without the reduced dataset, this process would require an experienced user to perform multiple dynamic segmentations that might break the curve into multiple shorter segments based on the location of the other roadway attributes. To simplify this process, roadway attributes were summarized along the curve, including the total number of lanes, the types of signs, grade, shoulder, and others. This dataset allows for a quick way to query the curves based on the roadway attributes.

The curve reduced dataset is valuable for any researcher investigating curves. This dataset has incorporated and summarized over 11 different features within the RID. This reduces a significant amount of work that a researcher would spend developing a similar dataset. The reduced dataset also uses information from the original collected data to provide attributes in the direction of travel of the curve instead of both directions of travel. The curves also provide a lookup value for finding the opposite direction curve.

Speed Limit Reduced Dataset

The intent of the speed limit reduced dataset was to provide the most accurate speed limits on all routes within the RID. The reduced dataset merged speed limit information from the mobile data collection, State-provided data, HPMS data, and link data. The merged datasets provide a single speed limit for all routes based on data source hierarchy. This resulted in the most accurate speed limit data being available for any given roadway.

Mobile HPMS Presence Reduced Dataset

The mobile HPMS presence reduced dataset represents roadway segments based on whether mobile or HPMS data are present. This dataset will allow users to more easily determine which data are available at a given location, without necessitating multiple dynamic segmentations or spatial joins. To create this feature class, mobile and HPMS feature classes were summarized based on presence. This process did not consider the attributes of the features, such as number of lanes, shoulder types, and other similar features. This reduced dataset only assessed whether there were any data available or present. An attribute field for each feature class analyzed (shoulder, rumble, lanes, HPMS, etc.) is included as well as a true/false value indicating whether the corresponding attribute data are available at the given location. The presence reduced dataset can be beneficial when analyzing locations to identify where data are located as well as where they are not. Without this reduced dataset, users would need to perform eight or more dynamic segmentations to determine the attributes for all mobile data collected, even if the user wants to eliminate locations with specific attributes. This process can require a significant amount of time, increase the opportunity for user error, and create a complex output with multiple attributes and short segments. Users can instead use the presence reduced dataset to perform a single dynamic segmentation and get a cleaner, easier-to-comprehend output.

The presence reduced dataset can be beneficial in exploratory analysis and for a full analysis. In an exploratory analysis, the dataset is valuable to get an initial estimate of how many data are available. This provides an initial estimate of how many miles or roadways have data that would be of interest. In addition, for some research questions the details are not required and only the presence of certain features is important.

Lane Reduced Dataset

The lane reduced dataset was created to provide the total number of lanes in both directions of travel. The lanes feature in the RID was collected for each direction and includes only the number of lanes in the given direction. This can often make it difficult and confusing to identify specific road types. For example, without the lane reduced dataset, a user wishing to identify two-lane roadways would have to query the data to find roadways with one through lane in a given direction and assume that the other direction also has only one through lane. The lane reduced dataset used dynamic segmentation to associate the different directions of travel and provide the total number of lanes in both directions while also retaining the original data for each direction. The lanes reduced dataset simplifies the process for users and eliminates the confusion that may be caused due to the direction collection of the lanes data.

Divided/Undivided Reduced Dataset

The divided/undivided reduced dataset provides a generalized, continuous, and consolidated representation of median presence and absence for all routes along which mobile data were collected. The reduced dataset provides more continuity along routes, since specific types of medians are of less interest. Three categories were established within the reduced dataset: (1) divided, (2) undivided, and (3) painted. The designation of undivided roadways is particularly useful, given the prior challenges associated with their systematic identification through use of the standard RID. Ultimately, this layer can be used for dynamic segmentation with other RID layers to determine whether the roadway/roadside characteristics selected are along divided, undivided, and painted roadways. Without this reduced dataset, multiple assumptions have to be made by the user, and these assumptions often vary from State to State based on the available data.

CHAPTER 3. POTENTIAL RESEARCH TOPICS

This chapter presents potential highway safety research topics that can be addressed with the support of the five new reduced datasets in combination with previously developed reduced datasets. The research topics are organized under the following seven general topic areas:

1. Crash prediction.
2. Safety performance impacts of horizontal curve features.
3. Safety performance impacts of intersection features.
4. Safety performance impacts of access management.
5. Risk factors for systemic safety analysis.
6. Crash assignments.
7. Driver awareness of signalized intersections when entering urban areas (following rural travel).

Within each topic area, there is a list of specific research topics and an indication of the applicable reduced datasets to support each topic.

CRASH PREDICTION

Crash prediction generally involves the development of regression models that relate expected crash frequency, by crash type and severity, at a road segment or intersection to characteristics of that same segment or intersection. Safety performance functions (SPFs) and severity distribution functions (SDFs) are two different types of regression models that play key roles in crash prediction. SPFs predict crash frequencies, while SDFs predict the proportions of crashes that are expected to fall into each defined level of severity. The SPF and the SDF in combination provide predictions of crash frequencies for each severity level of interest (e.g., predicted number of fatal crashes, predicted number of incapacitating injury crashes).

The RID reduced datasets can support the development of new SPFs and SDFs as well as the validation or calibration of already-existing SPFs and SDFs. Specific activities under this topic area include the development, validation, or calibration of the following:

- AADT-only intersection SPFs.
- Full-model intersection SPFs.
- AADT-only road segment SPFs.
- Full-model road segment SPFs.
- Intersection SDFs.
- Road segment SDFs.

The data required for AADT-only SPFs typically consist of crash counts of interest (e.g., all types and severities, roadway departure, fatal-plus-injury), AADT, and data to characterize the facility types for which the SPFs apply. For segments, the characterization is usually based on area type (i.e., urban, rural), number of lanes, and whether the facility is divided or undivided (e.g., rural, multilane divided highway). For intersections, it is based on area type, number of intersecting legs, and type of traffic control (e.g., urban, three-leg, stop-controlled intersections).

The data required for full-model SPFs are the same as the data needed for AADT-only SPFs, plus additional characteristics about the segments or intersections. Use of the RID for full-model SPFs may be particularly impactful given the higher levels of measurement accuracy and the availability of variables not typically available (e.g., horizontal curvature, vertical grade, traffic control devices).

The data required for SDF development consist of the same crashes and segments/intersections as the full-model SPF databases, but they are restructured so that the basic observation unit (i.e., database row) is the crash instead of the road segment or intersection. Table 2 links each of these development areas to the supporting reduced datasets described in chapter 2. In addition to the reduced datasets in table 2, users would need to link crashes to their respective homogenous segments and establish crash counts in a manner like the CCRDs and the ICRDs.

Once SPFs and SDFs are either developed, validated, or calibrated with the RID, RID users would also be able to compare what is predicted/expected at each respective RID site versus what was observed. The NDS could then be used to determine if there are noticeable differences in the behaviors of NDS drivers at locations that perform “worse than predicted/expected” versus those that perform “as predicted/expected” and “better than predicted/expected.”

Table 2. Supporting reduced datasets by crash prediction topics.

Research Topics	Intersection Crashes	Curve Crashes	Homogenous Segments	Curves	Intersections	Divided/Undivided	Lanes	Speed Limits	Intersection Width	AADT	Mobile HPMS Presence
AADT-only intersection SPFs	●	—	—	—	●	●	●	—	—	●	—
AADT-only segment SPFs	—	—	●	—	—	●	●	—	—	●	—
Full-model intersection SPFs	●	—	—	—	●	●	●	●	●	●	●
Full-model segment SPFs	—	●	●	●	—	●	●	●	●	●	●
Intersection SDFs	●	—	—	—	●	●	●	●	●	●	●
Segment SDFs	—	●	●	●	—	●	●	●	●	●	●

—No data.
●Reduced dataset supports the crash prediction research topic.

SAFETY PERFORMANCE IMPACTS OF HORIZONTAL CURVE FEATURES

FHWA estimates that more than 50 percent of traffic fatalities result from roadway departure crashes. Vehicles are more likely to depart the roadway at locations with horizontal curves. Some estimates show that the average crash rate along horizontal curve sections of rural, two-lane highways, for example, is three times higher than on tangent roadway sections and that the severity of roadway departure crashes along horizontal curves is greater than the severity of roadway departure crashes along tangent roadway sections. Quantifying the impacts of the presence and dimensions of horizontal curve characteristics remains an active area of research. A significant portion of previous research has been unable to account for impacts of curve radius, superelevation, and vertical grade because of the difficulty of obtaining a large sample of these particular data elements. The RID can help fill this gap.

With the reduced datasets discussed in this white paper, estimating the safety performance impacts of the following features and their potential interactions may be possible:

- Curve radius.
- Curve superelevation.
- Curve side friction demand (estimated using speed limit, radius, and superelevation).
- Curvature change rate.
- Curve direction.
- Roadway width and allocation of paved width (i.e., different combinations of lane width and paved shoulder width).
- Intersection presence.
- Lighting presence.
- Rumble strip presence.
- Delineation and signing.
- Barrier presence.

Table 3 provides a summary of relevant supporting reduced datasets for the horizontal curve research topics.

Table 3. Supporting reduced datasets for horizontal curve topics.

Reduced Datasets	Applicability
Intersection crashes	—
Curve crashes	●
Homogenous segments	—
Curves	●
Intersections	—
Divided/undivided	—
Lanes	—
Speed limits	—
Intersection width	—
AADT	●
Mobile HPMS presence	—

—No data.

●Reduced dataset supports the horizontal curve research topics.

Estimating the safety performance impacts of these features with the RID will occur with a cross-sectional study, as the RID mobile data represent a snapshot in time. Whether estimated regression parameters from cross-sectional studies can ever represent safety effects estimates is a matter of ongoing debate. More recent safety research has implemented cross-sectional study designs intended to mimic randomized experiments and strengthen confidence in safety effects estimates, such as the propensity scores-potential outcomes framework. In addition to estimating the safety performance impacts of horizontal curve features, the RID also serves as a database through which to continue these types of methodological advancements as well.

The ability to merge the RID with the NDS allows researchers to gain a more thorough understanding of their interpretations of crash-based findings and of the potential behavior mechanisms leading to those findings. Donnell et al. provide one such example of future research related to the measures of design consistency for which they obtained crash-based results with the RID, concluding that "the NDS itself could be used to try and verify the speed and lane keeping impacts of the horizontal curve and design consistency characteristics explored in this study to further support interpretations of the findings."(7)

SAFETY PERFORMANCE IMPACTS OF INTERSECTION FEATURES

When designing intersections, there is a need to provide a reasonable level of safety while maintaining operations for the intersecting roadways. Safety and operational performance depend on several factors, including intersection geometry and traffic control. There has been a great deal of research focused on the safety performance impacts of individual intersection features, such as the presence of turn lanes, type of signal phasing, and presence of horizontal curvature near the intersection.

The reduced datasets provide an opportunity to further explore the safety impacts of these features, with a particular interest in the impacts of different combinations of features (i.e., interactions). For example, right-turn lanes are one potential strategy to improve the safety and operational performance of intersections by separating right-turning vehicles from through vehicles. While right-turn lanes have several potential benefits (e.g., to separate turning vehicles from through vehicles, manage conflicts, and define intent of turning vehicles), these same turn lanes can also contribute to safety issues at the intersection. One concern is related to sightline obstructions, where a vehicle in the right-turn lane on the major road could obstruct the view of drivers on the minor road, potentially contributing to crashes. Other factors, such as the length of turn lanes, presence of horizontal curves, vehicle speeds, and traffic volumes, may exacerbate this issue. The reduced datasets will allow further research into the interaction effects among these various features.

With the reduced datasets discussed in this white paper, estimating the safety performance impacts of the following features and their potential interactions may be possible:

- Number of approach legs.
- Type of traffic control.
- Proximity to nearby intersections.
- Speed limit on approach legs.
- Intersection width.
- Lane arrangement.

Table 4 provides a summary of relevant supporting reduced datasets for the intersection research topics. The same discussion regarding cross-sectional study designs and the use of the NDS to inform the interpretation of findings that was relevant to horizontal curve features is also relevant here. One such example includes an evaluation of left-turn lane offset on driver behavior by Hutton et al.[10]

Table 4. Supporting reduced datasets for intersection topics.

Reduced Datasets	Applicability
Intersection crashes	●
Curve crashes	●
Homogenous segments	—
Curves	●
Intersections	●
Divided/undivided	●
Lanes	●
Speed limits	●
Intersection width	●
AADT	●
Mobile HPMS presence	●

—No data.
●Reduced dataset supports the intersection research topics.

SAFETY PERFORMANCE IMPACTS OF ACCESS MANAGEMENT

Access management is an important component in the operational and safety performance of roadways. For example, the safety performance of an intersection depends on factors such as the type of turning movements, presence and length of turn lanes, and the number and proximity of driveways. The safety performance of a corridor depends on factors such as the density and spacing of intersections, driveways, and median openings. One challenge to quantifying the safety impacts of access management decisions is that reliable information is relatively limited. There are multiple reasons for the limited quantity and quality of information on the safety effects of access management strategies. Part of the reason is the difficulty in developing high-quality CMFs and SPFs for access management strategies due to challenges related to data availability, sample size, and statistical methods. There is an opportunity to use the reduced datasets to better understand the safety performance impacts of access management strategies and fill critical knowledge gaps. Of particular interest is the effect of features such as medians, median openings, access density and spacing, and turn lanes, individually and in combination.

With the reduced datasets discussed in this white paper, estimating the safety performance impacts of the following features and their potential interactions is possible:

- Median presence.
- Driveway spacing or density.
- Driveway width.
- Median opening spacing or density.
- Median opening width.

Table 5 provides a summary of relevant supporting reduced datasets for the intersection research topics. In addition to the reduced datasets in table 5, users need to define segments or corridors over which to analyze the access characteristics and link crashes to their respective segments.

Table 5. Supporting reduced datasets for access management topics.

Reduced Datasets	Applicability
Intersection crashes	●
Curve crashes	—
Homogenous segments	—
Curves	—
Intersections	—
Divided/undivided	●
Lanes	—
Speed limits	—
Intersection width	●
AADT	●
Mobile HPMS presence	—

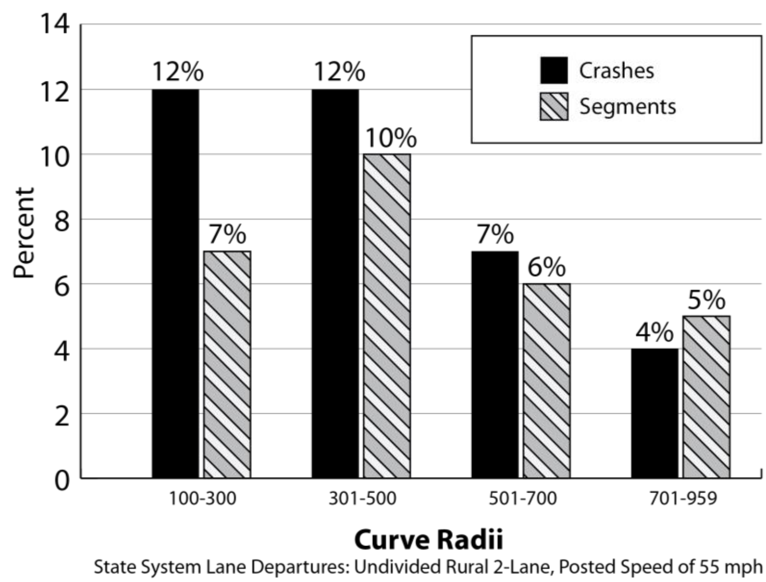
—No data.

●Reduced dataset supports the access management research topics.

RISK FACTORS FOR SYSTEMIC SAFETY ANALYSIS

There are two general approaches to road safety management: (1) selecting and treating sites based on the frequency and severity of crashes observed at specific sites (sometimes called the hot-spot approach) and (2) selecting and treating sites based on site-specific geometric, operational, and other attributes known to increase crash potential and/or severity (the systemic approach). The systemic approach analysis selects focus crash types, facility types, and risk factors. Focus crash types typically reflect prevalent severe crash types for a given jurisdiction. Focus facility types typically include the facility and site types where the focus crash types are most prevalent (e.g., rural, two-lane, undivided segments or urban, four-leg, signalized intersections). Risk factors are site-specific attributes associated with an increased potential of the focus crash types occurring. Risk factors may include site-specific crash history (if available), geometric and operational characteristics, and surrounding sociodemographic and environmental attributes.

FHWA's *Systemic Safety Project Selection Tool* provides information on the identification of focus crash types, focus facility types, and risk factors.(11) Risk factors may be identified using full-model SPFs and SDFs, such as those described in the Crash Prediction section of this white paper. Other approaches include regression trees or basic descriptive statistics approaches, such as the New York State DOT evaluation of curve radii as a potential risk factor shown in figure 4. The RID may offer a unique opportunity to increase our knowledge on what the key risk factors are for higher priority crash types that lead to significant numbers of fatalities (e.g., roadway departure on horizontal curves, angle crash at intersections). A multi-State analysis may increase the statistical power of findings and the availability of traffic control device information offers opportunities to analyze how risk factors change in the presence/absence of signing or delineation. Such a multi-State analysis was challenging without the RID due to the inconsistency in the availability of the same information across States.



Source: FHWA.(11)

Figure 4. Graph. New York State DOT evaluation of curve radii as a potential risk factor.(11)

CRASH ASSIGNMENT

The ICRDs and CCRDs provide users with flexibility to define different buffer distances around intersections and curves to capture associated crashes. There is no widely accepted process for determining buffer distances. For intersections, appropriate buffer distances may vary as a function of intersection characteristics, such as traffic control, approach speeds, number of lanes, and presence of turn lanes. Similarly, appropriate buffer distances for horizontal curves may also vary as a function of curve characteristics. The reduced datasets offer an opportunity to conduct a systematic exploration of buffer distances and develop recommendations for future research.

DRIVER AWARENESS OF SIGNALIZED INTERSECTIONS WHEN ENTERING URBAN AREAS (FOLLOWING RURAL TRAVEL)

This applied research topic was identified in *NCHRP Report 756: Highway Safety Research Agenda: Infrastructure and Operations*.[(12)] Traditionally, identification of appropriate and sufficient analysis sites has been quite challenging. Also challenging, and potentially costly, was monitoring driver responses at identified sites. The RID, in conjunction with the NDS, facilitates identification of appropriate first signalized intersections and assessment of corresponding driver response through use of corresponding NDS data.

Initially, standard RID datasets may be employed, in part, to identify qualifying first signalized intersections and the presence and type of advance treatments. Both HPMS "TOPS" and Incorporated Areas and Designated Places datasets may be used to differentiate urban and rural areas. Route intersection or the reduced intersection datasets may then be used to identify the first signalized urban intersection. Given the qualifying intersections, signs may be incorporated to identify the presence and type of any advanced treatments, such as real-time activated or static flashers.

Integration of several RID reduced datasets would be necessary to further support and enhance analysis. For example, intersection crashes provide historic crash experience at the qualifying intersections, while AADT provides the pertinent exposure data. Speed limits, coupled with homogenous segments, may be used to categorize qualifying intersections based on the types of facilities on which they are located. Speed limit may also be used with the NDS data to compare driver speeds and posted speed limits prior to the qualifying intersections and/or advanced treatments. Ultimately, through investigation of crash experience and NDS-based driver behavior, the safety effects of various advanced treatments strategies along different types of facilities may be quantified. Table 6 provides a summary of relevant supporting reduced datasets for the intersection awareness research topic.

Table 6. Supporting reduced datasets for first signalized intersection topic.

Reduced Datasets	Applicability
Intersection crashes	●
Curve crashes	—
Homogenous segments	●
Curves	—
Intersections	●
Divided/undivided	—
Lanes	—
Speed limits	●
Intersection width	—
AADT	●
Mobile HPMS presence	—

—No data.

●Reduced dataset supports the intersection awareness research topic.

CHAPTER 4. SUMMARY AND CONCLUSIONS

The RID is an accurate and detailed database of roadway, traffic, crash, and safety countermeasure information that has applicability to a diverse range of safety research questions, both as a stand-alone database and in combination with NDS data. The RID as a stand-alone database is particularly useful for research efforts that seek to relate the safety performance of road segments and/or intersections, measured by the frequency, type, and severity of crashes, to the segment and intersection traffic, traffic control, and geometric characteristics. By integrating data within a uniform schema over a multi-State context, the RID allows researchers to develop robust, directly comparable datasets that are often not achievable through traditional open data platforms or data requests. This uniformity and consistency, both temporally and spatially, helps address some of the challenges associated with obtaining data from multiple sources.

Furthermore, the reduced datasets within the RID are highly safety-relevant resources that allow users to streamline the data collection and integration process. The RID contains 11 reduced datasets as of April 2019, and these complementary datasets encompass several critical dimensions of safety analysis. Each dataset could stand alone as a unit of analysis, be integrated with original data collected by a researcher, or be integrated with other reduced datasets and the NDS. This flexibility allows the RID to answer questions about transportation infrastructure and driver behavior that most other datasets are incapable of addressing.

Chapter 3 discussed several potential applications of the RID reduced datasets and highlights the specific datasets pertinent to each research question. However, the relevance of the RID to safety research is not limited to these examples. By providing a robust, geographically diverse dataset in a readily available format, the RID allows the research community to spend its efforts expanding the breadth and detail of individual research questions. This foundation of data provided by the RID enables researchers to iterate and test multiple approaches within the context of the dataset, rather than spend significant research resources on collecting and assembling data in an ad hoc and piecemeal fashion.

ACKNOWLEDGEMENTS

The original maps in figure 1 through figure 3 are the copyright property of ArcMap Imagery Basemap. The map overlays showing example intersection width dimensions provided by the IWRD were developed for this research project. The overlays include the dark blue lines, light blue lines, and white dots showing the widths of intersections and access points.

REFERENCES

1. Transportation Research Board. "What is Happening During Safety Data Phase 1?" (website). National Academies of Science, Engineering, and Medicine, Washington, DC. Available online: http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/SHRP2DataSafetyPhase1.aspx. Last accessed on April 8, 2019.

2. Transportation Research Board, Strategic Highway Research Program, and Virginia Tech Transportation Institute. (2018). "InSight Data Access Website: SHRP2 Naturalistic Driving Study" (website). Virginia Tech Transportation Institute, Blacksburg, VA. Available online: https://insight.shrp2nds.us/. Last accessed on April 8, 2019.

3. National Highway Traffic Safety Administration. (2019). "Query FARS Data" (website). Washington, DC. Available online: https://www-fars.nhtsa.dot.gov/QueryTool/QuerySection/SelectYear.aspx. Last accessed on April 8, 2019.

4. Federal Highway Administration. (2019). "Horizontal Curve Safety" (website). Washington, DC. Available online: https://safety.fhwa.dot.gov/roadway_dept/countermeasures/horicurves/. Last accessed on April 8, 2019.

5. National Highway Traffic Safety Administration. (2011). "Model Performance Measures for State Traffic Records Systems" (website). Washington, DC. Available online: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811441. Last accessed on April 8, 2019.

6. Smadi, O., Hawkins, N., Hans, Z., Bektas, B., Knickerbocker, S., Nlenanya, I., Souleyrette, R., and Hallmark, S. (2015). *Naturalistic Driving Study: Development of the Roadway Information Database*, SHRP2 Report S2-204A-RW-1, Transportation Research Board, Washington, DC.

7. Donnell, E., Porter, R., Li, L., Hamilton, I., Himes, S., and Wood, J. (In press). *Reducing Roadway Departure Crashes at Horizontal-Curve Sections on Two-Lane Rural Highways*, FHWA-SA-19-005, Federal Highway Administration, Washington, DC.

8. American Association of State Highway and Transportation Officials. (2018). *A Policy on Geometric Design of Highways and Streets: Seventh Edition*, Washington, DC.

9. Environmental Systems Research Institute (Esri) (2018). ArcGIS Desktop: Release 10.6.1. Redlands, CA.

10. Hutton, J.M., Bauer, K.M., Fees, C.A., and Smiley, A. (2015). "Evaluation of Left-Turn Lane Offset Using the Naturalistic Driving Study Data." *Journal of Safety Research, 54*(September 2015), pp. 5-e1–15, Elsevier, Amsterdam, Netherlands.

11. Preston, H., Storm, R., Dowds, J.B., Wemple, B., and Hill, C. (2013). *Systemic Safety Project Selection Tool*, FHWA-SA-13-019, Federal Highway Administration, Washington, DC.

12. Zegeer, C.V. (2013). *NCHRP Report 756: Highway Safety Research Agenda: Infrastructure and Operations*, Transportation Research Board, Washington, DC.

Recycled
Recyclable

HRDS-20/12-19(WEB)E