Speed–Safety Analyses Using Linked National Performance Management Research Dataset (NPMRDS) and SHRP2 Roadway Information Database (RID) Data

U.S. Department of Transportation
Federal Highway Administration

FOREWORD

This report documents a case study in which data from the Transportation Research Board's second Strategic Highway Research Program Roadway Information Database (RID) were linked with data from the Federal Highway Administration's National Performance Management Research Data Set (NPMRDS).[1,2] Speed and speed–safety relationships were then investigated using the linked NPMRDS-RID data. The study focused on urban interstates and divided major arterials in Washington State. The NPMRDS contains travel-time data for highway segments from which operating speeds along the segments (links) were estimated. Relationships between operating speeds and posted speeds for passenger cars during noncongested conditions were determined through regression analysis. In addition, linear regression was conducted to address the research question: How is crash severity related to the differential between the posted speed and the 85th-percentile operating speed for passenger cars on urban interstates and major divided arterials in Washington State during noncongested conditions?

The target audience for this report includes researchers examining speed–safety relationships and highway safety practitioners.

James S. Pol, P.E., PMP
Acting Director, Office of Safety
Research and Development

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-HRT-18-046	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Speed–Safety Analyses Using Linked National Performance Management Research Data Set (NPMRDS) and SHRP2 Roadway Information Database (RID) Data		5. Report Date November 2018	
		6. Performing Organization Code	
7. Author(s) Michael Dimaiuta, Mohamad Banihashemi, Bruce Spear, Robert Hull, Pamela Beer, Omar Smadi (ORCID: 0000-0002-3147-9232), and Zach Hans		8. Performing Organization Report No.	
9. Performing Organization Name and Address Genex Systems 11848 Rock Landing Drive Newport News, VA 23606 Cambridge Systematics 101 Station Landing, Ste. 410 Medford, MA 02155 Center for Transportation Research and Education Iowa State University 2711 S. Loop Drive, Ste. 4700 Ames, IA 50010		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. Contract DTFH61-14-D00043; Task Order DTFH61-14-D00043-T-5010	
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Safety Research and Development 6300 Georgetown Pike McLean, VA 22101		13. Type of Report and Period Covered Research Report; September 2015–December 2017	
		14. Sponsoring Agency Code HRDS-10	
15. Supplementary Notes Abdul Zineddin (HRDS-10; Abdul.Zineddin@dot.gov), Office of Safety Research and Development, served as the Task Order Task Manager for the Federal Highway Administration.			

16. Abstract

The case study documented in this report involves linking the second Strategic Highway Research Program (SHRP2) Roadway Information Database (RID) with the National Performance Management Research Data Set (NPMRDS) and then using the linked data to examine speed and speed–safety relationships.[1,2] This report includes the following information:

- An overview of the SHRP2 safety data.
- An overview of the RID and the NPMRDS.
- The methodology developed to link the RID with the NPMRDS.
- Additional data preparation required to conduct analyses using the linked data.
- A summary of the analyses conducted.
- A description of the value of the results related to highway safety.
- Proposed future research directions.

17. Key Words Safety, SHRP2, Roadway Information Database, geographic information system, NPMRDS		18. Distribution Statement No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161. http://www.ntis.gov	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 104	22. Price N/A

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

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

APPROXIMATE CONVERSIONS FROM SI UNITS

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

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

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AADT	annual average daily traffic
AIC	Akaike information criterion
ANOVA	Analysis of Variance
CMF	crash modification factor
csv	comma-separated value
dbf	database file
DCM	Design Consistency Module
DVM	Driver/Vehicle Module
FHWA	Federal Highway Administration
FI	fatal and injury
GIS	geographic information system
GPS	Global Positioning System
HPMS	Highway Performance Monitoring System
HSM	*Highway Safety Manual*
IHSDM	Interactive Highway Safety Design Model
LRS	linear referencing system
MPO	metropolitan planning organization
NCHRP	National Cooperative Highway Research Program
NDS	Naturalistic Driving Study
NHS	National Highway System
NPMRDS	National Performance Management Research Data Set
RID	Roadway Information Database
RSA	Road Safety Audit
shp	shapefile
SHRP2	second Strategic Highway Research Program
SPF	safety performance function
TAM	Traffic Analysis Module
TMC	Traffic Message Channel
TT	travel time
v	version
vpd	vehicles per day
xlsx	spreadsheet format file extension

EXECUTIVE SUMMARY

Phase 1 of this project (Task Order DTFH61-09-D-00020-T-14001) identified available data sources that can be linked with the second Strategic Highway Research Program (SHRP2) Roadway Information Database (RID) to offer additional insights into traffic safety performance and included a detailed assessment of the analytical possibilities associated with these linkages.[1]

In phase 1, an Innovative Data Source Case Study Outline, which proposed linking the National Performance Management Research Data Set (NPMRDS) with the RID, was presented.[2] The objectives for Task Order DTFH61-14-D00043-T-5010 ("GIS [geographic information system] Linkages with the SHRP2 Roadway Information Database (RID): Phase 2") were the following:

- To complete the analysis described in the phase 1 Case Study Outline to demonstrate the value of GIS linkages with the RID in support of highway safety.

- To assist in improving the usability of the RID by the research community.

This report summarizes the project team's findings.

RID

SHRP2 "was established by Congress to investigate the underlying causes of highway crashes and congestion in a short-term program of focused research. The objective is to identify countermeasures that will significantly improve highway safety through an understanding of driving behaviors."[3]

SHRP2 safety research centers on the Naturalistic Driving Study (NDS).[3] The SHRP2 NDS website states the following:

> The main goal of the NDS is to address the role of driver performance and behavior in traffic safety. This includes developing an understanding of how the driver interacts with and adapts to the vehicle, traffic environment, roadway characteristics, traffic control devices, and environment. It also includes assessing the changes in collision risk associated with each of these factors and interactions. This information will support the development of new and improved countermeasures with greater effectiveness.[4]

Six field study sites were selected for the NDS: Tampa, FL; Bloomington, IN; Durham, NC; Buffalo, NY; Central Pennsylvania; and Seattle, WA.[3] Two projects gathered roadway information for the six NDS sites, leading to the development of the RID.

RID Structure and Data Elements

The RID was built on a milepoint-based linear referencing system (LRS), which was based on the Esri® StreetMap Premium™ for ArcGIS® all-roads street-centerline network, using a common roadway-centerline network covering all six study areas.[5] Roadway features, such as

those collected by mobile vans under SHRP2 Project S04B (Mobile Data Collection), were conflated to this LRS.[6,7] Data sources added to the RID include the following:

- **Highway Performance Monitoring System (HPMS):**[8] Data were submitted to the Federal Highway Administration (FHWA) by the State transportation departments of each of the six NDS study sites for submittal year 2012.[3] The main HPMS data elements included in the RID are functional classification, traffic volume, number and type of lanes, lane width, shoulder width and type, median width and type, and some geometry and grade information.[8]

- **Supplemental State roadway data (phase 1):** Additional roadway-inventory data were obtained from each of the State transportation departments. These data vary from State to State in terms of content and coverage.

- **Supplemental State roadway data (phase 2):** These data include crash histories, traffic, weather, work zones, changes to infrastructure, and Federal Railroad Administration– grade crossings.

- **Additional supplemental data:** These data include aerial imagery, safety enforcement laws, and active safety campaigns.

The RID is maintained as a set of geodatabases—one for each of the six study sites, plus a seventh containing roadway-centerline data for adjacent States.

NPMRDS

In July 2013, FHWA acquired a national travel-time (TT) dataset from HERE® (formerly NAVTEQ) to support Federal mobility applications.[9] The NPMRDS version 1 contains passenger and freight TT datasets for the National Highway System and other roadways.[10]

The NPMRDS includes three categories of average TTs obtained from vehicle-probe data: all vehicles, passenger vehicles, and freight trucks.

NPMRDS data are made available to States and metropolitan planning organizations to "use for their performance management activities."[2] The FHWA Office of Operations controls access to the NPMRDS database.

All NPMRDS data have been available since July 2013 and are updated monthly. The period for which both RID crash data and NPMRDS data are available is July through December 2013. NPMRDS coverage overlaps with all six RID States (Florida, Indiana, North Carolina, New York, Pennsylvania, and Washington).

The NPMRDS consists of several components, including the following:

- Traffic Message Channel (TMC) static file.
- TT data.
- NPMRDS Links:
  - NPMRDS Links shapefile.
  - NPMRDS Links to TMC look-up table.
- NPMRDS supporting documentation.

Each roadway segment in the NPMRDS is defined by a unique identifier known as the TMC code. TMC codes were developed to provide an efficient way to transmit location information that is independent of the accuracy or level of resolution of the map on which the data are displayed. A TMC segment corresponds to a section of roadway between two well-defined end points in a specific direction of travel.

TTs are defined for TMC segments. The TT data represent the average time each vehicle type (i.e., passenger vehicle, truck, or all vehicles) takes to traverse a specific roadway segment on a specific day and time period (e.g., Friday, January 30, 8:00 to 8:05 a.m.).

The NPMRDS Links shapefile is a geospatial database that contains the geometric alignments of the NPMRDS Links derived from the HERE® roadway-centerline network (i.e., the road geometry for the highways covered in the database).[9] The NPMRDS Links and their geometric alignments are very similar to, but do not match exactly, the roadway-centerline segments in the RID.

LINKAGE OF NPMRDS AND RID DATA

The NPMRDS-RID linkage process had the following 10 steps:

1. Perform the initial setup (including the creation of a new geodatabase that contains the desired feature classes from the RID as well as link TTs from the NPMRDS).

2. Identify common links between the NPMRDS and RID.

3. Import the NPMRDS Links–TMC look-up table.

4. Identify appropriate TMC records.

5. Identify the directional relationship between TMCs and links.

6. Refine/limit common links and independently manage directional data.

7. Assign TMCs to limited common links.

8. Import TT data.

9. Calculate weekday free-flow speed.

10. Combine TT and speed data with linkage data.

The main goal of the linkage process was to build a geodatabase that contains common links from the RID and NPMRDS accompanied by TTs, speeds, and speed limits. The final product of these steps was a geodatabase containing selected RID and NPMRDS feature classes as well as some intermediate feature classes and tables produced in this process.

The NPMRDS provides TT information for TMC segments. Using the length of the TMC segments, the operating speed on roads can be computed. TT data are collected in near real time (every 5 min), which offers a unique temporal aspect of road service conditions. The RID contains posted-speed data on the roadways but not actual operating-speed data. By linking NPMRDS "continuous"-travel-speed data with RID crash and roadway data, advances in studying the speed–safety relationship are possible.

ANALYSES PERFORMED USING THE LINKED NPMRDS-RID DATA

Estimated and free-flow 85th-percentile operating speed (V_{85}) and average operating speed (V_{avg}) were derived from TTs extracted from the NPMRDS for links matched with the RID. The objective was to compute the actual free-flow speeds on links in the Washington State RID using TT data from the NPMRDS.

The following analyses were conducted using the linked NPMRDS-RID data:

- **Analysis 1:** Studied the relationship between the noncongested V_{85} and the speed limit (posted speed) during noncongested conditions.

- **Analysis 2:** Studied the relationship between crash severity and the differential between posted and operating speeds during noncongested conditions.

Analysis 1

The following research questions were addressed in analysis 1:

- What is the relationship between V_{avg} and the posted speed for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

- What is the relationship between V_{85} and the posted speed for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

Analysis 2

The objective of analysis 2 was to investigate whether crash severity can be explained by the magnitude of the differential between V_{85} and the speed limit.

The research question addressed in analysis 2 was as follows: How is crash severity related to the differential between the posted speed and V_{85} for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

SUMMARY OF RESULTS

Summary of Analysis 1 Results

Analysis 1 results include regression models for estimating V_{85} and V_{avg} during off-peak (noncongested) periods as a function of the posted speed and weighted average of degree of curvature.

V_{85} *As a Function of Posted Speed and Weighted Average of Degree of Curvature*

The results suggest that, on sections with very flat, horizontal curves (weighted-average radius of 5,000 ft), V_{85} exceeds the speed limit when the posted speed is less than 65 mph. V_{85} exceeds the speed limit by as much as 30 percent at lower speeds (i.e., 15 percent of drivers exceed the posted speed by as much as 30 percent at lower speeds), with the differential decreasing as the speed limit increases. For speed limits of 65 mph and above, V_{85} is lower than the speed limit.

For sections with sharper curves (weighted-average radius of 1,000 ft), V_{85} exceeds the speed limit when the posted speed is less than 60 mph; for a speed limit of 40 mph, the differential is nearly 25 percent. As with flatter sections, the differential between V_{85} and posted speeds decreases as the speed limit increases. For speed limits of 60 mph and above, V_{85} is lower than the speed limit.

For sections with sharp curves (weighted-average radius of 500 ft), V_{85} is lower than the speed limit for all posted speeds, with the differential increasing as the speed limit increases.

Therefore, the amount by which the speed limit is exceeded by V_{85} depends, in part, on how sharp the curves are. When speed limits are above 65 mph, drivers appear less likely to exceed the posted speed, even when the road alignment is relatively straight.

V_{avg} *As a Function of Posted Speed and Weighted Average of Degree of Curvature*

The results indicate that average speeds are consistently below the posted speeds even on relatively straight roads. For sections with weighted-average curve radii of 5,000 ft, the differential between V_{avg} and the posted speed ranges from 7 to 10 mph, depending on the posted speed. The slopes of V_{avg} versus posted speed plots are similar for sections with sharper curves.

Summary of Analysis 2 Results

Analysis 2 results include a model for the relationship of crash severity versus the differential of posted and operating speeds in noncongested conditions. Crash severity was represented by the ratio of fatal and injury (FI) crashes to total crashes.

The results suggest that, as the operating versus posted speed differential increases, the ratio of FI crashes to total crashes decreases. Although this relationship is somewhat counterintuitive, there is a potential explanation. The analysis 1 results show that differentials between operating speed and posted speed are greater on sections with lower posted speeds. Since crash severity is influenced by speed (with higher speeds generally resulting in more severe crashes), an expectation is that crashes are relatively less severe at lower speeds than at higher speeds. Since the greater speed differentials occur on sections with lower posted speeds, it would reason that

the ratio of FI crashes to total crashes could be lower on those sections compared to sections with smaller speed differentials (i.e., sections with higher posted speeds).

INTRODUCTION

The case study conducted in this task order involves linking the second Strategic Highway Research Program (SHRP2) Roadway Information Database (RID) with the National Performance Management Research Database (NPMRDS) and then using the linked data to examine speed and speed–safety relationships.[1,2] This report summarizes the project team's findings. The report provides the following information:

- An overview of the SHRP2 safety data.[4]
- An overview of the RID and NPMRDS.
- The methodology developed to link the RID with the NPMRDS.
- Additional data preparation required to conduct analyses using the linked data.
- A summary of the analyses conducted.
- A description of the value of the results related to highway safety.
- Proposed directions for future research.

BACKGROUND: THE SPEED–SAFETY RELATIONSHIP

Speed is widely recognized as having an impact on safety. However, the effects of speed on safety are complex and only partially understood. Despite much research, only a limited number of consistent and reliable speed–safety relationships have been established. There are very few points of consensus on the effects of speed on crash probability, though there is general agreement that the risk of injuries and fatalities increases with speed.[11]

The lack of quantified speed–safety relationships is reflected in the first edition of the *Highway Safety Manual* (HSM) (e.g., the crash prediction models in HSM part C do not include speed among the input variables).[12] The HSM does note that "the relationship between speed and safety depends on human behavior and driver adaptation to roadway design, traffic control, and other roadway conditions" and suggests the following:

> [T]he relationship between speed and safety can be examined during the "pre-event" and the "event" phases of a crash. The "pre-event" phase considers the probability that a crash will occur, specifically how this probability depends on speed. The "event" phase considers the severity of a crash, specifically the relationship between speed and severity. Identifying the errors that contribute to the cause of crashes helps to better identify potential countermeasures.[12]

Studying the relationship between speed and safety performance requires detailed speed and crash data and a way to link these two with reliable precision.

OBJECTIVE

The process described in this report is for a specific example in which free-flow speeds are derived from NPMRDS travel-time (TT) data for links matched with the RID for urban interstates and major arterials in Washington State. The resultant linked database contains actual free-flow speeds on links in the Washington State RID using TT data from the NPMRDS.

The NPMRDS provides TT information for Traffic Message Channel (TMC) segments. Using the length of the TMC segments, the operating speed on roads can be computed. TT data are collected in "near real time" (every 5 min), which offers a unique temporal aspect of road service conditions. The RID contains posted-speed data on the roadways but not actual operating-speed data. By linking NPMRDS "continuous"-travel-speed data with RID crash and roadway data, advances in studying the speed–safety relationship are possible.

The objective of the study was to address the following research questions using the linked NPMRDS-RID data:

- What is the relationship between the average operating speed (V_{avg}) and posted speed for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

- What is the relationship between the 85th-percentile operating speed (V_{85}) and posted speed for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

- How is crash severity related to the differential between the posted speed and V_{85} for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

SHRP2 SAFETY/NATURALISTIC DRIVING STUDY OVERVIEW

SHRP2 "was established by Congress to investigate the underlying causes of highway crashes and congestion in a short-term program of focused research. The objective is to identify countermeasures that will significantly improve highway safety through an understanding of driving behaviors."[3]

SHRP2 focus areas include safety, renewal, reliability, and capacity. SHRP2 safety research centers on the Naturalistic Driving Study (NDS).[3] The central goal of the NDS is the following:

> to address the role of driver performance and behavior in traffic safety. This includes developing an understanding of how the driver interacts with and adapts to the vehicle, traffic environment, roadway characteristics, traffic control devices, and the environment. It also includes assessing the changes in collision risk associated with each of these factors and interactions. This information will support the development of new and improved countermeasures with greater effectiveness.[4]

In the SHRP2 NDS, over 3,400 volunteer drivers had their vehicles fitted with cameras, radar, and other sensors to capture data as they went about their usual driving tasks.[3,4] Video images of the front and rear views, the passenger side view, the driver's face and hands, and the cabin were recorded. Additionally, rates of acceleration, lateral and vertical motion, the presence of alcohol within the cabin, position information, turn-signal actuation, and other variables, such as steering wheel angle, speed, seatbelt use, and airbag deployment, were recorded through various sensors.[4]

Six field study sites were selected: Tampa, FL; Bloomington, IN; Durham, NC; Buffalo, NY; Central Pennsylvania; and Seattle, WA. Two additional projects gathered the roadway information necessary to relate drivers' actions to the roadway characteristics.[13] The resultant RID was populated with roadway and other contextual information from the six NDS field study sites. The relationship between the NDS and RID is shown conceptually in figure 1.[3,1]

SHRP2 NDS Database and RID

NDS
- Approx. 1,950 Data Acquisition Systems.
- Approx. 3,500 participants.
- Approx. 5 million trips.
- Approx. 50 million mi traveled.
- Recorded continuously.
- Passenger cars, vans, SUVs, pickup trucks.

RID
- New data collected and QA:
  - 12,500 centerline mi.
- Consistent among six States.
- Acquired data (State transportation departments, others):
  - 200,000 centerline mi.
  - Roadway, weather, traffic.

Source: FHWA.
SUV = sport utility vehicle; QA = quality assured.

Figure 1. Illustration. SHRP2 NDS database and RID.[3]

RID OVERVIEW

The SHRP2 Safety Research Program produced the RID through Project S04, which consisted of the following two projects:[1,7]

- Project S04A (Roadway Information Database Development and Technical Coordination and Quality Assurance of the Mobile Data Collection (Project S04B)).

- Project S04B (Mobile Data Collection).

The RID is a database of roadway characteristics that can be linked to the SHRP2 NDS database.[3] As noted by Campbell in *TR News*, "The RID combines roadway data from the State highway departments and other sources with data collected by mobile vans, which measure roadway characteristics while traveling at posted speed limits on routes selected by SHRP2."[13] All the databases, including van-collected data, are projected onto Esri® StreetMap Premium™ for ArcGIS® road network, which is based on a version of the HERE® road network.[5,9]

The roadway data collected via mobile vans for the SHRP2 project include "the number of lanes, lane type and width, grade, superelevation, beginning and end points of a curve, curve radius, lighting, rumble strips, median type, width of the paved shoulder, all signs and their locations, location of intersections, number of approaches, and traffic control devices."[13] These data were collected on selected roads within the six SHRP2 NDS sites (Florida, Indiana, North Carolina, New York, Pennsylvania, and Washington State).[3] A full list of roadway data elements collected and data formats used can be found on the SHRP2 website "Roadway Data Elements Collected."[14]

The RID also includes supplemental data—such as roadway inventories, crash histories, traffic, weather, work zones, 511 information, and safety enforcement laws—acquired from Government, public, and private sources.[15]

Figure 2 summarizes the RID data, including lists of data elements collected from mobile-van data, acquired roadway data, and acquired supplemental data.

RID
Geospatial database to manage and access disparate datasets

Mobile-Van Data	Acquired Roadway Data	Acquired Supplemental Data

Mobile-Van Data

- New data collected for SHRP2.
- Quality assured.
- Approx. 25,000 driven mi/12,500 centerline mi across the 6 NDS sites (FL: 4, 366 mi; IN: 4,635 mi; NC: 4,558 mi; NY: 3,570 mi; PA: 3,670 mi; WA: 4,277 mi.
- Includes:
  - Horizontal curvature (radius, length, PC, PT, direction).
  - Grade.
  - Cross slope.
  - Lanes (number, width, and type).
  - Shoulder type/curb and paved width.
  - Intersection location, number of approaches, and control (including ramp termini).
  - All MUTCD signs.
  - Barriers.
  - Median presence and type.
  - Rumble strip presence/location.
  - Lighting presence.

Acquired Roadway Data

Existing roadway inventory data acquired from agencies, e.g., State transportation departments from the 6 NDS sites (data items vary by State).
- Approx. 200,000 centerline mi.
- Includes HPMS files for the 6 NDS sites, plus:
  - Functional classification.
  - Signals.
  - Intersections.
  - Access control.
  - Pavement condition.
  - Bridge location.
  - Vertical alignment.
  - Interchanges.
  - Rest areas.
  - Terrain.
  - Tunnels.
  - FRA grade crossings.

All mobile-van and acquired roadway data are referenced to a common base map.

Acquired Supplemental Data

Existing data and information from State transportation departments, public agencies, and private sources, including:

- Approx. 200,000 centerline mi.
- Crash history data.
- AADT.
- Traffic data–continuous counts.
- Aerial imagery.
- Speed limit data and laws.
- Cell phone and text messaging laws.
- Automated enforcement laws.
- Graduated driver licensing laws.
- State motorcycle helmet use laws.
- Seatbelt use laws.
- Local climatological data.
- Winter road conditions.
- Work zones.
- 511 information.

Source: FHWA.

Approx. = approximately; PC = point of curve; PT = point of tangent.

Figure 2. Illustration. SHRP2 RID overview.

SHRP2 PROJECT S04B: MOBILE DATA COLLECTION

Under SHRP2 Project S04B, roadway data were collected at highway speed using Automatic Road Analyzer vehicles on selected roads within the six SHRP2 NDS sites.[7,3]

The objective of the mobile-data-collection project was to collect high-quality data on those roads most frequently driven by NDS participants and on roads of greatest interest to safety researchers.[3] The mobile-data-collection effort included the following:[1]

- 25,076 total mi of roadway data collected (12,500 plus centerline mi).
- Total number of assets, including grade and cross-slope points: 7,524,310. These assets are as follows:
    - 44,114 curves.
    - 6,129 mi of barriers.
    - 43,195 intersections.
    - 7,376 mi of highway lighting.
    - 10,756 mi of medians.
    - 33,013 mi of paved and unpaved shoulders (including inside and outside shoulders on four-lane divided roads).
    - 11,852 mi of rumble strips.
    - 518,570 total signs with *Manual on Uniform Traffic Control Devices* classification.[16]
    - Video log.

RID STRUCTURE

The RID is maintained as a set of geodatabases—one geodatabase for each of the six study-area States, plus a seventh containing roadway-centerline data for adjacent States.

Each roadway-data item included in the RID geodatabases is represented as a linear or point feature with a geographic position relative to all other roadway features in the geodatabase specified by a linear measure along a specific route. Every road segment in each study area is associated with a uniquely defined route, and the collection of all routes and their linear measures comprises a linear referencing system (LRS) for the study area.

The LRS Route feature class represents the common framework to which all other road inventory features are attached. Individual routes are constructed by stringing together roadway-centerline segments that share a common route number or road name.

RID DATA ELEMENTS

The following are sources of the final RID geodatabases for the six States and their data elements (the asterisk indicates conflated base-map routes):

- Esri® roadway-centerline geometry and attribute data (used for base-map routes).[5]
- Mobile-van data (i.e., data collected by the SHRP2 contractor).*
- Highway Performance Monitoring System (HPMS) data.[8]*
- Supplemental State roadway data (phase 1).*
- Supplemental roadway data (phase 2).*
- Additional supplemental data.

The supplemental State roadway data vary significantly from State to State and include both new data items and HPMS data items that cover roadway segments beyond those required for HPMS submittal.[8]

The RID developer added two sets of State-provided data to the RID. The first set was conflated with the LRSs and became part of the geodatabases (referred to here as supplemental roadway data (phase 2)). The components of this set of data differ by State; multiyear crash and traffic data as well as 511 data are the main components of these datasets.

The additional supplemental data have different formats and contain mainly aerial imagery, crash data documentation and descriptions, weather data, State laws regarding speed, and safety campaign data. As with the phase 2 data, the contents differ from State to State. However, these additional supplemental data are not conflated with the base-map routes.

NPMRDS OVERVIEW

In July 2013, the Federal Highway Administration (FHWA) acquired a national TT dataset from HERE® to support Federal mobility applications.[9] The NPMRDS contains passenger and freight TT datasets for the National Highway System (NHS) and other roadways.[2,10] NPMRDS TT data are gathered through traffic sensors, Global Positioning System (GPS)–probe devices, historical databases, and incident reports. Some of the data are distributed back into the traffic operations system in real time to help inform road users or to direct traffic when needed.

NPMRDS data are made available to States and metropolitan planning organizations (MPOs) to "use for their performance management activities."[2] The FHWA Office of Operations controls access to the NPMRDS database.

NPMRDS data have been available since July 2013 and are updated monthly. For 2012, the TT data are also available for all interstate highways. The period for which both RID crash data and NPMRDS data are available is July through December 2013.[1] NPMRDS coverage overlaps with all six RID States (Florida, Indiana, New York, North Carolina, Pennsylvania, and Washington).

DATA COVERAGE

The NPMRDS includes three categories of average TTs obtained from vehicle-probe data: all vehicles, passenger vehicles, and freight trucks. The data are collected for both freight and passenger traffic for the following road classifications (representing approximately 220,000 centerline mi of road):

- U.S. Interstate System.

- The NHS, including NHS intermodal connectors.[10]

- Roadway facilities classified as principal arterials not included in the current NHS.[10]

- Strategic Defense Highway Network roadways.

- Border crossings on principal arterials.

- Minor arterials within a 5-mi radius on either side of the border at the top 20 United States/Canada border crossings as measured by average daily truck trips.

- Minor Arterials within a 5-mi radius on either side of the border at the top six United States/Mexico border crossings as measured by average daily truck trips.

The NPMRDS TT data are distributed in monthly releases. In each monthly release, data for all 50 States; Washington, DC; and Puerto Rico are divided into the following regions:

- Northeast.
- South.
- Midwest.
- West.
- Borders (roadways that border Mexico and Canada).
- United States (i.e., the entire United States).

The monthly release includes TT data for all of these regions.

NPMRDS DATA COMPONENTS

The NPMRDS consists of several components, including the following:

- TMC static file.
- TT data.
- NPMRDS Links:
    - NPMRDS Links shapefile.
    - NPMRDS Links to TMC look-up table.
- NPMRDS supporting documentation.

Figure 3 shows a high-level layout of the NPMRDS data components.
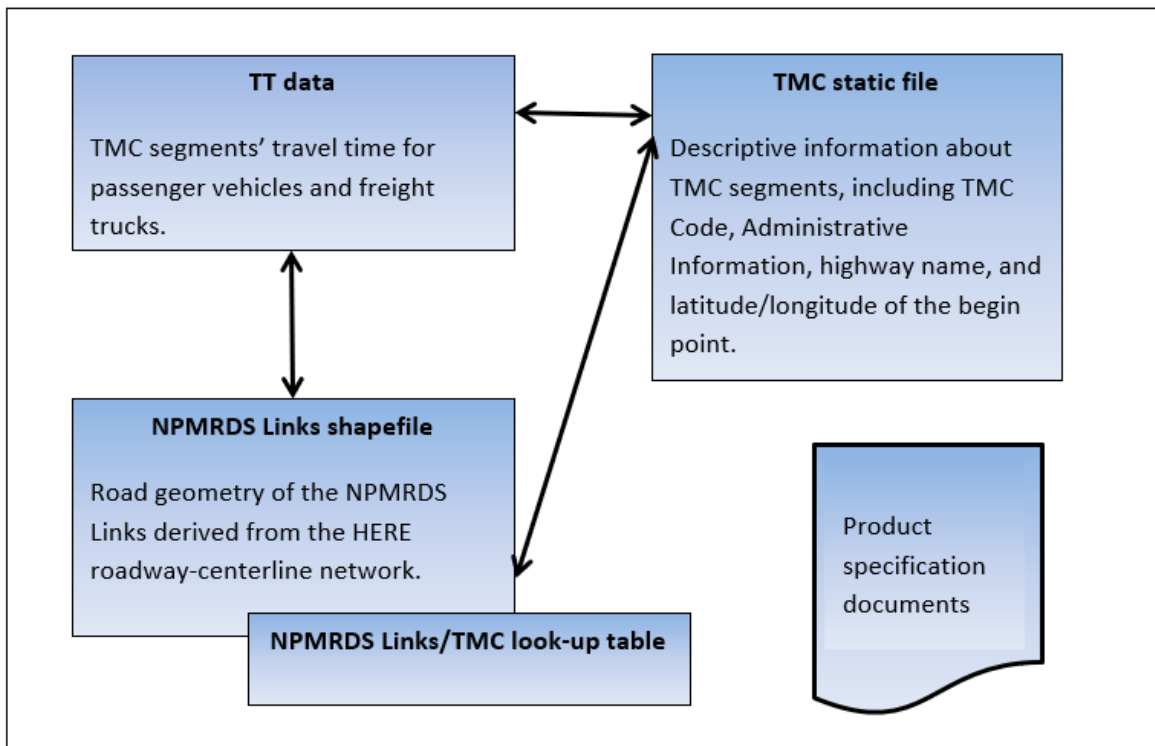
Source: FHWA.

Figure 3. Illustration. General layout of the NPMRDS data.

TMC

Each roadway segment in the NPMRDS is defined by a unique identifier known as the TMC code. TMC codes were developed to provide an efficient way of transmitting location information that is independent of the accuracy or level of resolution of the map on which the data are displayed.

A TMC segment corresponds to a section of roadway between two well-defined end points in a specific direction of travel. On limited-access divided highways, roadway segments typically go from one interchange to the next in each travel direction (e.g., I-95 northbound from exit 39 to exit 40). On surface streets, roadway segments typically go between major intersections (e.g., Main Street between First and Second Avenues). TMC lengths vary from a fraction of a mile to several miles, but most are between 0.5 and 2.0 mi.

TMC STATIC FILE

A static file (comma-separated value (csv) format) contains descriptive information about each TMC segment in the coverage area. The fields in the static file are described in the data-specification document; they do not change from release to release and remain the same unless there is a need to update. Some examples of a need to update the static file include when a road name changes or the road geometry moves (latitude/longitude). Figure 4 and table 1 show a view of the static file in spreadsheet format and the static file record layout, respectively.

	TMC	ADMIN_LEVEL_1	ADMIN_LEVEL_2	ADMIN_LEVEL_3	DISTANCE	ROAD_NUMBER	ROAD_NAME	LATITUDE	LONGITUDE	ROAD_DIRECTION
1										
246820	114N04098	USA	Washington	King	2.79553	I-90		47.44332	-121.66814	Eastbound
246821	114N04099	USA	Washington	King	2.09622	I-90		47.46375	-121.70895	Eastbound
246822	114N04100	USA	Washington	King	2.34013	I-90		47.47366	-121.75273	Eastbound
246823	114N04101	USA	Washington	King	3.49941	I-90		47.48522	-121.79303	Eastbound
246824	114N04102	USA	Washington	King	1.19326	I-90		47.51358	-121.84879	Eastbound
246825	114N04103	USA	Washington	King	3.32853	I-90		47.50981	-121.87739	Eastbound
246826	114N04104	USA	Washington	King	2.3423	I-90		47.52416	-121.93681	Eastbound
246827	114N04105	USA	Washington	King	2.23801	I-90		47.53404	-121.97587	Eastbound
246828	114N04106	USA	Washington	King	0.45077	I-90		47.53421	-122.01737	Eastbound
246829	114N04107	USA	Washington	King	1.58446	I-90		47.53821	-122.03335	Eastbound
246830	114N04108	USA	Washington	King	2.02886	I-90		47.54626	-122.05488	Eastbound
246831	114N04109	USA	Washington	King	1.94493	I-90		47.56303	-122.09491	Eastbound
246832	114N04110	USA	Washington	King	0.34599	I-90		47.57554	-122.12425	Eastbound
246833	114N04111	USA	Washington	King	1.70535	I-90		47.57838	-122.13255	Eastbound
246834	114N04112	USA	Washington	King	0.93497	I-90		47.58027	-122.16522	Eastbound
246835	114N04113	USA	Washington	King	0.66352	I-90		47.58004	-122.18127	Eastbound
246836	114N04114	USA	Washington	King	1.2156	I-90		47.57861	-122.20205	Eastbound
246837	114N04115	USA	Washington	King	0.51238	I-90		47.58488	-122.22516	Eastbound
246838	114N04116	USA	Washington	King	0.80892	I-90		47.58894	-122.2345	Eastbound
246839	114N04117	USA	Washington	King	1.72565	I-90		47.59041	-122.25029	Eastbound
246840	114N04118	USA	Washington	King	0.0776	I-90		47.59053	-122.30244	Eastbound
246841	114N04119	USA	Washington	King	0.70089	I-90		47.59055	-122.30434	Eastbound
246842	114N04120	USA	Washington	King	0.75295	I-90		47.59509	-122.31499	Eastbound
246843	114N04122	USA	Washington	King	0.29324	I-405		47.46199	-122.26784	Southbound
246844	114N04123	USA	Washington	King	0.40062	I-405		47.46261	-122.25837	Southbound

Source: FHWA.
ADMIN = administrative; USA = United States of America.

Figure 4. Screen shot. NPMRDS static file in spreadsheet format.[2]

Table 1. NPMRDS static file record layout.

Field Name	Type	Example	Description
TMC	Text	D01N04474	Traffic Location code in the format of CLLDTTTTT Where: C is the Country Code (1 digit). LL is the Location Table Number (2 digits). D is the TMC path direction (P or N). P refers to positive direction of the TMC. N refers to negative direction of the TMC. TTTTT is the TMC location code (5 digits). If no data exist for a TMC for an epoch, there will be no entry in the data file for that combination of TMC/day of week/epoch.
ADMIN_LEVEL_1	Text	United States	The Country where the listed Traffic Location Code is located
ADMIN_LEVEL_2	Text	Illinois	The State/Province where the listed Traffic Location Code is located
ADMIN_LEVEL_3	Text	DuPage	The County where the listed Traffic Location Code is located
DISTANCE	Float	3.27285	The length of the TMC, measured in miles to 5 decimal places
ROAD_NUMBER	Text	I-90	Road Number taken from Location Table. When a Road has both a route number and local name, the route number is placed in the Road Number column and the local name in the Road Name column.
ROAD_NAME	Text	SOUTH ARCHER AVE.	Road Name taken from Location Table. When a Road has both a route number and local name, the route number is placed in the Road Number column and the local name in the Road Name column.
LATITUDE	Float	37.87472	World Geodetic System 1984 (WGS84) coordinate to five decimal places Note: Latitudes and longitudes provided will be for the beginning of the TMC.
LONGITUDE	Float	−122.18671	WGS84 coordinate to 5 decimal places Note: Latitudes and longitudes provided will be for the beginning of the TMC.
ROAD_DIRECTION	Text	NORTHBOUND	Road direction represents the direction of travel based on the road sign.

TT DATA

TTs are defined for TMC segments. The TT data represent the average time each vehicle type (i.e., passenger vehicle, truck, or average of all vehicles) takes to traverse a specific roadway segment on a specific day and time period (e.g., Friday, January 30, 8:00 to 8:05 a.m.). Each day is divided into 288 time periods, or epochs, of 5 min each (e.g., epoch 0 represents the time period from 12:00 to 12:05 a.m.).

Figure 5 and table 2 show a sample of the TT data in spreadsheet format and the TT data record layout.

	A	B	C	D	E	F
				Travel Time Data File		
1	TMC	DATE	EPOCH	Travel_TIME_ALL_VEHICLES	Travel_TIME_PASSENGER_VEHICLES	Travel_TIME_FREIGHT_TRUCKS
2	114N04104	7262013	276	130	130	
3	114N04104	7272013	59	133		133
4	114N04104	7272013	90	134	127	170
5	114N04104	7272013	183	119	119	
6	114N04104	7282013	152	135	135	
7	114N04104	7282013	183	123	123	
8	114N04104	7292013	28	130		130
9	114N04104	7262013	121	140	138	141
10	114N04104	7292013	59	133	133	
11	114N04104	7292013	90	146	146	
12	114N04104	7292013	214	119	119	
13	114N04104	7292013	245	144		144
14	114N04104	7302013	28	146	146	
15	114N04104	7302013	59	144	144	
16	114N04104	7302013	90	134	134	
17	114N04104	7302013	121	144	144	
18	114N04104	7302013	183	119	116	138
19	114N04104	7302013	245	119	119	
20	114N04104	7302013	276	152	152	
21	114N04104	7312013	59	141		141
22	114N04104	7312013	121	133	133	
23	114N04104	7312013	183	134	134	
24	114N04104	8012013	114	135	135	

Source: FHWA.

Figure 5. Screen shot. NPMRDS TT data in spreadsheet format.[2]

Table 2. NPMRDS TT data record layout.

Field Name	Type	Example	Description
TMC	Text	D01N04474	Traffic Location code in the format of CLLDTTTTT, where: C is the Country Code (1 digit). LL is the Location Table Number (2 digits). D is the TMC path direction (P or N). P refers to positive direction of the TMC. N refers to negative direction of the TMC. TTTTT is the TMC location code (5 digits). If no data exist for a TMC for an epoch, there will be no entry in the data file for that combination of TMC/day of week/epoch.
DATE	Text	04022012	Day Month Year (DDMMYYYY)
EPOCH	Integer	48	Value from 0 through 287 that defines the 5-min period the average speed applies (local time), where: 0 = 00:00:00 to 00:04:59. 1 = 00:05:00 to 00:09:59. 2 = 00:10:00 to 00:14:59. … 287 = 23:55:00 to 23:59:59.
Travel_TIME_ALL_ VEHICLES	Integer	44	In seconds TTs are calculated as the ratio of the segment length to the average speed on the segment. Average segment speed is determined from a combination of the passengers' and freight trucks' individual GPS-probe speed observations.
Travel_TIME_ PASSENGER_ VEHICLES	Integer	76	In seconds TTs are calculated as the ratio of the segment length to the average speed. Average segment speed is determined from only passengers' individual GPS-probe speed observations.
Travel_TIME_ FREIGHT_TRUCKS	Integer	1	In seconds TTs are calculated as the ratio of the segment length to the average speed. Average segment speed is determined from only freight trucks' individual GPS-probe speed observations.

NPMRDS LINKS

In addition to the TT and TMC static file data, the following two layers of information are published for the NPMRDS:

- **NPMRDS Links shapefile:** This geospatial database contains the geometric alignments of the NHS–NPMRDS Links derived from the HERE® roadway-centerline network (i.e., the road geometry for the highways covered in the database).[10,9] The database is provided in shapefile (shp) format—a common exchange format for geospatial data. The NPMRDS Links and their geometric alignments are very similar to, but do not match exactly, the roadway-centerline segments in the RID. Based on an initial investigation, about 90 percent of the NPMRDS Links are identical to the RID Links. The majority of the rest also match the RID begin and end points but have slight differences in their alignments. Only a small portion are completely different from the RID Links.

- **NPMRDS Links–TMC look-up table:** The look-up table (database file (dbf) format) relates the NPMRDS Links to TMC segments. A Link may be associated with more than one TMC, and a TMC may be associated with more than one Link. The dbf look-up table, therefore, provides the "crosswalk" that indicates which TMC codes should be assigned to each shapefile link.

Table 3 lists attributes that are published for these two layers.

Table 3. NPMRDS attributes.

Field Name	Name	File Format
FEATURE_ID	Feature ID	.shp
SHAPE	Shape	.shp
LINK_ID	Link ID	.dbf
ST_NAME	Street Name	.shp
ROUTE_TYPE	Route Type	.shp
DIR_TRAVEL	Direction of Travel	.shp
FRONTAGE	Frontage	.shp
RAMP	Ramp	.shp
CONTROLLED_ACCESS	Controlled Access	.shp
ISO_COUNTRY_CODE	Country	.shp
DISTANCE	Distance	.shp
TMC	TMC	.dbf

ID = identification; ISO = International Organization for Standardization; .shp = shapefile.

Figure 6 and figure 7 show these two layers in ArcGIS® and in a database format, respectively.[17] Some fields' names and the order in which they appear in the shapefile are slightly different from what is in the documentation shown in table 3. Detailed descriptions of the fields and possible values are available in the shapefile document.

Source: Screen shot created by FHWA using ArcGIS® software by Esri®. ArcGIS® is the intellectual property of Esri® and is used herein under license.

FID = Feature identification; ID = identification; ST = street; DIR = direction; ISO = International Organization for Standardization; RD = road; RT = route; B = both; N = negative.

Figure 6. Screen shot. NPMRDS shapefile in ArcGIS®.[16]

Source: FHWA.

XLINK_PVID = link number; DIR = direction; F = from; T = to.

Figure 7. Screen shot. NPMRDS Links to TMC look-up table.[2]

OBTAINING THE NPMRDS DATA

NPMRDS data are updated monthly and made available to States and MPOs to "use for their performance management activities."[2] The FHWA Office of Operations controls access to the NPMRDS database. FHWA offices, States, and MPOs, as well as their contracted programs, may request access to the data download site.

The RID data collection period generally covers November 2011 to March 2013 with small variance among the six RID States (Florida, Indiana, North Carolina, New York, Pennsylvania, and Washington). With the exception of Florida, crash-history data are available for 2013. The period for which both RID crash data and NPMRDS data are available is July through December 2013. RID data features vary among the six States; Washington data were used for the case study.

The project team contacted the FHWA Office of Operations and HERE® to obtain access to the download site and then downloaded the necessary data for the case study.[9] The databases for July through December 2013 for the West region, which includes Washington State data, were obtained.

PLAN FOR CONDUCTING THE NPMRDS-RID DATA LINKAGE

The initial plan for establishing data linkage included the following steps:

1. Compare the NPMRDS and RID Links.[2,1]

2. Match the NPMRDS Links to the RID Links and create a new feature class.

3. Create a new feature class for TMC.

4. Attach TMC identifiers to each link in the new feature class.

5. Define a process to attach TT data associated with TMCs to the new feature class for different epochs.

However, after advancing in the linkage process, it became apparent that, due to the nature of the TMCs, creating a new feature class for TMCs was not necessary for achieving the project objective. In addition, creating a new feature class requires significant programming outside ArcGIS®.[17] As a substitute, a new geodatabase was built that contains the desired feature classes from the RID as well as link TTs from the NPMRDS. The major steps of the process were then revised to the following:

1. Perform the initial setup (including the creation of the new geodatabase).
2. Identify common links between the NPMRDS and RID.
3. Import the NPMRDS Links–TMC look-up table.
4. Identify appropriate TMC records.
5. Identify the directional relationship between TMCs and links.
6. Refine/limit common links and independently manage directional data.
7. Assign TMCs to limited common links.
8. Import TT data.
9. Calculate weekday free-flow speed.
10. Combine TT and speed data with linkage data.

The following sections describe these steps in detail.

NPMRDS AND RID NETWORKS

The NPMRDS and RID networks are compatible in node coordinates and geospatial characteristics. However, not all Link IDs match. For this study, only links with the same Link IDs in both databases are used in the process.

RID FEATURE CLASSES AND NPMRDS DATA FILES

The following feature classes from the original RID (WA_DataWarehouse.gdb (.gdb = geodatabase) version 1.1 (v1.1)) and the RID reduced datasets are used in the linkage process:

- Links.
- Routes.
- Alignment.
- SpeedLimit.

The following NPMRDS data files are used in the process:

- NPMRDS shapefile: NHS_NPMRDS_Shape_file_HERE_Q2_2013.shp.

- NPMRDS and TMC look-up table: NPMRDS_TMC_LUT_2013Q2.dbf.

- NPMRDS TT data for July through December 2013:
  FHWA_TASK2_4_WA_MM_2013_TT, where MM = 07 to 12, representing the months
  July to December.

For this study, the NPMRDS TMC TTs were limited to the period of July through December 2013. The TT data file, FHWA_TASK2_4_WA_MM_2013_TT, carries these data. Parts of the process conducted on these data can be repeated for other periods if needed.

NPMRDS-RID LINKAGE PROCESS

The process described in this document is for a specific example in which estimated free-flow V_{85} and V_{avg} are derived from TTs extracted from the NPMRDS for links matched with the RID.[2,1] The objective of this example is to compute the actual free-flow speeds on links in the Washington State RID using TT data from the NPMRDS.

There were 10 distinct steps in the linkage process. The main goal of the process was to build a geodatabase that contains common links between the RID and the NPMRDS accompanied by TTs, speeds, and speed limits. ArcGIS® ArcMap™ 10.1 was the primary tool employed in the linkage process.[16] Other tools, such as Microsoft® Excel™, were also utilized. The final product of these steps was a geodatabase containing selected RID and NPMRDS feature classes as well as some intermediate feature classes and tables. Figure 8 shows the final geodatabase list of feature classes and tables.

New_RID_NPMRDS_Linkage.gdb
- Crash_2011_2013_Reduced
- Crash_2011_2013_Reduced_SR_Int_Art_30ft
- Crash2011_LRS_Reduced
- Crash2011_Other_Reduced
- Crash2012_LRS_Reduced
- Crash2012_Other_Reduced
- Crash2013_LRS_Reduced
- Crash2013_Other_Reduced
- Limited_LinkTMCRef
- Limited_LinkTMCRef_Negative
- Limited_LinkTMCRef_Positive
- LimitedCommonLinks_Negative
- LimitedCommonLinks_Positive
- LimitedCommonLinksTMCs_Negative
- LimitedCommonLinksTMCs_Positive
- Links_N
- Links_Negative_Speeds
- Links_P
- Links_positive_Speeds
- LinkTMCRef
- Negative_Cts_GT50_TT
- Negative_Cts_GT50_TT_Speeds
- Negative_GT50_Cts_TT
- Negative_GT50_Cts_TT_Speeds
- Negative_GT50_CtsTT__Speeds_Fixed
- Negative_GT50_TT
- Negative_GT50_TT_Speeds
- Positive_Cts_GT50_TT
- Positive_Cts_GT50_TT_Speeds
- Static_File
- Static_TMC_Count
- UniqueLinks
- UniqueMissingLinks
- WA_07_12_2013_Weekday_Freeflow_TT_Speed
- WA_Alignment
- WA_FunctionalClass_SR
- WA_Lane
- WA_Links
- WA_Location
- WA_NHS_NPMRDS_Links
- WA_Routes
- WA_Routes_One_Way
- WA_Sections
- WA_SpeedLimit_Reduced
- WA_Tops
- WA_Tops_Div_Urban_Int_Art
- WA_Tops_Div_Urban_Int_Art_Reduced

Source: Screen shot created by FHWA using ArcMap™ software by Esri®. ArcMap™ is the intellectual property of Esri® and is used herein under license.

Figure 8. Screen shot. New_RID_NPMRDS_Linkage geodatabase.[17]

Each step, including details about specific files and actions, is explained in the following sections.

STEP 1: PERFORM INITIAL SETUP

The initial setup consists of the following steps:

- Open a new ArcGIS® map document.[17]
- Add selected RID and NPMRDS feature classes into the map.
- Create a new geodatabase.
- Import feature classes in the new geodatabase with different names.

Input

- Washington RID Alignment feature class.
- Washington RID Links feature class.
- Washington RID Routes feature class.
- Washington RID SpeedLimit feature class.
- NPMRDS shapefile NHS_NPMRDS_Shape_file_HERE_Q2_2013.shp.

Output

- NPMRDS_Linkage.mxd map.
- RID_NPMRDS_Linkage geodatabase.
- WA_NHS_NPMRDS_Links feature class.
- WA_Alignment feature class.
- WA_Routes feature class.
- WA_Links feature class.
- WA_SpeedLimit_Reduced feature class.

Description

Open a new map in geographic information system (GIS)/mapping software and name it NPMRDS_Linkage; add all four feature classes of the RID (Alignment, Routes, Links, and SpeedLimit_Reduced) and the NPMRDS shapefile to the map. The NPMRDS shapefile includes links for the entire United States. Conduct a Select by Location operation on the links from the shapefile within 100 ft of the links in the RID Links feature class chosen as the source-layer features. The objective of this step is to limit the nationwide data to only those links proximate to RID data. Out of 1,792,650 links, 34,028 are selected, and the links are saved as a new feature class named WA_NHS_NPMRDS_Links.

Create a new geodatabase named RID_NPMRDS_Linkage in the ArcMap™ Catalog window.[17] Import the RID feature classes as well as the WA_NHS_NPMRDS_Links feature class (figure 9) into the new geodatabase. Choose WA_Alignment, WA_Routes, WA_Links, and WA_SpeedLimit_Reduced as the names of the imported RID feature classes in the new geodatabase.
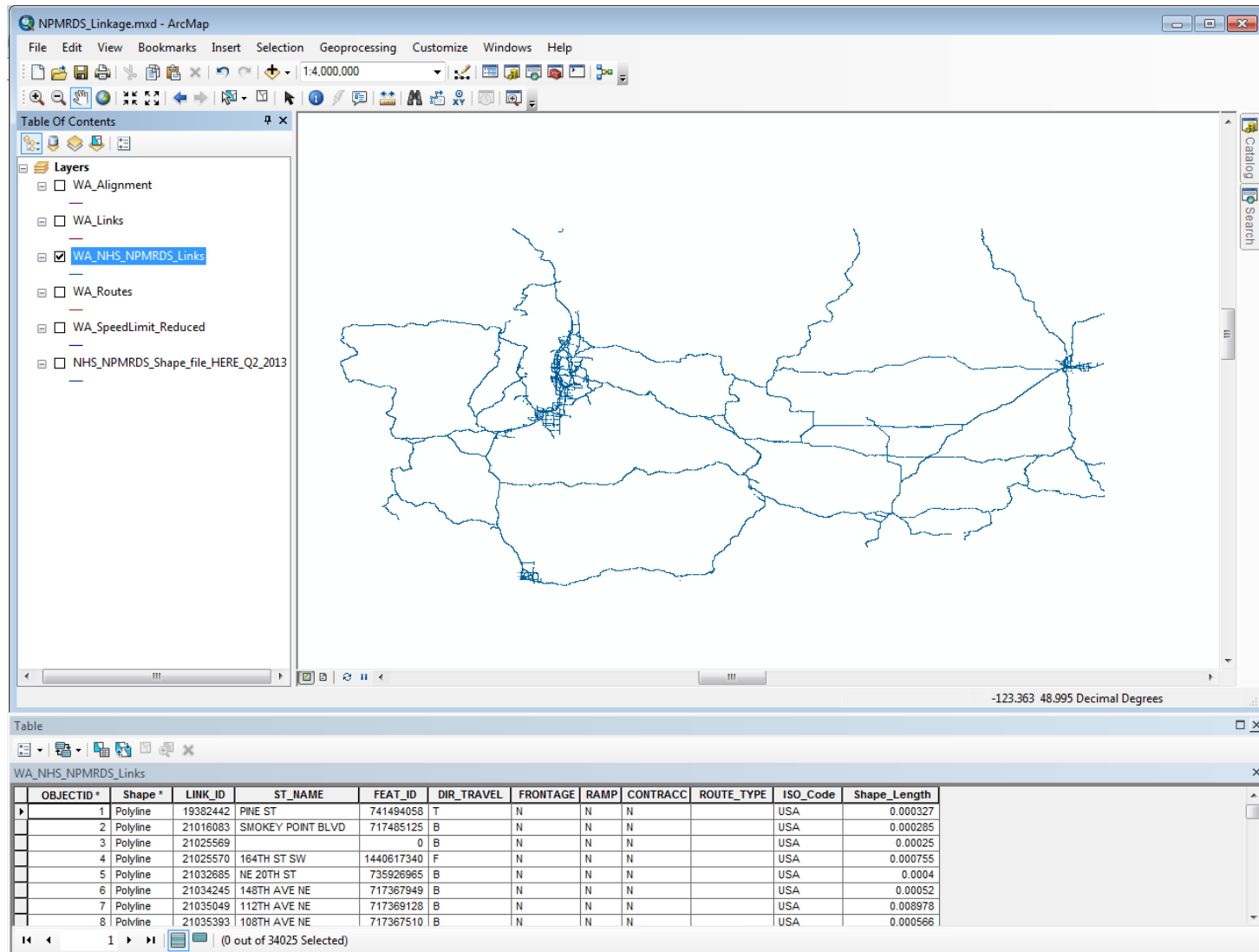
Source: Screen shot created by FHWA using ArcMap™ software by Esri®. ArcMap™ is the intellectual property of Esri® and is used herein under license. StreetMap data source: Tele Atlas.

Figure 9. Screen shot. WA_NHS_NPMRDS_Links feature class represented in ArcMap™.[(17)]

STEP 2: IDENTIFY COMMON LINKS BETWEEN THE NPMRDS AND RID

Join the RID and NPMRDS Links together and identify common links. In this step, records of matching links are saved as a separate feature class.

Input

- WA_Links feature class (output from step 1).
- WA_NHS_NPMRDS_Links feature class (output from step 1).

Output

- WA_NHS_NPMRDS_Links feature class (updated to include RID Links).
- JoinedLinks feature class.
- CommonLinks feature class.
- MissingLinks feature class.
- UniqueLinks table.
- UniqueMissingLinks table.

Description

Use ArcGIS® to join the RID Links feature class (WA_Links) to the NPMRDS Links from Washington State (WA_NHS_NPMRDS_Links).[16] Use the RID feature LINKID and the NPMRDS feature LINK_ID (from WA_NHS_NPMRDS_Links) in this operation with the option of Keeping All Target Features. In this process, the links from the first feature class (WA_Links) that do not match any link in the second feature class (from WA_NHS_NPMRDS_Links) are assigned values of NULL for all attributes of the second feature class, including LinkID.

Duplicate the WA_NHS_NPMRDS_Links in the new geodatabase as a new feature class and rename it JoinedLinks (34,028 features; see figure 10). Select features of the joined feature class with LINKID not equal to NULL (WA_Links.LINKID IS NOT NULL) and save it as a new feature class in the new geodatabase named CommonLinks (29,420 features (figure 11)). Then, use Switch Selection and save the result as a new feature class in the new geodatabase named MissingLinks (4,608 features).

Open the CommonLinks feature class and use the Summarize option to create a dBASE table named UniqueLinks by summarizing LINK_ID ( ). This table has all links and the number of occurrences of each link in the original table (all are 1, so no link is repeated).

Open the MissingLinks feature class and use the Summarize option to create a dBASE table named UniqueMissingLinks by summarizing LINK_ID ( ). This table has all links and the number of times each link appears in the original table (all are 1, so no link is repeated).

Source: Screen shot created by FHWA using ArcMap™ software by Esri®. ArcMap™ is the intellectual property of Esri® and is used herein under license. StreetMap data source: Tele Atlas.

Figure 10. Screen shot. JoinedLinks feature class represented in ArcMap™.[17]
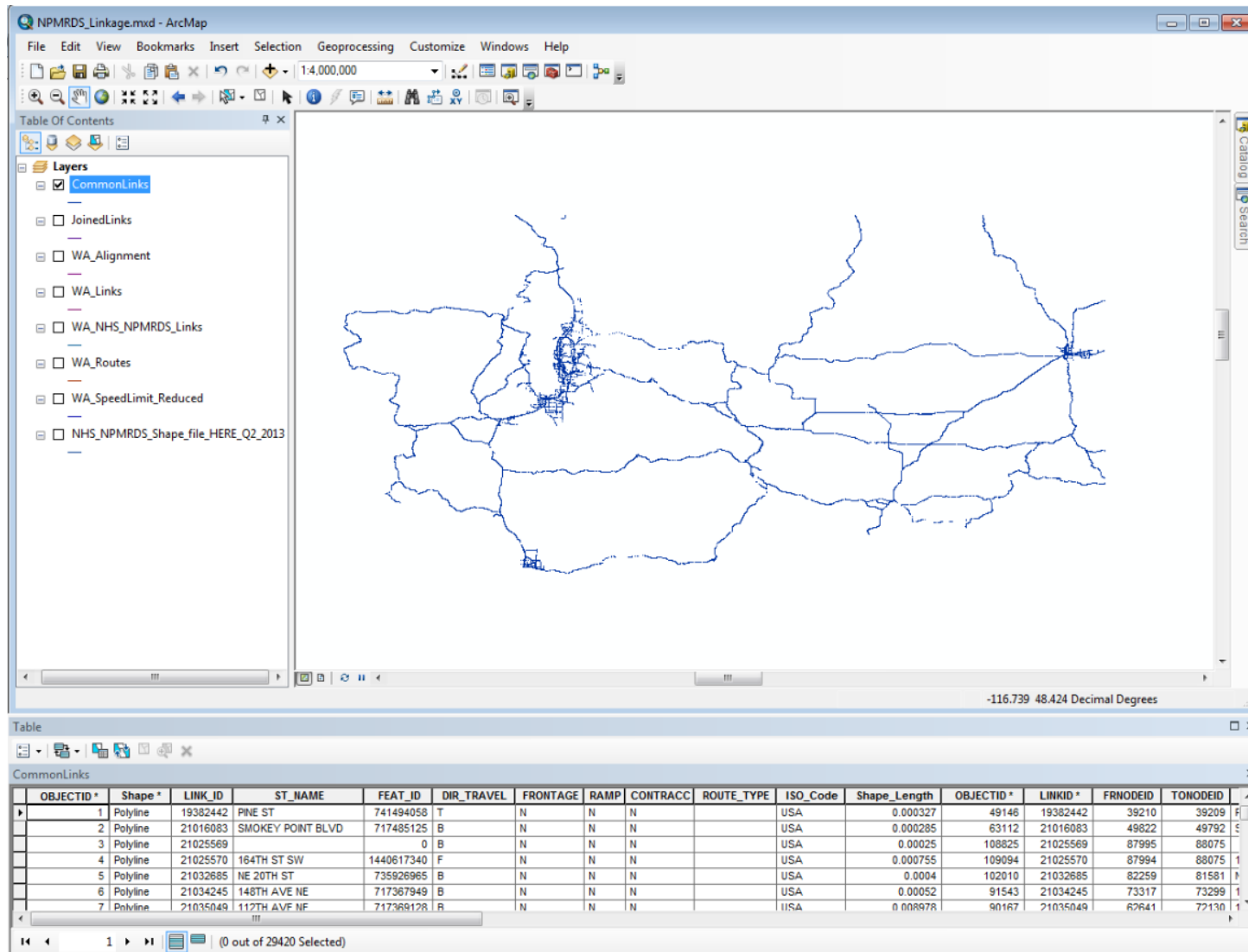
Source: Screen shot created by FHWA using ArcMap™ software by Esri®. ArcMap™ is the intellectual property of Esri® and is used herein under license. StreetMap data source: Tele Atlas.

Figure 11. Screen shot. CommonLinks feature class represented in ArcMap™.[(17)]

STEP 3: IMPORT NPMRDS LINKS–TMC LOOK-UP TABLE

Each TMC is associated with one or more NPMRDS Links. The NPMRDS Links–TMC look-up table (NPMRDS_TMC_LUT_2013Q2.dbf) shows the relations (figure 7). Import this table into the linkage geodatabase (LinkTMCRef).

Input

- NPMRDS_TMC_LUT_2013Q2.dbf.

Output

- LinkTMCRef table.

Description

Import NPMRDS_TMC_LUT_2013Q2.dbf into the new geodatabase through the ArcGIS® Catalog import option and rename it LinkTMCRef (2,609,031 rows (figure 12)).[17]

Source: Screen shot created by FHWA using ArcMap™ software by Esri®. ArcMap™ is the intellectual property of Esri® and is used herein under license.

Figure 12. Screen shot. LinkTMCRef table in ArcMap™.[(17)]

STEP 4: IDENTIFY APPROPRIATE TMC RECORDS (I.E., IDENTIFY TMCS THAT HAVE ALL THEIR LINKS MATCHED TO RID LINKS)

The relationship between TMCs and NPMRDS Links is many-to-many, meaning that one TMC may include one or more links, and one link may appear on more than one TMC. Figure 13 shows a sample of this relationship where one TMC (114N12094) relates to 25 NPMRDS Links (XLINK_PVID column), and NPMRDS Link 19359405 relates to two TMCs (114N12094 and 114P12095). Also, not all NPMRDS Links are matched to RID Links. The main goal in this step is to identify TMCs that have all of their links matched to RID Links.

Input

- LinkTMCRef table (output from step 3).
- CommonLinks feature class (output from step 2).

Output

- Limited_LinkTMCRef table.

Description

Copy LinkTMCXRef (via Data Management Tools/General/Copy) and name the copy Limited_LinkTMCXRef. Join UniqueLinks (LINK_ID) to Limited_LinkTMCXRef (XLINK_PVID) and keep only the matching records (right-click on Limited_LinkTMCXRef and choose the "Joins and Relates" option; 48,406 records (figure 14)).

	A	B	C	D	E	F	G
1	XLINK_PVID	TMC	DIR		XLINK_PVID	TMC	DIR
2	19359405	114N12094	T		19359405	114N12094	T
3	21333187	114N12094	T		19359405	114P12095	F
4	21333205	114N12094	T				
5	21333249	114N12094	T				
6	21333318	114N12094	F				
7	21333593	114N12094	T				
8	21333627	114N12094	T				
9	21333770	114N12094	T				
10	21337218	114N12094	T				
11	21337231	114N12094	T				
12	21416568	114N12094	T				
13	31331205	114N12094	T				
14	31331209	114N12094	T				
15	31331210	114N12094	T				
16	124481226	114N12094	T				
17	124507656	114N12094	T				
18	124507657	114N12094	T				
19	124507672	114N12094	F				
20	124507673	114N12094	T				
21	124507751	114N12094	F				
22	124507752	114N12094	F				
23	124507753	114N12094	T				
24	124507755	114N12094	T				
25	757310395	114N12094	T				
26	757310396	114N12094	T				

Source: FHWA.

XLINK_PVID = link number; DIR = direction; T = to; F = from.

Figure 13. Screen shot. Sample of "many-to-many" relationship between NPMRDS Links and TMCs.

Source: Screen shot created by FHWA using ArcMap™ software by Esri®.
ArcMap™ is the intellectual property of Esri® and is used herein under license.

Figure 14. Screen shot. Sample of the Limited_LinkTMCRef table in ArcMap™.[17]

STEP 5: IDENTIFY DIRECTIONAL RELATIONSHIP BETWEEN TMCS AND LINKS

The fourth value in each TMC record is a letter representing the direction of the TMC. The values can be P (positive) or N (negative), where P indicates that the latitude/longitude values for that record refer to the beginning of the TMC, while N indicates that the latitude/longitude is related to the end of the TMC. There are also directional relationships between TMCs and all their corresponding links. These relations are carried out by the Dir attribute, with values of T (to), F (from), and B (both).

In this step, categorize the links into positive links and negative links. Positive links are those that correspond to one positive TMC (and no more than one negative TMC); negative links are those that correspond to one negative TMC. Eliminate the rest of the links (e.g., links with two or more positive TMCs, links with one positive TMC and more than one negative TMC, and links with no positive TMCs and more than one negative TMC).

Input

- Limited_LinkTMCRef table (output from step 4).

Output

- Limited_LinkTMCRef table (revised).
- Limited_LinkTMCRef.xlsx. (.xlsx = spreadsheet format file extension)
- Links_N table.
- Links_P table.
- LimitedCommonLinks_Negative feature class.
- LimitedCommonLinks_Positive feature class.

Description

Export Limited_LinkTMCXRef into dbf (by choosing dBASE table type) and add to the map. Delete the original Limited_LinkTMCXRef from the map. Add a field to the new Limited_LinkTMCXRef and name it TMC_Dir. Calculate values using the MID ( ) function (Mid([TMC],4,1)). This function returns P or N values showing the direction of the TMC. Figure 15 shows this table sorted by TMC.

Source: Screen shot created by FHWA using ArcMap™ software by Esri®.
ArcMap™ is the intellectual property of Esri® and is used herein under license.

Figure 15. Screen shot. Limited_LinkTMCXRef table.[17]

Import Limited_LinkTMCXRef (from dbf) to the new geodatabase (table to table). Export it as text, open using spreadsheet software, and save as a spreadsheet file (Limited_LinkTMCRef.xlsx). Create a pivot table (i.e., a tool that allows the user to reorganize and summarize selected columns and rows of data in a spreadsheet or database table to obtain a desired report) of Limited_ LinkTMCXRef in a spreadsheet using XLINK_PVID as rows, TMC_DIR as columns, and the count of any field as the values. Each row represents a unique XLINK_PVID. The three columns (N, P, and Grand Total) represent the number of occurrences of each XLINK_PVID and a TMC of the corresponding direction. Figure 16 shows a sample of this pivot table.

Source: FHWA.

Figure 16. Screen shot. Limited_LinkTMCXRef pivot table.

Copy and paste the pivot table (as values) to another Microsoft® Excel™ worksheet. This is referred to as LinkTMC_Dir. Add a new attribute named USE_DIR to LinkTMC_Dir. Populate USE_DIR with the TMC direction each link represents. In some cases, a link represents multiple TMCs in the same direction of travel. Remove these links from consideration. The following is the derivation protocol for USE_DIR:

- USE_DIR = N, where N = 1 and P is NULL.
- USE_DIR = P, where N is NULL and P = 1.
- USE_DIR = B, where N = 1 and P = 1.
- USE_DIR = NULL, where N > 1 or P > 1.

Select all records from the LinkTMC_Dir where USE_DIR = N or B. The results represent the links corresponding to only one TMC (in the negative direction). Throughout the remainder of this document, it will be referred to as LinkTMC_DirN. Import the LinkTMC_DirN table into the new geodatabase and name it Links_N.

Select all records where USE_DIR = P or B. The results represent the links corresponding to only one TMC (in the positive direction). Throughout the remainder of this document, it will be referred to as LinkTMC_DirP. Import LinkTMC_DirP into the new geodatabase and name it Links_P.

STEP 6: REFINE/LIMIT COMMON LINKS AND INDEPENDENTLY MANAGE DIRECTIONAL DATA

In this step, common links are separated based on their negative or positive correspondences with TMCs.

Input

- CommonLinks.

Output

- LimitedCommonLinks_Negative feature class (modified).
- LimitedCommonLinks_Positive feature class (modified).

Description

Make a copy of CommonLinks and name it LimitedCommonLinks_Negative. Then, join Links_N to this table (XLINK_PVID to LINK_ID) and keep only the matching records. Import the LimitedCommonLinks_Negative to the new geodatabase as a new feature class. The links in this feature class represent those corresponding to negative-direction TMCs (23,731 features).

Make a copy of CommonLinks and name it LimitedCommonLinks_Positive. Then, join Links_P to this table (XLINK_PVID to LINK_ID) and keep only the matching records. Import the LimitedCommonLinks_Positive to the new geodatabase as a new feature class. The links in this feature class represent those corresponding to positive-direction TMCs (23,819 features).

STEP 7: ASSIGN TMCS TO LIMITED COMMON LINKS

Assign TMCs to the common links for positive and negative TMCs separately.

Input

- Limited_LinkTMCRef.xlsx (output from step 5).
- LimitedCommonLinks_ Positive feature class (output from step 6).
- LimitedCommonLinks_ Negative feature class (output from step 6).

Output

- Limited_LinkTMCXRef_Positive table.
- Limited_LinkTMCXRef_Negative table.
- LimitedCommonLinksTMC_Negative feature class.
- LimitedCommonLinksTMC_Positive feature class.

Description

Import positive and negative TMCs from Limited_ LinkTMCXRef.xlsx as tables to the new geodatabase and name them Limited_ LinkTMCXRef_Positive and Limited_LinkTMCXRef_Negative, respectively (24,265 and 24,141 records, respectively).

Join Limited_LinkTMCXRef_Negative to LimitedCommonLinks_Negative (XLINKID to LINKID) and import that as a new feature class to the new geodatabase and name it LimitedCommonLinksTMC_Negative (23,731 features).

Join Limited_ LinkTMCXRef_Positive to LimitedCommonLinks_Positive (XLINKID to LINKID) and import that as a new feature class to the new geodatabase and name it LimitedCommonLinksTMC_Positive (23,819 features).

Note that all previous steps (1 through 7) are done only once as part of the data-linkage process. Conversely, the remaining steps (8 through 10) represent a specific integration example (i.e., selecting times of the day when "free-flow" conditions would be expected); however, different times may be selected, depending on the research question being addressed. Steps 8 through 10 may be repeated as desired by the user.

STEP 8: IMPORT TT DATA

FHWA's *2014 Urban Congestion Trends: Improved Data for Operations Decision Making* report defines the free-flow TT for each TMC path as "the 15th percentile TT during traditional off-peak times (weekdays between 9:00 AM to 4:00 PM and 7:00 PM to 10:00 PM)."[18] These time periods are assumed to represent off-peak, noncongested times of the day where free-flow speeds are most likely to occur.

Import TT data for July through December 2013, separate Weekday-Free-Flow Travel Times, and save tables in the linkage geodatabase. Use weekday epochs of 108 to 191 (corresponding to

9:00 a.m. to 4:00 p.m.) and epochs of 228 to 263 (corresponding to 7:00 to 10:00 p.m.) for this purpose.

Input

- FHWA_TASK2_4_WA_MM_2013_TT, where MM = month of the year, with values of 07 through 12 corresponding to the months of July through December.

Output

- FHWA_TASK2-4_WA_MM_2013_TT.accdb containing the query Weekday-WA_MM_2013-9-4&19-22, where MM = 07 to 12.

- Negative_TMCs_WD_WA_MM_Split_into_12.xlsx, where MM = 07 to 12.

- Positive_TMCs_WD_WA_MM_Split_into_14.xlsx, where MM = 07 to 12.

- Negative_TMCs_WD_WA_All_6_Months_Split_##_of_12.xlsx, where ## = 01 to 12.

- Positive_TMCs_WD_WA_All_6_Months_Split_##_of_14.xlsx, where ## = 01 to 14.

- Negative_TMCs_WD_WA_All_6_Months_Split_All_12_of_12_Summary.xlsx.

- Positive_TMCs_WD_WA_All_6_Months_Split_All_14_of_14_Summary.xlsx.

- Negative_TMCs_WD_WA_All_6_Months_Split_All_12_of_12_Summary.csv.

- Positive_TMCs_WD_WA_All_6_Months_Split_All_14_of_14_Summary.csv.

Description

Import NPMRDS TT data for Washington State covering July through December 2013 to database software and create queries to separate Weekday Free-Flow hours (9:00 a.m. to 4:00 p.m. and 7:00 to 10:00 p.m.).[2] Manage the data in a spreadsheet and import the created tables into the new geodatabase as follows:

- Import TT files for July through December 2013 (such as FHWA_TASK2-4_WA_07_2013_TT.CSV) into database software and create database files out of those (e.g., FHWA_TASK2-4_WA_07_2013_TT.accdb).

- The epochs that will represent free-flow speed are 108 to 191 (from 9:00 a.m. to 4:00 p.m.) and 228 to 263 (from 7:00 to 10:00 p.m.) on weekdays. In the database software, separate these epochs from the rest and create tables out of these data (name the queries Table-Weekday-WA_MM_2013-9-4&19-22, where MM = months from 07 to 12). TTs during these periods are considered as free-flow, and the speeds estimated based on these times are considered as free-flow speeds for this study.

- For each month from July through December (07, 08, 09, 10, 11, and 12), split the TTs for "passenger cars only" into 12 groups for negative TMCs and 14 groups for positive TMCs. (The number of groups was chosen to make the size manageable in the spreadsheet.) Name the produced sample files Negative_TMCs_WD_WA_MM_Split_into_12.xlsx and Positive_TMCs_WD_WA_MM_Split_into_14.xlsx.

- Create a separate file for each of the TMC splits (12 for negative TMCs and 14 for positive TMCs) to contain TTs. Name the file sets Negative_TMCs_WD_WA_All_6_Months_Split_##_of_12.xlsx (where ## = 01 to 12) and Positive_TMCs_WD_WA_All_6_Months_Split_##_of_14.xlsx (where ## = 01 to 14). In these files, combine the TTs for all 6 mo for the same groups of TMCs and derive the statistics for each TMC, as shown in table 4.

Table 4. Statistics derived for each TMC.

Statistic—Abbreviation	Statistic—Full Name
15 Per. TT	15th-percentile travel time
Avg. TT	Average travel time
TT SD	Travel time standard deviation
Min. TT	Minimum travel time
Max. TT	Maximum travel time
# of Records	Number of records

- Copy the summary of statistics for all TMC groups in a single spreadsheet for each of the negative and positive TMCs. Then, create summary worksheets for negative TMCs (Negative_TMCs_WD_WA_All_6_Months_Split_All_12_of_12_Summary.xlsx) and positive TMCs (Positive_TMCs_WD_WA_All_6_Months_Split_All_14_of_14_Summary.xlsx).

- Create csv files (Positive_TMCs_WD_WA_All_6_Months_Split_All_14_of_14_Summary.csv and Negative_TMCs_WD_WA_All_6_Months_Split_All_12_of_12_Summary.csv) from the above spreadsheet files. These files contain the following for each TMC:

  o TMC.
  o 15th-percentile TT.
  o Average TT.
  o TT standard deviation.
  o Minimum TT.
  o Maximum TT.
  o Number of records.

Note that the TT statistics listed in step 8 are calculated using only TMCs with 50 or more TT records. The selection of 50 records as the criterion is arbitrary and based on the judgment of the project team.

STEP 9: CALCULATE WEEKDAY FREE-FLOW SPEED

Calculate speed using TMC lengths for all epochs representing Weekday Free-Flow hours (9:00 a.m. to 4:00 p.m. and 7:00 to 10:00 p.m.).

Input

- Limited_LinkTMCRef (output from step 5).

- FHWA_Monthly_Static_File_Q22013.csv.

- Positive_TMCs_WD_WA_All_6_Months_Split_All_14_of_14_Summary.csv (output from step 8).

- Negative_TMCs_WD_WA_All_6_Months_Split_All_12_of_12_Summary.csv (output from step 8).

Output

- TMC_Counts table.
- Static_File table.
- Static_TMC_Count table (created from the join of TMC_Counts and Static_File).
- Positive_GT50_Cts_TT (TT for 3,064 TMCs).
- Negative_GT50_Cts_TT (TT for 2,671 TMCs).
- Positive_GT50_Cts_TT_Speeds table (TT and speed for 2,592 TMCs).
- Negative_GT50_Cts_TT_Speeds table (TT and speed for 2,594 TMCs).

Note that GT50_Cts in the file names above refers to greater than 50 counts of TT records.

Description

Make a summary of TMCs in the Limited_LinkTMCRef and create the TMC_Counts table in the map. Use the static file FHWA_Monthly_Static_File_Q22013.csv to prepare the Washington State (WA) static file in a spreadsheet. Import the table and name it Static_File. This table contains the lengths of the TMCs, which are used later for the calculation of speed from TTs (where speed = length × 3,600/TT in seconds) (5,860 records).

Join TMC_Counts to Static_File and keep only the matched cases (5,310 out of 5,860). Import this table to the new geodatabase and name it Static_TMC_Count (5,310 rows).

Import the Positive_TMCs_WD_WA_All_6_Months_Split_All_14_of_14_Summary.csv and Negative_TMCs_WD_WA_All_6_Months_Split_All_12_of_12_Summary.csv files that contain

the TMC TTs into the new geodatabase and name them Negative_GT50_Cts_TT and Positive_GT50_Cts_TT. Information for 2,671 TMCs and 3,064 TMCs is contained in these files, respectively.

Join Static_TMC_Count to Negative_GT50_Cts_TT and Positive_GT50_Cts_TT, keeping only the matched cases (2,594 TMCs and 2,592 TMCs remain, respectively).

Create new tables in the geodatabase from the above joined tables and name them Negative_GT50_Cts_TT_Speeds and Positive_GT50_Cts_TT_Speeds. Add fields to the new tables to calculate speeds derived from the TTs. The speeds calculated from the 15th-percentile TTs are used to estimate V_{85}, which represents the operating speed. Speeds calculated from average TTs are used as estimates of V_{avg} in this study.

STEP 10: COMBINE TT AND SPEED DATA WITH LINKAGE DATA

Join the TT and speed statistics with the TMC–Link data produced in step 7. The final feature classes are for positive and negative TMCs Links.

Input

- Positive_GT50_Cts_TT_Speeds table (TT and speed for 2,592 TMCs; output from step 9).

- Negative_GT50_Cts_TT_Speeds table (TT and speed for 2,594 TMCs; output from step 9).

- LimitedCommonLinksTMC_Positive feature class (output from step 7).

- LimitedCommonLinksTMC_Negative feature class (output from step 7).

Output

- Links_Positive_Speeds feature class.
- Links_Negative_Speeds feature class.

Description

Join Negative_GT50_Cts_TT_Speeds to the LimitedCommonLinksTMCs_Negative feature class and join Positive_GT50_Cts_TT_Speeds to the LimitedCommonLinksTMCs_Positive feature class through the TMC columns, keeping only the matched cases.

Make new feature classes from the joined data and name them Links_Positive_Speeds and Links_Negative_Speeds. Delete unneeded columns. These feature classes include a column named Count_TMC that shows the number of links associated with each TMC. Sort the data by TMCs. When the number of rows for a TMC is equal to the values in Count_TMC, all links associated with that TMC are available in this feature class. These TMCs are the candidates to be used in the analyses.

SUMMARY OF THE NPMRDS-RID DATA LINKAGE PROCESS

The preceding section of this report provides a step-by-step guide for linking the NPMRDS database and RID to add TT and speed information to the RID geodatabase. The unmatched links between the NPMRDS and RID as well as the complicated relations between TMCs and links make it very difficult and costly to produce a perfect linkage. However, the proposed method is a practical approach that provides a close-to-perfect linkage with minimal effort.

TT and speed information from the final output feature classes (Links_Positive_Speeds and Links_Negative_Speeds) can be used at the link level. A column named Num_of_Links carries the number of links for each TMC. When this number is equal to the number of available links for the same TMC, the count of links for that TMC is complete (i.e., there is no link missing for the TMC). These TMCs can be used alongside the WA_Alignment feature class to build stretches of highway with known alignment, TT, and speed.

Figure 17 shows RID Links in the Washington RID that are matched to a TMC from the NPMRDS. Figure 18 provides a zoomed-in view of a portion of the Washington State network. Links represented by thin (green) lines show the full RID network; links represented by thick, patterned (blue–black) lines are TMCs with speed data; links represented by thin (red) lines with tick marks are TMCs without speed data. Gaps in the coverage (i.e., RID Links missing a TMC) can be easily seen. Similarly, figure 19 shows a zoomed-in view of two freeway sections and their interchange connection; most of the RID Links missing a TMC are within the interchange area.
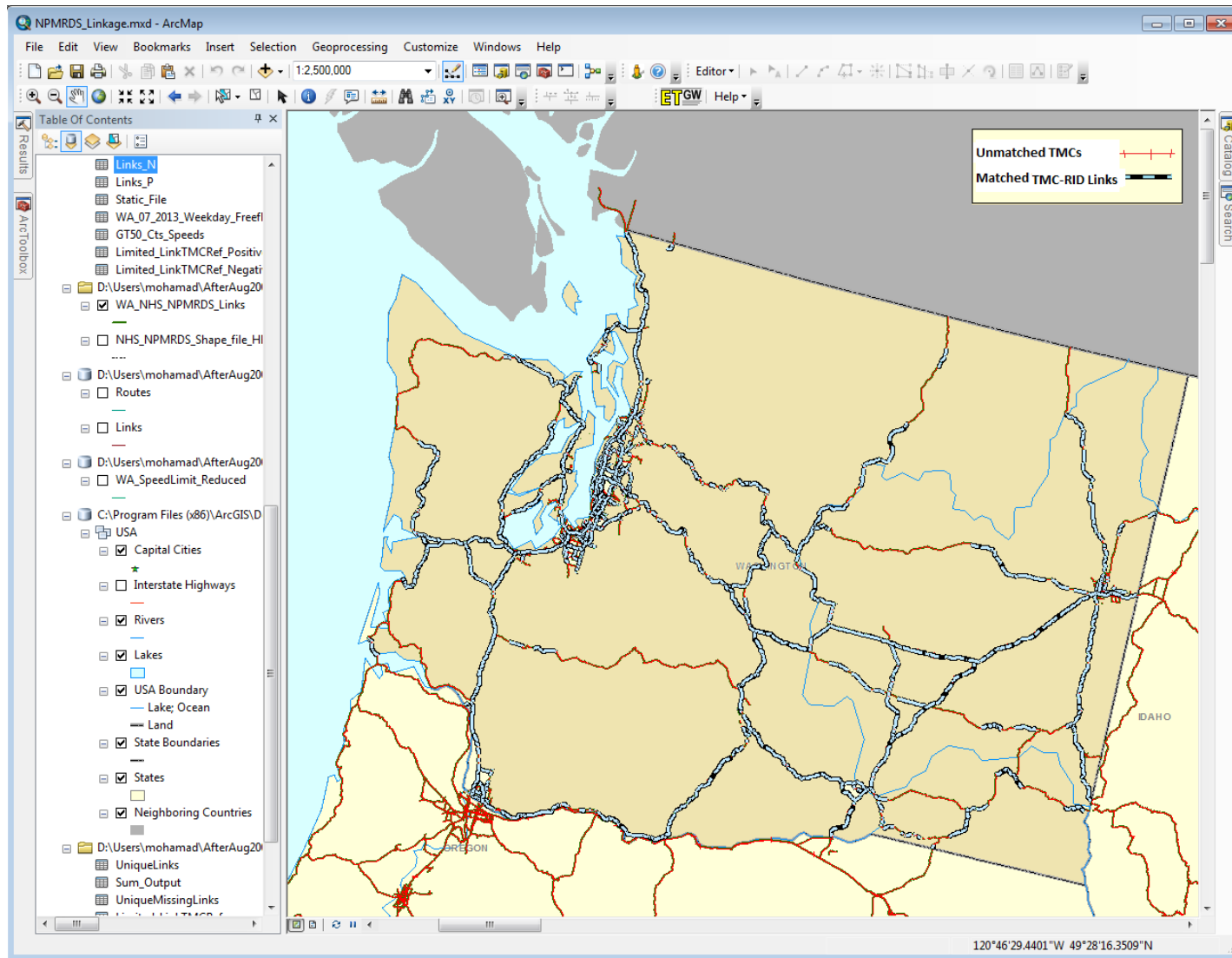
Source: Screen shot created by FHWA using ArcMap™ software by Esri®. ArcMap™ is the intellectual property of Esri® and is used herein under license. StreetMap data source: Tele Atlas.

Figure 17. Screen shot. RID Links in Washington State matched to TMCs from NPMRDS data.[(17)]
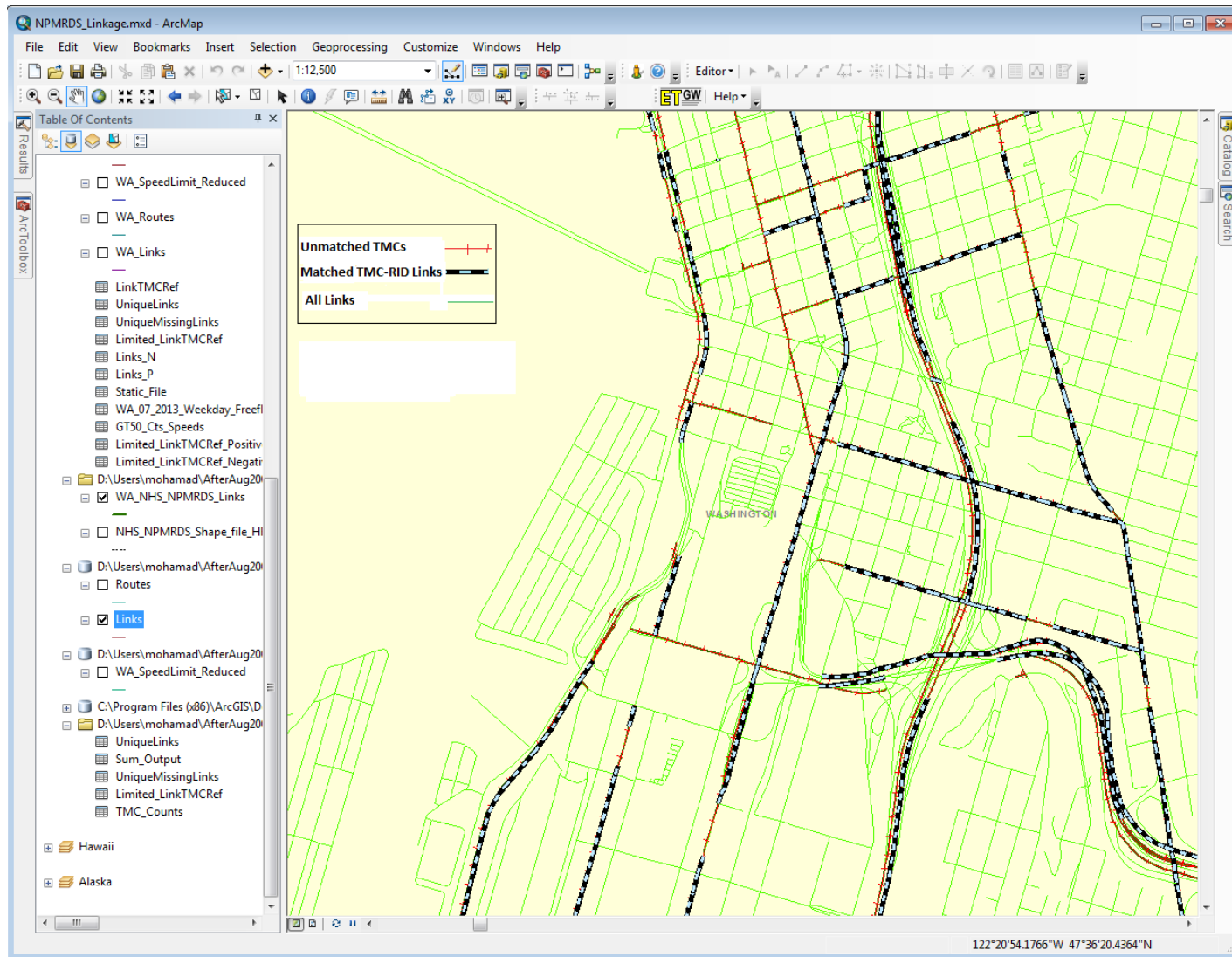
Source: Screen shot created by FHWA using ArcMap™ software by Esri®. ArcMap™ is the intellectual property of Esri® and is used herein under license. StreetMap data source: Tele Atlas.

Figure 18. Screen shot. RID Links matched to TMCs for a small section in Washington State.[17]
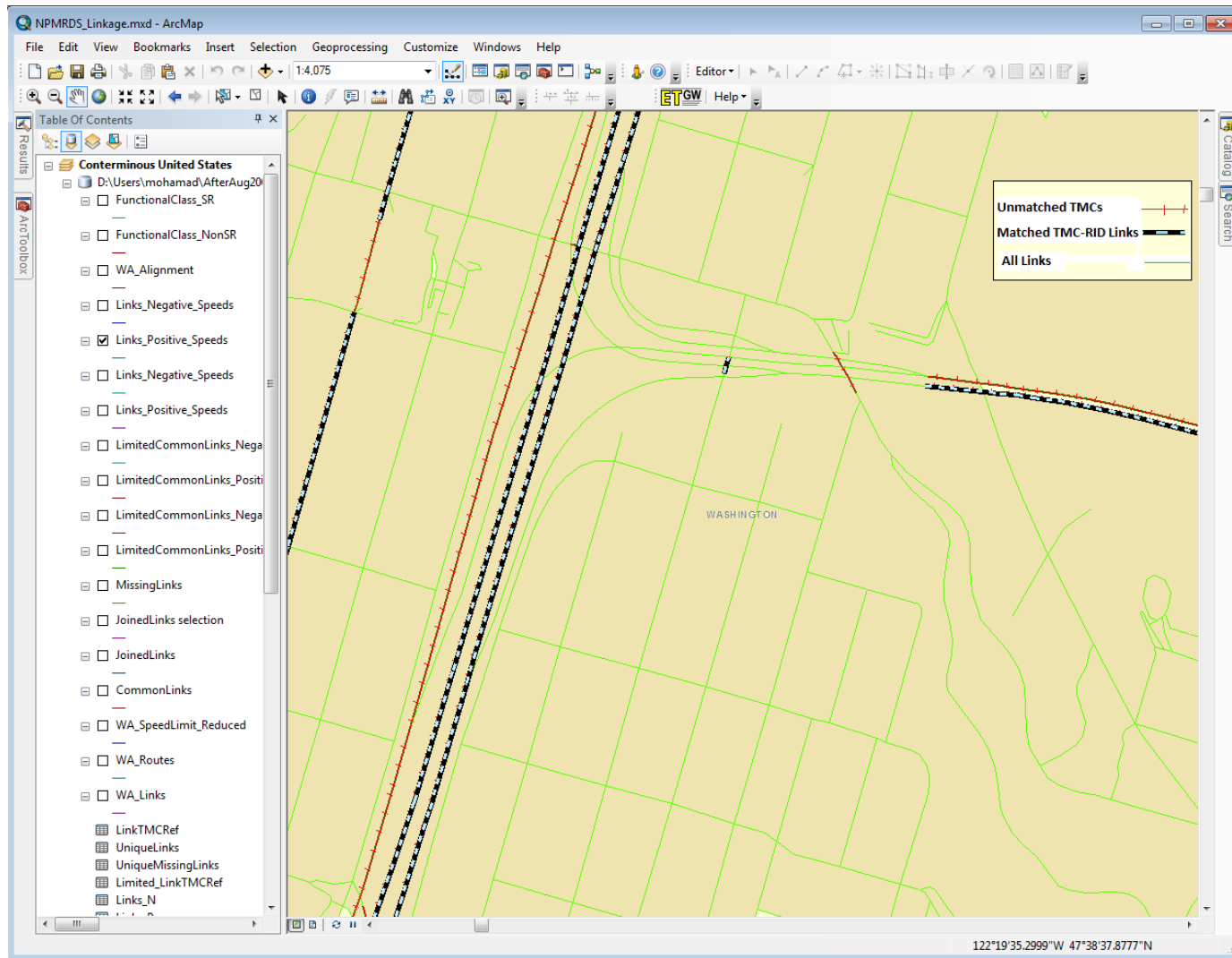
Source: Screen shot created by FHWA using ArcMap™ software by Esri®. ArcMap™ is the intellectual property of Esri® and is used herein under license. StreetMap data source: Tele Atlas.

Figure 19. Screen shot. RID Links matched to TMCs for a freeway interchange in Washington State.[17]

ADDITIONAL DATA PREPARATION FOR THE ANALYSES

Additional preparation of the linked data for conducting the analyses includes the following elements:

- Limiting the network to urban interstates and major arterials.
- Reflecting speed data on the network and building the case study geodatabase.
- Preparing alignment data.
- Preparing crash data.
- Preparing speed data and putting all data together.

The appendix describes each data preparation step in more detail.

LIMITING THE NETWORK TO URBAN INTERSTATES AND MAJOR ARTERIALS

The definition of free-flow conditions applied to NPMRDS data in this study is from the FHWA report, *2014 Urban Congestion Trends: Improved Data for Operations Decision Making*.[2,17] Based on this definition, the network under study is limited to interstate highways and major arterials in urban areas. Development of the WA_Tops_ Urban_Int_Art feature class (i.e., urban interstates and arterials) is described below. This feature class is used to filter the speed data for further analysis.

The initial plan was to use FunctionalClass_SR in the development of the WA_Tops_ Urban_Int_Art feature class. However, further investigation showed that FunctionalClass_SR is missing significant lengths of urban interstates in the WA database. Although this was reported to the RID developer and corrected, the correction was completed after development of the WA_Tops_ Urban_Int_Art feature class for this study.[1]

The project team initially considered using only the "oneway" attribute in Routes to target one-way highways representing divided highways. However, some problems were discovered with the conflation of HPMS data into Expressways in the WA RID.[8] Therefore, the plan was changed to filter urban interstates and major arterials.

REFLECTING SPEED DATA ON THE NETWORK AND BUILDING THE CASE STUDY GEODATABASE

Once the urban interstates and major arterials network is built, the next steps include the following elements:

- Reflecting the speed data contained in the linked data feature classes (Links_Positive_Speeds and Links_Negative_Speeds) on the Washington State urban interstates and arterials feature class (WA_Tops_Urban_Int_Art).

- Building a subset of the Alignment for the case study analyses.

In addition to the RID feature classes, this process also uses the Alignment data collected from mobile vans (i.e., the original alignment data) to relate the two directions of travel of the same highway. The last step is to create a new geodatabase for use in the case study analyses.

PREPARATION OF ALIGNMENT DATA

The appendix describes in detail the final preparation of alignment data for the analyses.

PREPARATION OF CRASH DATA

WA RID crash data for 2011 to 2013 are used in this study. For divided highways, crashes are conflated mostly on one direction of the highway. Therefore, to use WA RID crashes for any evaluation, both directions of the highway must be combined.

The appendix describes in detail the preparation of crash data for the analyses.

PREPARATION OF SPEED DATA AND PUTTING ALL DATA TOGETHER

The RID WA_SpeedLimit_Reduced feature class contains many records with nonvalid speed-limit values. Delete the invalid records to produce a feature class (WA_SpeedLimit_Reduced_non_Zero) containing only valid speed-limit values (11,741 of 268,550 records remain).

Conduct dynamic segmentation to produce two tables (WA_Align_SL_P_Speed_Intersect and WA_Align_SL_N_Speed_Intersect) that carry the speed limit and operating speeds for links of different directions of the chosen highways. These tables are used in the analyses and contain all necessary information, including the following elements:

- Alignment information from the RID Alignment feature class.

- Speed-limit information from the RID WA_SpeedLimit_Reduced feature class.

- V_{85} and V_{avg} from the NPMRDS (averaged along TMCs, dynamically segmented over Alignment records).

- Route, Fname, Visi_from, Visi_to, and vis_dir from the mobile-van data.

In these files, the RouteID, FrMeasure, and ToMeasure columns identify the actual segments of highways, and vis_dir with values of 5 or 6 determines the directions of travel.

> Note that, for analyses that require additional RID data, further dynamic segmentation can be conducted.

Figure 20 shows a new geodatabase created after conducting the above data preparation for the analyses. The appendix describes in detail the preparation of speed data and putting all data together.
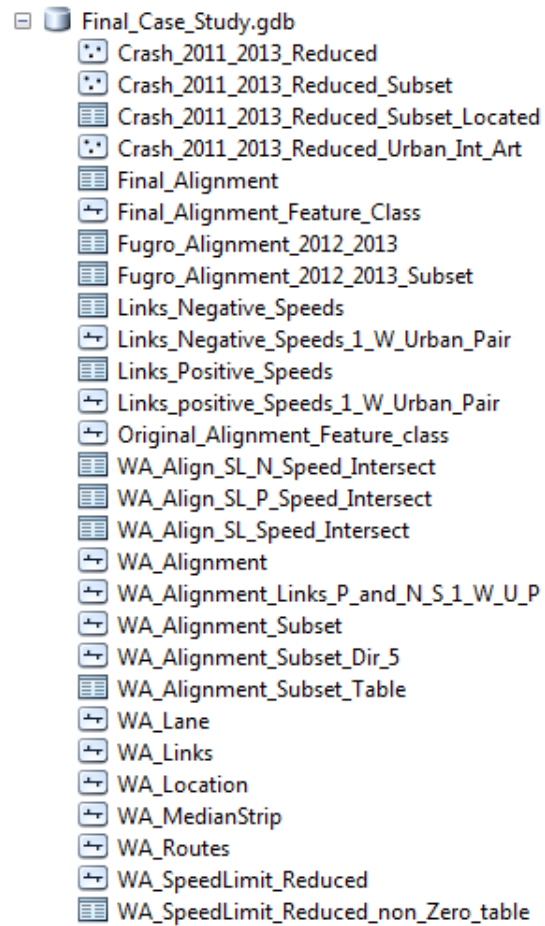
Figure 20. Screen shot. Final Case Study geodatabase.[17]

ANALYSES USING THE LINKED NPMRDS-RID DATA

The following analyses were conducted using the linked NPMRDS-RID data:[2,1]

- **Analysis 1:** Studied the relationship between V_{85} and the speed limit during noncongested conditions.

- **Analysis 2:** Studied the relationship between crash severity and the differential between posted and operating speeds during noncongested conditions.

The data sources for both analyses are the same. However, the analysis unit differs. Analysis 1 does not require any data that must be aggregated for both directions of travel (e.g., crash data). Therefore, data from each direction of travel can be used independently to conduct the analysis.

Crash data are used in analysis 2. Since the RID crash data are not conflated accurately to both directions of travel (i.e., crashes are mostly conflated to one side of a divided highway), the two directions must be combined and segments of highways defined as the analysis unit.

85TH-PERCENTILE OPERATING SPEED

V_{85} defines the speed that 85 percent of drivers will drive at or below under free-flow conditions.[19] V_{85} is a value that is used by many States and cities for establishing regulatory speed zones. Use of the V_{85} concept is based on the theory that the large majority of drivers are reasonable and prudent, do not want to have a crash, and desire to reach their destination in the shortest possible time, and the speed at or below which 85 percent of people drive at any given location under good weather and visibility conditions may be considered as the maximum safe speed for that location.[20]

ANALYSIS 1: NONCONGESTED OPERATING SPEED (V_{AVG} AND V_{85}) AS A FUNCTION OF SPEED LIMIT

The following research questions were addressed in analysis 1:

- What is the relationship between V_{avg} and the posted speed for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

- What is the relationship between V_{85} and the posted speed for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

A combination of the data gathered in the WA_Align_SL_P_Speed_Intersect and WA_Align_SL_N_Speed_Intersect tables (see the Preparation of Speed Data and Putting All Data Together section of this report) was used in this analysis. These tables each carry data for nonoverlapping sections of the target highways by direction. Directions identified by the vis_dir variable with values of 5 and 6 are identified in the mobile-van data based on the order in which these directions are surveyed by instrumented vans. This does not necessarily follow the order of directions in the RID or the actual highways on the ground.

The focus for this analysis is on the horizontal alignment of segments, posted speed, V_{avg}, and V_{85}. Since the highways under study are divided, these values are available for both directions independently. Therefore, the decision was made to use both directions independently in the analysis.

Data reduction and analysis steps can be summarized as follows:

1. Export the WA_Align_SL_P_Speed_Intersect and WA_Align_SL_N_Speed_Intersect tables outside ArcGIS®.[17]

2. Combine both tables in one worksheet in a spreadsheet tool. There are 2,650 records in the combined worksheet.

3. Add a Degree of Curvature column to the data and calculate the degree of curvature for each section (tangents are assigned a value of 0).

4. Add another column with the values of Degree of Curvature Squared.

5. Conduct the following four regression analyses:

   o V_{85} as the dependent variable; Speed Limit and Degree of Curvature as independent variables.

   o V_{85} as the dependent variable; Speed Limit and Degree of Curvature Squared as independent variables.

   o V_{avg} as the dependent variable; Speed Limit and Degree of Curvature as independent variables.

   o V_{avg} as the dependent variable; Speed Limit and Degree of Curvature Squared as independent variables.

Analysis 1 Results

Analysis 1 results include regression models for estimating V_{85} and V_{avg} during off-peak (noncongested) periods as a function of posted speed and weighted average of degree of curvature.

For estimating both V_{85} and V_{avg}, the results were slightly better when Degree of Curvature Squared was included as an independent variable (rather than Degree of Curvature).

V_{85} As a Function of Posted Speed and Weighted Average of Degree of Curvature

The analysis results for estimating V_{85} are shown in this section and in table 5, figure 21, and figure 22.

Overall fit statistics for the regression analysis results for V_{85} based on speed limit and degree of curvature squared include the following:

- Multiple R: 0.738545.
- R^2 (coefficient of determination): 0.545448.
- Adjusted R^2: 0.545105.
- Standard error: 3.824613.
- Observations: 2,650.

Table 5 shows Analysis of Variance (ANOVA) statistics.

Table 5. ANOVA statistics for V_{85} as a function of speed limit and degree of curvature squared.

ANOVA Statistic	Regression	Residual	Total	Intercept	Speed Limit	Degree of Curvature Squared
Degrees of freedom	2	2,647	2,649	-	-	-
Sum of squares	46,462.090	38,719.42	85,181.51	-	-	-
Mean square	23,231.040	14.62766	-	-	-	-
F-test statistic (F)	1,588.158	-	-	-	-	-
Significance F	0	-	-	-	-	-
Coefficients	-	-	-	32.746750	0.502617	−0.093590
Standard error	-	-	-	0.590481	0.010042	0.007115
t Stat	-	-	-	55.457710	50.050630	−13.153700
P-value	-	-	-	0	0	2.51E-38
Lower 95%	-	-	-	31.588900	0.482925	−0.107540
Upper 95%	-	-	-	33.904600	0.522308	−0.079640

-No data.

The resulting model is a relatively good estimation of V_{85} (e.g., adjusted R^2 value of 0.545). Although the effect of degree of curvature is small, it is beneficial to include it—especially to account for the effect of curvature of sharper curves.

$$V_{85th-off-Peak} = 32.8 + 0.50 \times V_{Posted} - 0.1 \times (D_w)^2$$

Figure 21. Equation. V_{85} during off-peak periods.

Where:
   $V_{85th-off-Peak}$ = V_{85} during off-peak period, excluding nighttime (miles per hour).
   V_{Posted} = speed limit (miles per hour).
   D_W = weighted average of degree of curvature; if there are multiple curve and tangent sections, this variable represents the weighted average based on the lengths of the curves and tangents. Tangents have 0 degree of curvature.

Figure 22 shows the variation of V_{85} with respect to the two independent variables in a graphical form for curves with radii of 500; 1,000; and 5,000 ft. These radii correspond to degree of curvature squared of 131.00, 32.80, and 1.31, respectively.

The results suggest that, on sections with very flat, horizontal curves (weighted-average radius of 5,000 ft), V_{85} exceeds the speed limit when the posted speed is less than 65 mph. V_{85} exceeds the speed limit by as much as 30 percent at lower speeds (i.e., 15 percent of drivers exceed the posted speed by as much as 30 percent at lower speeds), with the differential decreasing as the speed limit increases. For speed limits of 65 mph and above, V_{85} is lower than the speed limit.

For sections with sharper curves (weighted-average radius of 1,000 ft), the 85th-percentile driver exceeds the speed limit when the posted speed is less than 60 mph; for a speed limit of 40 mph, the differential is nearly 25 percent. As with flatter sections, the differential between V_{85} and posted speeds decreases as the speed limit increases. For speed limits of 60 mph and above, V_{85} is lower than the speed limit.

For sections with sharp curves (weighted-average radius of 500 ft), V_{85} is lower than the speed limit for all posted speeds, with the differential increasing as the speed limit increases.

Therefore, given the above, a conclusion is that the amount by which the speed limit is exceeded depends, in part, on how sharp the curves are. When speed limits are above 65 mph, drivers appear less likely to exceed the posted speed, even when the road alignment is relatively straight.
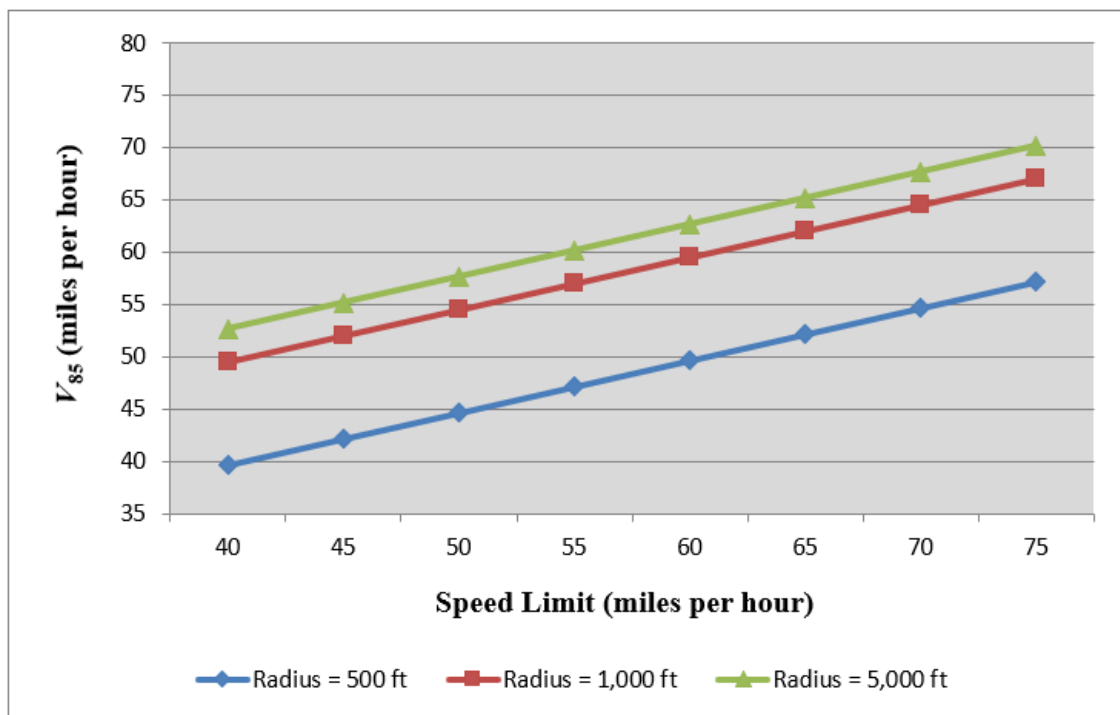
Source: FHWA.

Figure 22. Graph. V_{85} as a function of speed limit and curvature.

V_{avg} As a Function of Posted Speed and Weighted Average of Degree of Curvature

While the analysis results for V_{avg} were not as promising as for V_{85} in terms of goodness of fit (e.g., adjusted R^2 value of 0.394 versus 0.545), they would still be valuable when other means to estimate average speeds are lacking. Analysis results for estimating V_{avg} are shown below and in table 6, figure 23, and figure 24.

Overall fit statistics for the regression analysis results for V_{avg} based on speed limit and degree of curvature squared include the following:

- Multiple R: 0.627781.
- R^2: 0.394109.
- Adjusted R^2: 0.393651.
- Standard error: 8.722414.
- Observations: 2,650.

Table 6 shows ANOVA statistics.

Table 6. ANOVA statistics for V_{avg} as a function of speed limit and degree of curvature squared.

ANOVA Statistic	Regression	Residual	Total	Intercept	Speed Limit	Degree of Curvature Squared
Degrees of freedom	2	2,647.0	2,649.0	-	-	-
Sum of squares	130,993.40	201,385.1	332,378.5	-	-	-
Mean square	65,496.69	76.08051	-	-	-	-
F-test statistic (F)	860.8867	-	-	-	-	-
Significance F	1E-288	-	-	-	-	-
Coefficients	-	-	-	−1.041450	0.890098	−0.077780
Standard error	-	-	-	1.346652	0.022902	0.016227
t Stat	-	-	-	−0.773360	38.865240	−4.793470
P-value	-	-	-	0.439376	7.9E-262	1.73E-06
Lower 95%	-	-	-	−3.682050	0.845190	−0.109600
Upper 95%	-	-	-	1.599146	0.935006	−0.045960

-No data.

As with the model for V_{85}, although the effect of degree of curvature is small, including it is beneficial—especially to account for the effect of curvature of sharper curves.

$$V_{avg-Off-Peak} = -1.04 + 0.89 \times V_{Posted} - 0.08 \times (D_W)^2$$

Figure 23. Equation. V_{avg} during off-peak periods.

Where $V_{avg-Off-Peak}$ is average speed during off-peak period, excluding nighttime (miles per hour).

Figure 24 shows the variation of V_{avg} with respect to the two independent variables in a graphical form for curves with radii of 500; 1,000; and 5,000 ft. These radii correspond to degree of curvature squared of 131.00, 32.80, and 1.31, respectively.

The results indicate that V_{avg} is consistently below the posted speeds, even on relatively straight roads. For sections with average weighted curve radius of 5,000 ft, the differential between V_{avg}

and posted speed ranges from 7 to 10 mph, depending on the posted speed. The slopes of V_{avg} versus posted speed plots are similar for sections with sharper curves.
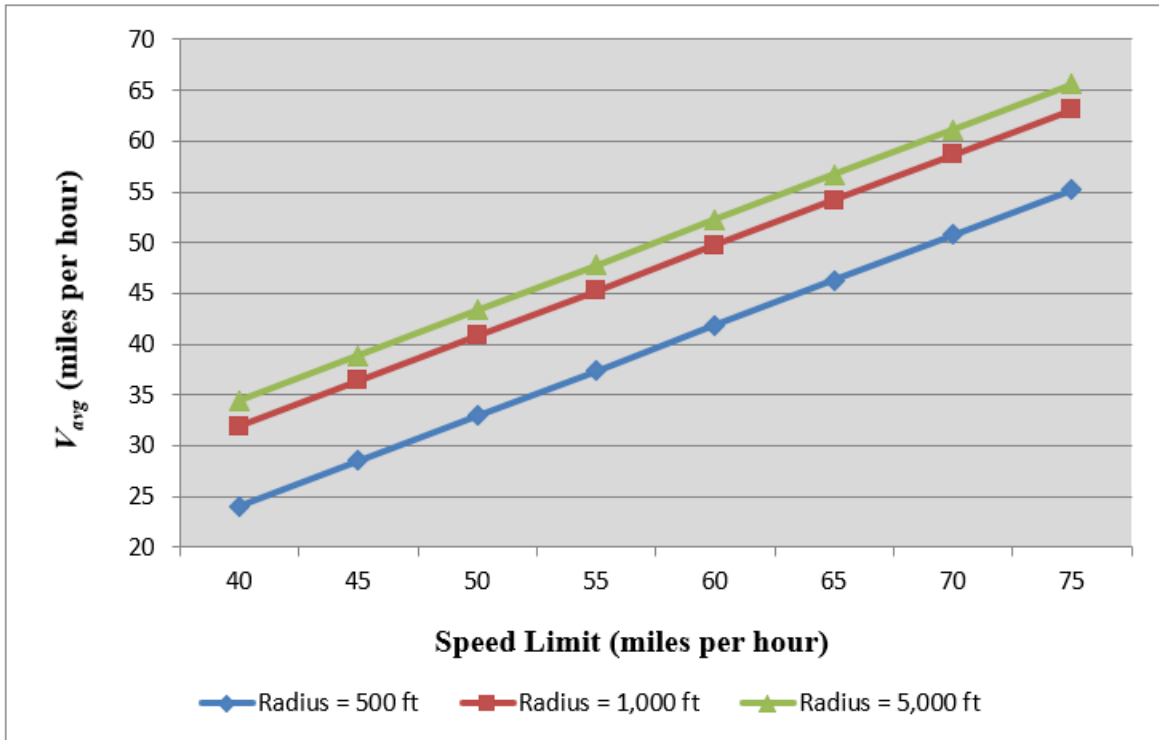
Source: FHWA.

Figure 24. Graph. V_{avg} as a function of speed limit and curvature.

ANALYSIS 2: SEVERITY OF CRASHES AS A FUNCTION OF THE DIFFERENTIAL BETWEEN THE SPEED LIMIT AND V_{85}

The research question addressed in analysis 2 was as follows: How is crash severity related to the differential between the posted speed and V_{85} for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

The objective of analysis 2 was to investigate whether crash severity can be explained by the magnitude of the differential between V_{85} and the speed limit. Since other highway characteristics that affect crash frequency (e.g., annual average daily traffic (AADT)) are not considered in this study, the ratio of fatal and injury (FI) crashes to total crashes is used. This measure represents the relative crash severity for each analysis unit.

For each highway, most crash data in the Washington State RID geodatabase are conflated to only one direction of travel. This limitation forces the use of both directions of travel combined when conducting analysis using crash data from this database.

Data reduction and analysis steps can be summarized as follows:

1. Continue to utilize the same spreadsheet tool worksheet used for analysis 1. There are 2,650 records in the combined worksheet.

2. Use Crash.dat and Road.dat files (see appendix, "Preparation of Crash Data") outside of the spreadsheet tool with a C programming code to identify crashes and crash severities for all 2,650 records (segments) identified in step 1.

3. For each segment, add the number of total and FI crashes for the worksheet.

4. Sort the data based on Route, Vis_Dir, Visi_From, and FrMeasure. This sorting brings together the sections of highways that form the Both-Direction Highway Segments identified in table 10 (see appendix).

5. Each Both-Direction Highway Segment identified in table 6 comprises some rows of data with Dir 5 and some with Dir 6 (i.e., the two directions of travel for the same section of a highway). Aggregate the data for each segment. Speed values for each segment are the weighted average of the speed values of individual sections, and crashes are the sum of crashes for all sections in both directions.

6. For each analysis unit (Both-Direction Highway Segment), calculate the ratio of FI crashes to total crashes. This ratio is used as a measure for the severity of crashes on these highways.

7. For each analysis unit, calculate the speed differential between V_{85} and speed limit.

Table 7 shows the summary of the data prepared for analysis 2. The goal is to investigate the relationship between the ratio of FI crashes to total crashes and the speed differential between V_{85} and the speed limit.

Table 7. Speeds and FI–total crashes ratio by analysis unit.

Route	Section	Average Speed Limit (mph)	Average V_{85} (mph)	Speed Differential (mph)	FI–Total Crashes Ratio
3	1	54.7	58.9	4.1	0.270
3	2	60.0	63.8	3.8	0.321
3	3	59.3	62.6	3.4	0.321
5	1	70.0	70.3	0.3	0.316
5	2	64.1	66.0	1.8	0.245
5	3	62.6	65.5	2.9	0.313
5	4	57.0	63.8	6.8	0.349
5	5	60.0	63.5	3.5	0.316
5	6	59.8	65.4	5.7	0.171
5	7	60.0	60.2	0.2	0.574
5	8	60.0	63.7	3.7	0.364
5	9	60.0	63.8	3.8	0.249
5	10	59.1	64.5	5.5	0.258
5	11	60.0	63.9	3.9	0.313
5	12	62.1	64.1	2.0	0.175
16	1	57.1	62.2	5.2	0.312
16	2	60.0	63.7	3.7	0.446
16	3	54.7	58.3	3.6	0.291
18	1	49.9	61.4	11.5	0.531
18	2	60.0	62.4	2.4	0.389
18	3	60.0	65.6	5.6	0.357
18	4	60.0	65.6	5.6	0.238
99	1	59.6	60.0	0.4	0.346
99	2	40.0	51.2	11.1	0.299
99	3	39.3	45.8	6.5	0.349
99	4	39.6	44.8	5.2	0.269
101	1	59.2	63.3	4.0	0.372
167	1	60.0	63.1	3.1	0.420
167	2	59.8	61.3	1.5	0.474
405	1	59.9	61.3	1.4	0.321
405	2	60.0	61.5	1.5	0.379
410	1	55.0	60.4	5.4	0.333
410	2	55.0	57.8	2.8	0.455
512	1	59.2	62.0	2.8	0.396
512	2	60.0	62.6	2.6	0.138
520	1	53.9	60.3	6.4	0.433
520	2	59.8	65.7	5.9	0.224
520	3	60.0	65.2	5.2	0.295
522	1	59.1	61.3	2.1	0.441
526	1	58.7	58.4	−0.3	0.533

mph = miles per hour.

Analysis 2 Results

Linear regression was conducted to investigate whether the FI–total crashes ratio can be expressed as a function of speed differential between V_{85} and the speed limit.

The results were not satisfactory with regard to the goodness-of-fit measures, and no conclusion could be derived. Considering that the analysis units listed in table 7 are of different lengths, another regression analysis was conducted using the segment lengths as weight factors (weighted linear regression). These results were slightly better but still not satisfactory (e.g., a P-value of 0.225 for ln(speed differential), where less than 0.050 is desirable). Counter to expectations and engineering judgment, the results suggest that, as the speed differential increases, the ratio of FI crashes to total crashes decreases. The analysis results for estimating the FI–total crash ratio are shown below and in table 8, figure 25, and figure 26.

Overall fit statistics for the regression analysis results include the following:

- Multiple R: 1.0.
- R squared: 0.9.
- Adjusted R squared: 0.9.
- Standard error: 0.1.
- Observations: 40.0.
- Akaike information criterion (AIC): −153.048.
- AIC with a correction for small sample sizes: −152.382.

Table 8 shows ANOVA statistics, with alpha equal to 0.05, and figure 25 presents the results in an equation.

Table 8. Weighted regression analysis results for FI–total crash ratio.

ANOVA Statistic	Regression	Residual	Total	Section Length (Miles)	LN (Speed Differential)
Degrees of freedom	2	38	40	-	-
Sum of squares	14.830860	0.788732	15.61959	-	-
Mean square	7.415429	0.020756	-	-	-
F-test statistic	357.265	-	-	-	-
P-value	2.3E-25	-	-	-	-
Significant?	Yes	-	-	-	-
Coefficients	-	-	-	0.345025	−0.015670
Standard error	-	-	-	0.018590	0.012700
t Stat	-	-	-	18.559250	−1.233540
P-value	-	-	-	1.2E-20	0.224951
Lower	-	-	-	0.307390	−0.0413800
Upper	-	-	-	0.382659	0.010044
VIF	-	-	-	1.403400	1.403400

-No data; VIF = variance inflation factor.

$$Ratio_{\frac{FI\,Crashes}{Total\,Crashes}} = 0.345 - 0.0157 \times \ln(Speed\,Differential)$$

Figure 25. Equation. Ratio of FI to total crashes as a function of speed differential.

Where:

$Ratio_{\frac{FI\,Crashes}{Total\,Crashes}}$ = ratio of FI crashes to total crashes.

Speed Differential = speed differential between V_{85} and speed limit.

Figure 26 shows the results of the regression analysis in graphical form.
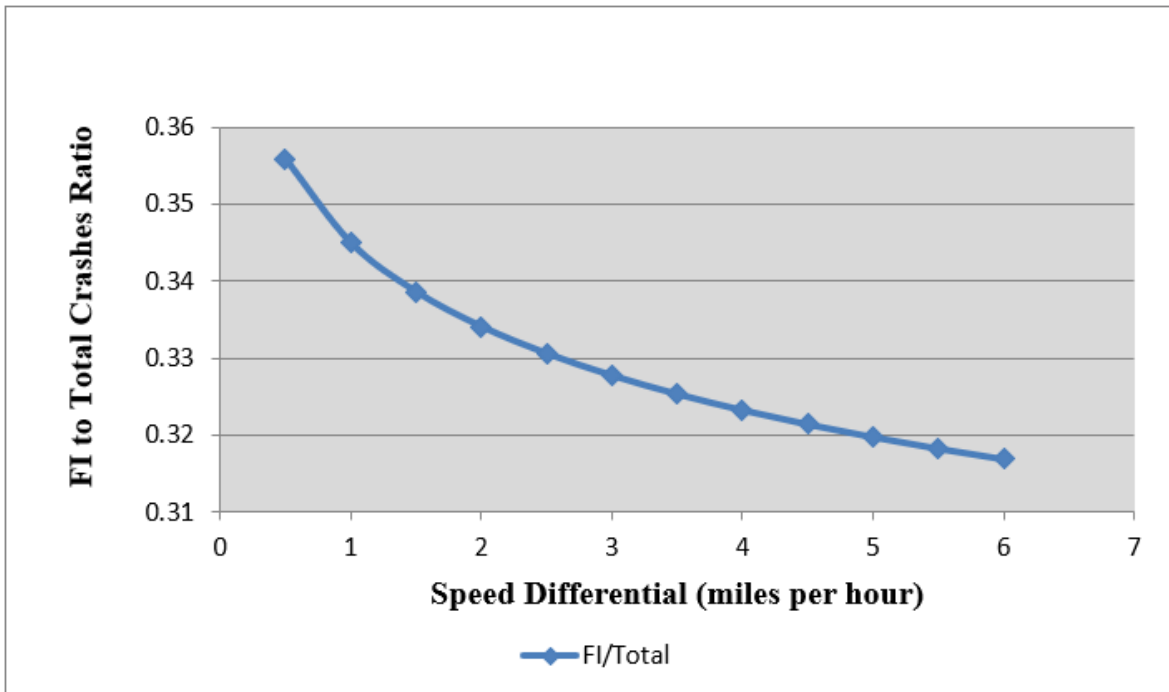
Source: FHWA.

Figure 26. Graph. Relationship of speed differential and FI–total crashes ratio.

Although the results were unanticipated, there is a potential explanation. The analysis 1 results show that differentials between operating speed and posted speed are greater on sections with lower posted speeds. Since crash severity is influenced by speed (with higher speeds generally resulting in more severe crashes), an expectation is that crashes are relatively less severe at lower speeds than at higher speeds. Since the greater speed differentials occur on sections with lower posted speeds, it would reason that the ratio of FI crashes to total crashes could be lower on those sections compared to sections with smaller speed differentials (i.e., sections with higher posted speeds).

SUMMARY

This section contains a summary of the analyses performed and the results.

ANALYSES PERFORMED

The NPMRDS and RID data for urban interstates and major arterials in Washington State were linked together.[2,1] The NPMRDS data covered the period of July through December 2013, and the RID crash data were extracted for the period of 2011 to 2013.

The following analyses were performed using the linked NPMRDS-RID data:

- **Analysis 1**: Studied the relationship between estimated V_{avg} and V_{85} and the speed limit during noncongested conditions.

- **Analysis 2:** Studied the relationship between crash severities and the differential between posted and operating speeds in noncongested conditions.

The following research questions were addressed in analysis 1:

- What is the relationship between V_{avg} and the posted speed for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

- What is the relationship between V_{85} and the posted speed for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

The following research question was addressed in analysis 2:

- How is crash severity related to the differential between the posted speed and V_{85} for passenger cars on urban interstates and major arterials in Washington State during noncongested conditions?

Another analysis option considered was to investigate secondary crashes that occur as the result of congestion following an initial crash. However, crash data in the RID are not necessarily conflated to the proper direction of travel in the Washington State database. This limitation precluded any analysis that required crash data by direction of travel.

ANALYSES RESULTS

Summary of Analysis 1 Results

Analysis 1 results include regression models for estimating V_{85} and V_{avg} for passenger cars during off-peak (noncongested) periods as a function of posted speed and weighted average of degree of curvature.

V_{85} *As a Function of Posted Speed and Weighted Average of Degree of Curvature*

The results suggest that, on sections with very flat, horizontal curves (weighted-average radius of 5,000 ft), V_{85} exceeds the speed limit when the posted speed is less than 65 mph. V_{85} exceeds the speed limit by as much as 30 percent at lower speeds (i.e., 15 percent of drivers exceed the posted speed by as much as 30 percent at lower speeds), with the differential decreasing as the speed limit increases. For speed limits of 65 mph and above, V_{85} is lower than the speed limit.

For sections with sharper curves (weighted-average radius of 1,000 ft), V_{85} exceeds the speed limit when the posted speed is less than 60 mph; for a speed limit of 40 mph, the differential is nearly 25 percent. As with flatter sections, the differential between V_{85} and posted speeds decreases as the speed limit increases. For speed limits of 60 mph and above, V_{85} is lower than the speed limit.

For sections with sharp curves (weighted-average radius of 500 ft), V_{85} is lower than the speed limit for all posted speeds, with the differential increasing as the speed limit increases.

Therefore, a conclusion is that the amount by which the speed limit is exceeded by V_{85} depends, in part, on how sharp the curves are. When speed limits are above 65 mph, drivers appear less likely to exceed the posted speed, even when the road alignment is relatively straight.

V_{avg} *As a Function of Posted Speed and Weighted Average of Degree of Curvature*

The results indicate that average speeds are consistently below the posted speeds, even on relatively straight roads. For sections with an average-weighted curve radius of 5,000 ft, the differential between V_{avg} and the posted speed ranges from 7 to 10 mph, depending on the posted speed. The slopes of V_{avg} versus posted speed plots are similar for sections with sharper curves.

Summary of Analysis 2 Results

Analysis 2 results include a model for the relationship of crash severity versus the differential of posted and operating speeds for passenger cars in noncongested conditions. Crash severity was represented by the ratio of FI crashes to total crashes.

The results suggest that, as the operating versus posted speed differential increases, the ratio of FI crashes to total crashes decreases. Although this initially seems counterintuitive, there is a potential explanation. The analysis 1 results show that differentials between operating speed and posted speed are greater on sections with lower posted speeds. Since crash severity is influenced by speed (with higher speeds generally resulting in more severe crashes), an expectation is that crashes are relatively less severe at lower speeds than at higher speeds. Since the greater speed differentials occur on sections with lower posted speeds, it would reason that the ratio of FI crashes to total crashes could be lower on those sections compared to sections with smaller speed differentials (i.e., sections with higher posted speeds).

CONCLUSIONS

LESSONS LEARNED AND CHALLENGES TO USING THE LINKED RID-NPMRDS DATA FOR SAFETY ANALYSES

Limitations of the data presented several challenges in linking the databases and performing analyses. This section discusses the following:

- Challenges related to the RID.[1]
- Challenges related to the NPMRDS.[2]
- Challenges related to linking the RID and the NPMRDS.
- Challenges related to using the linked RID-NPMRDS data for speed–safety analyses.
- Other lessons learned.

Challenges Related to the RID

The RID characteristics and the way different features are conflated to the geodatabase base map create challenges in using these data for safety analysis, including the following:

- Conflation of crash data to the base map for divided highways: On divided highways, crashes are conflated mostly on one side of the highway. Therefore, when conducting any type of analysis related to crashes, both directions combined must be used.

- In the RID, the two different directions of a divided highway have different RouteIDs, and there is nothing that easily relates the two RouteIDs of the same highway to each other. While there might be a way to relate the two directional travelways of a divided highway through the RouteID, the means were not evident to the project team. Therefore, in this study, the matching process was done manually.

- Missing attribute data: For example, the RID WA_SpeedLimit_Reduced feature class contains many records with nonvalid speed-limit values.

- FunctionalClass_SR missing significant lengths of urban interstates in the WA database: Although this was reported to the RID developer and corrected, the correction was completed after development of the Washington State urban interstates and arterials feature class for this study.

Challenges Related to the NPMRDS

Challenges related to the NPMRDS (v1 was used in this research) include the following:

- TMCs for different directions of travel on divided highways not usually related to the same sections of the divided highways: As a result, when the NPMRDS was linked to the RID, the speed data were created in two feature classes: one Positive (TMCs with positive direction with respect to the links' directions) and one Negative (TMCs with negative direction with respect to the links' directions).

- The many-to-many relationship between NPMRDS geospatial links included in the shapefile and TMCs defined by the static file:

  o The many-to-many relationship means that one TMC is usually made up of several NPMRDS Links, and an NPMRDS Link may be part of two or more TMCs. The possibility that one link may be part of two or more TMCs comes from the route-based nature of the TMCs. The understanding is that, where two or more sign routes share the same physical roadway (e.g., I-95 and U.S. Route 17 in Richmond, VA), each sign route has its own unique TMC code assigned to the roadway-centerline segment representing that roadway.

  o Links with multiple TMC codes occur for two different reasons. First, they may represent a single roadway-centerline that allows bidirectional travel (e.g., a two-lane road with no median). However, they may also represent a road that has separate TMC codes for a road that has a local street name, a county route number, a State route number, and possibly even a U.S. primary route number. Each of these TMC codes can include a different number of links and therefore be associated with different travel times.

- Missing attribute data: Gaps and outliers exist in the TT data, which need to be resolved before these data can be used in any performance evaluation.

- The large size of the NPMRDS database: Since the NPMRDS provides TTs every 5 min for each TMC, it contains a large volume of data that can be cumbersome to manipulate and difficult to mine.

Challenges Related to Linking the RID and the NPMRDS

Challenges to linking the RID and NPMRDS data included the following:

- Not all NPMRDS TMCs could be linked to the RID base-map Links.

- Identifying the location of the TMC segment on the RID base map requires simultaneous matching of the TMC segments to the NPMRDS Links identified in the shapefile and matching of the NPMRDS Links to the RID Links. The many-to-many relationships between NPMRDS geospatial links and TMCs create some challenges in doing so.

- Different approaches are possible for linking the data sources. Before attempting the linkage, it was not evident which method is most efficient.

- Each side of a divided highway is represented by different links. It can be difficult to determine which links should be paired to represent both sides of a divided highway section.

- Dynamic segmentation was used to overcome the challenge of reflecting speeds on the highway analysis units.

Several other technical challenges could arise when linking the RID with other data sources. For example, the RID includes Additional Supplemental Data that are not yet linked geospatially with the RID GIS data warehouse. In addition, the RID data are conflated to a centerline GIS map that may vary from the current ground-truth and/or the centerline system of other databases, which could make reconciliation among the different centerlines and LRSs challenging.

Challenges Related to Using the Linked RID-NPMRDS Data for Speed–Safety Analyses

Challenges to using the linked RID-NPMRDS data for speed–safety analyses include the following:

- NPMRDS TT data are provided for TMCs, many of which can be quite long (at least 1 to 2 mi). Therefore, speeds derived from the TTs were average speeds over a relatively long segment. This precluded the estimation of spot speeds (e.g., the speed on a particular curve) and necessitated the assumption of a uniform operating speed across the TMC for each epoch (5-min period). The effect of specific elements on speed is then difficult or impossible to discern, which reduces or eliminates the possibility of examining relationships between individual design elements (e.g., single horizontal curves) and speed/safety. For example, the effect of a single horizontal curve on operating speed cannot be determined when a TMC includes several curves. In this study, a weighted average of degree of curvature was used for each analysis unit rather than examining the effect of each curve.

- The start and end locations of NPMRDS TMCs for opposite directions of a divided highway do not always match. This is an issue when analyzing both directions of a highway.

- RID crash data are not conflated by direction. Therefore, when conducting analysis 2, both directions combined had to be used.

- The appropriate analysis unit for each case might not be evident. A subset of the Alignment feature class was created that covers the links for which the speed data were estimated. These alignments became the analysis units used in analysis 2. For analysis 1, combinations of adjacent links belonging to the same TMC were the analysis units. In this case, since neither crash data nor AADT were used in the analysis, the two directions of travel could be considered independently.

- For certain analyses, data may need to be processed or manipulated manually rather than through an automated process.

- Different approaches can be taken to define free-flow speeds or noncongested conditions. In this study, the definition from FHWA's *2014 Urban Congestion Trends: Improved Data for Operations Decision Making* report was used (i.e., 15th-percentile TT during traditional off-peak periods on weekdays: 9:00 a.m. to 4:00 p.m. and 7:00 to 10:00 p.m.).[18]

- For the analysis 1 results, one consideration is whether the "free-flow" speed assumption is a good one for the selected highways. Results showing V_{avg} and V_{85} below posted speeds could reflect the effect of traffic. That is, conditions that are assumed to be "free-flow" (noncongested) might include some non-free-flow (congested) conditions.

Other Lessons Learned

Other lessons learned include the following:

- The definition of an analysis unit has consequences with regard to the type of analysis to be performed. Selecting an appropriate analysis unit for the study is key.

- A small sample size can be an issue when examining the effect of very specific conditions on a particular facility type.

- Some results may initially seem unexpected or counterintuitive (e.g., the crash severity versus speed differential findings in analysis 2) and need further investigation to verify or refute.

- Data sources are updated periodically. NPMRDS v1 was used for this study; however, NPMRDS v2 was made available after most of the project was complete. While it was too late to use in this study, any future research involving the NPMRDS should use the new, enhanced data.

- Over time, conditions represented in static data sources (like the RID) may change.

BENEFITS OF LINKING THE RID WITH OTHER DATA SOURCES

An overarching benefit of linking the RID with other data sources is the potential for identifying temporal effects of various features. The RID is static in some ways because data were collected as a snapshot of a specific time period. However, many other data sources are updated continuously on 5-min/1-h/monthly/annual intervals. To a certain degree, linkage with the other data sources will enable time-sensitive analyses.

In conducting any analyses using the RID, it must be assumed that key roadway characteristics based on the RID remain unchanged (e.g., no changes in road alignments or speed limits). To the extent that these attributes may have changed over the time period covered by a linked database (e.g., crashes or NPMRDS TTs), the resulting analysis may lose some of its validity.

RECOMMENDED ENHANCEMENTS TO THE RID

Based on the project team's experience with linking the RID to another database and using the linked data for safety analyses, the following selected enhancements are recommended for the RID:

- Conflate crashes by direction: On divided highways, crashes are conflated mostly on one side of the highway. If crashes were conflated by direction, then crash-related analyses could be conducted for each direction separately. At present, both directions combined must be used.

- Relate the two-directional travelways of a divided highway through the RouteID (or other means).

- Link the RID Additional Supplemental Data geospatially with the RID GIS data warehouse.

FUTURE RESEARCH DIRECTIONS

The initial findings presented in this report are but a first step toward realizing the full potential of GIS linkages of innovative data sources with the RID.[1] This section discusses future research possibilities, including studies supported by the linked NPMRDS-RID data, the RID alone, and the RID linked with NDS data.[2,3]

ANALYTICAL POSSIBILITIES SUPPORTED BY THE LINKED NPMRDS-RID DATA

This section discusses potential future research using the linked NPMRDS-RID data.

Use of the New NPMRDS v2 for Future Studies

NPMRDS v1 was used to conduct this study. However, in February 2017, NPMRDS v2 was made available. The NPMRDS v2 dataset has several major improvements, including the following:[21]

- Additional data fields:
    - Data Density indicator for each dataset (passenger, truck, and combined), specifying the number of reporting vehicles (probes) contributing to the speed and TT record.

- Separate TTs on internal and external TMC paths.

- Faster availability (new data available within 5 business d after the end of each month).

- Greater coverage, including Interstate Trucks, Interstate All-Vehicles, and Non-Interstate All Vehicles.

- 14 HPMS attributes (including volume) conflated to the TMC network.[8]

Use of NPMRDS v2 would support a wider range of analyses and eliminate at least some of the challenges faced in using the NPMRDS v1 data.

Estimating Operating Speed by Using Traffic and Roadway Characteristics Data

Estimating operating speed could include investigating the influence of roadway geometry features on operating speeds on various facility types not covered in this study (e.g., rural two-lane highways, rural multilane highways).

Investigating Crashes in Noncongested Conditions Versus Congested Conditions in Large Urban Areas

Investigating crashes in noncongested conditions versus congested conditions in large urban areas would include addressing the following research questions:

- Is there a normal pattern for congested crashes (i.e., those occurring in congested traffic conditions at lower levels of service) versus noncongested crashes (i.e., those occurring in traffic conditions at higher levels of service)?

- Are specific road characteristics correlated with relatively higher levels of noncongested crashes (e.g., substandard geometry) or congested crashes?

- What countermeasures are suitable for targeting the different types of crashes?

Causes of noncongested crashes and congested crashes are expected to differ. It is reasonable to expect that the countermeasures for these crash types should also differ. If a location is often congested, the treatment for its safety concern may be focused on congestion relief. On the other hand, if substandard geometry results in a higher level of noncongested crashes, then countermeasures to improve the road geometry should be considered.

However, the crash-history data usually do not include information on the traffic conditions immediately preceding the crash event. NPMRDS TT data are collected every 5 min, offering a continuous view into road operating conditions. Combining NPMRDS TT data and RID crash-history data could support the identification of noncongested crashes versus congested crashes.

Investigating the Combined Effect of Horizontal Curve Length and Radius on Operating Speeds

It is unclear whether the results of analyses performed in this study would vary based on how the curves are distributed along the section. In this project, a 1-mi section consisting of a 0.75-mi tangent and a 0.25-mi-long curve with a 500-ft radius is considered equivalent (with regard to the weighted degree of curvature) to a 1-mi section with a 0.5-mi tangent and a 0.5-mi-long curve with a 1,000-ft radius. However, it is unclear whether that is a good assumption. Future research could test that by examining the combined effect of curve length and radius.

ANALYTICAL POSSIBILITIES SUPPORTED BY THE RID

This section summarizes analytical possibilities supported by the RID, without and with linkage to the NDS.[3]

Analytical Possibilities Using the RID Without Linkage to the NDS

The RID represents a comprehensive set of roadway-inventory databases with geospatial properties that make them unique compared to most of the other available roadway-inventory databases. The LRS embedded in these geodatabases makes it relatively easy to combine features from different feature classes using dynamic segmentation.

Analytical possibilities for which the RID may be used can be categorized in the following groups of analyses:

- Micro-Level safety analysis:
  - Calibrating existing HSM crash prediction models.[12]
  - Revising crash modification factors (CMFs) in the existing crash prediction models.
  - Developing new CMFs for existing crash prediction models.
  - Developing new safety performance functions (SPFs).
  - Applying the HSM Part B Process (Micro-Level).[12]
  - Identifying or estimating highway safety surrogate measures.
  - Assisting efforts to conduct Road Safety Audits (RSAs).
  - Testing the sensitivity of HSM predictive methods.[12]

- Macro-Level safety analysis:
  - Applying the HSM Part B Process (Macro-Level).[12]
  - Evaluating the safety of transportation systems in the planning stage.
  - Performing large-scale assessment of risk factors.

- Alignment extraction models.

Micro-Level Safety Analysis

Subsets of the RID that contain more detailed data can support micro-level safety analysis. These subsets can provide researchers with rich cross-sectional data that can be used in modeling, calibration, and fine tuning of existing models. There is strong skepticism among many highway safety researchers about using cross-sectional data. However, most agree that this type of data can be used successfully for modeling purposes if certain considerations are met. Examples of these considerations can be found in "Recommended Protocols for Developing Crash Modification Factors," which is part of the National Cooperative Highway Research Program (NCHRP) Project 20-07, Task 314 report.[22]

The project team has identified the following examples of micro-level analyses that can use these cross-sectional databases.

Calibration of Existing HSM Crash Prediction Models[12]

HSM part C contains crash prediction models for different types of highway facilities. The general structure of these mathematical models includes an SPF and a series of CMFs.[12] These models are applied to sites (i.e., homogenous highway segments/ramps, intersections, or ramp terminal intersections). The SPFs are functions of AADT and segment length (for highway segments) and predict the number of crashes for the site for various roadway characteristics under a set of baseline conditions (e.g., lane width = 12 ft, and shoulder width = 6 ft for rural two-lane highways). The SPFs are multiplied by the CMFs to adjust the prediction for deviation of the sites from the baseline conditions. Appropriate use of the HSM crash prediction models also requires calibration.[12] The HSM recommends that each jurisdiction develop calibration factors. This process requires applying the HSM models to a relatively large set of highways/sites. Table 9 shows the number of datasets needed for calibration of crash prediction models by facility type.

Table 9. HSM part C models that need calibration.[12]

Facility Type	Segment Models	Intersection/ Ramp Terminal Models
Rural two-lane highways	1	3
Rural multilane highways	2	3
Urban/suburban arterials	5	4
Freeway segments	4	-
Freeway speed-change lanes	4	-
Collector–Distributor roads	4	-
Freeway entrance ramps	4	-
Freeway exit ramps	4	-
Ramp terminals	-	4

-No data.

Calibration factors may be developed for each of the States for which the RID was created. The highway/intersection/ramp terminal data relevant to each of the HSM models plus crash data are used in this process.[12] The appendix to HSM part C lists data elements used for calibrating each crash prediction model. Data are split into two categories: Required and Desired. The Required data are needed for a valid calibration. The HSM provides default values for Desired data elements when actual data are not available.[12] Crash data must also be available for the calibration sites. In this process, the relevant model is applied to the data for all sites of the same type, and the total predicted number of crashes is calculated. The calibration factor is then the ratio of the number of observed crashes to the number of crashes predicted for all sites.

Revising the CMFs of the Existing Crash Prediction Models

This is an open discussion in the HSM community.[12] If there is any process that can be used to revise an existing CMF, the RID is an alternative database that could be used for this purpose. This process can be conducted not just for the RID States but for all States by using the entire RID since the six RID study sites collectively represent a reasonable cross section of the United States.

Development of New CMFs for Existing Crash Prediction Models

Many of the HSM crash prediction models lack CMFs for certain highway/intersection features that seemingly are significant enough to affect safety.[12] Examples of such features are horizontal curvature and superelevation for rural multilane highways and pavement friction and weather for all highway segment models. (An example was presented at the 94th Transportation Research Board annual meeting.)[23] The feature of interest may already be part of the RID or may be added to the RID by linking it to other databases.

Development of New SPFs

The RID can be used to develop new SPFs for the HSM crash prediction models.[12] The diversity of the areas represented by the six SHRP2 States will likely provide a better national representation than the existing models. The following steps in developing SPFs are outlined in the FHWA document, *Safety Performance Function Decision Guide: SPF Calibration versus SPF Development*:[24]

1. Identify facility type.
2. Compile necessary data.
3. Determine functional form.
4. Develop SPF.
5. Conduct model diagnostics.
6. Re-estimate SPF.

The roadway/intersection characteristics affecting crashes the most depend on the type of facility. The current HSM SPFs are functions of AADTs (and the length for segments), developed with data from sites with baseline conditions.[12] Almost all of the AADT relationships are nonlinear and in exponential form. For each State, RID data with baseline conditions relevant to each SPF can be used to adjust the AADT exponents. Some States are already doing this using other sources of data.

Applying the HSM Part B Process (Micro-Level)

HSM part B ("Roadway Safety Management Process") presents procedures for monitoring and reducing crash frequency on existing roadway networks. The following six steps of this process are described in the HSM, chapters 4 through 9:[12]

- Network Screening—reviewing a transportation network to identify and rank sites based on the potential for reducing average crash frequency.

- Diagnosis—evaluating crash data, historic site data, and field conditions to identify crash patterns.

- Select Countermeasures—identifying factors that may contribute to crashes at a site and selecting possible countermeasures to reduce the average crash frequency.

- Economic Appraisal—evaluating the benefits and costs of potential countermeasures and identifying individual projects that are cost effective or economically justified.

- Prioritize Projects—evaluating economically justified improvements at specific sites and across multiple sites to identify a set of improvement projects to meet objectives such as cost, mobility, or environmental impacts.

- Safety Effectiveness Evaluation—evaluating effectiveness of a countermeasure at one site or multiple sites in reducing crash frequency or severity.

For processes where the number of crashes is predicted or changes in the predicted number of crashes due to safety countermeasures are evaluated, crash prediction models or CMFs/functions are used. These models/factors/functions are developed as functions of highway characteristics— either based only on AADTs (and length, for highway segments) or as functions of more detailed characteristics of the highway facility. The latter case can be considered a micro-level analysis, with the former more of a macro-level analysis (see "Macro-Level Safety Analysis").

The RID can be used to apply the micro-level version of HSM part B to the six States for which these data are collected (i.e., to conduct network-level quantitative safety evaluation using crash

prediction models developed based on highway geometric characteristics and AADTs).[12] In this process, HSM part C crash prediction models—or any other set of crash prediction models that are developed based on detailed highway characteristics (versus average characteristics that would be used in a macro-level analysis)—can be used in conjunction with the data available in these databases to estimate crashes for highway facilities.[12] The CMFs that represent the safety effects of countermeasures can also be used with these data to estimate the benefits of using a countermeasure in terms of reduction in expected crashes.

Highway Safety Surrogate Measures

Surrogate measures are used in qualitative or quantitative safety evaluation of highways and intersections. The RID can be a suitable source of data for identifying or estimating these surrogate measures. In this process, the correlation of highway or intersection characteristics with crash data is studied, and the characteristic most strongly correlated with crash data is selected as a surrogate measure. There should also be a validation process to confirm this surrogacy. In some cases, a mathematical relationship between the surrogate measure and crashes can be established. When this is not possible, threshold values might be established for the surrogate measure to identify critical situations from a safety perspective. One example surrogate measure is the differential between the speed on a curve and the speed on its preceding tangent on rural two-lane highways, which is strongly correlated with the safety of these curves.[25]

RSA

An RSA is a proactive qualitative process that is an important component of many States' highway safety programs. The RID, in conjunction with its video data (or still images matching the Location data points), can be used in the initial stages of the RSA process. A good example of using still images along with highway segments and intersections data can be found in the Pennsylvania Department of Transportation Video Log.[26]

Test Sensitivity of HSM Predictive Method

The predictive method in the HSM relies on fairly detailed knowledge of roadway alignment and cross-sectional features.[12] In many cases, practitioners are required to have functional layouts of roadways in order to apply the predictive method. Research could be conducted to explore whether the RID and associated crash data could be used to test the sensitivity of the predictive method of different roadway functional classifications to all of the CMFs associated with the functional classification. This would reveal whether simplified predictive equations could be available for predesign alternatives analysis.

Macro-Level Safety Analysis

The project team has identified the following types of macro-level safety analysis models that can take advantage of the RID.

Applying the HSM Part B Process (Macro-Level)

The RID can be used to apply the macro-level version of HSM part B to the six States for which these data are collected (i.e., to conduct network-level quantitative safety evaluation using crash prediction models developed based on average values for highway geometric characteristics and

AADTs).[12] In this process, HSM part C crash prediction models—or any other set of crash prediction models that are developed based on average highway characteristics (versus detailed characteristics that would be used in a micro-level analysis)—can be used in conjunction with the data available in these databases to estimate crashes for highway facilities.

SafetyAnalyst™ (AASHTOWare®, an American Association of State Highway and Transportation Officials application) is a software implementation of HSM part B, which contains several macro-level crash prediction models developed based on average values of highways' geometric characteristics and AADTs.[27,12] The SafetyAnalyst™ tool can use the RID as the source of data for validation and/or calibration.

Safety Evaluation of Transportation Systems in the Planning Stage

The NCHRP Report 546, *Incorporating Safety Into the Long-Range Planning Process*, demonstrated that crashes could be predicted using traditional transportation planning data, such as socioeconomic data, transportation network data, data on travel demand, and surrounding land use data.[28] The target audience for this research was MPOs and large cities or counties with travel demand models and modeling expertise. Conceptually, if crashes are shown to be sensitive to land use and transportation network decisions, then funding priorities and transportation land use and planning policies can be evaluated and established to reduce crash frequency and severity.

The RID is a rich dataset of crash, roadway-geometry, and traffic-volume data that may be integrated with travel demand model data and/or outputs to estimate planning-level crash prediction models. Variables such as vehicle miles traveled by roadway functional classification, intersection type and density, speed limit, or miles of roadway by functional classification have been shown to influence crash frequency and severity. Data could be used at city, county, regional, or statewide scale and researched for usefulness.

Large-Scale Assessment of Risk Factors

Systemic safety analysis considers crash history on an aggregate basis to identify high-risk roadway characteristics. Systemic safety analysis is a complement to spot-specific network screening analysis. While traditional spot-specific safety analysis results in safety investments at high-crash locations, the systemic approach leads to widespread implementation of projects to reduce the potential for severe crashes. The systemic approach to safety involves improvements that are widely implemented based on high-risk roadway features correlated with particular severe crash types or frequencies.

For example, the data on crash and roadway characteristics can be analyzed to identify that, in a rural environment, crashes are overrepresented on curves with a particular radius in a particular traffic volume range, or in an urban environment, pedestrian crashes are overrepresented at intersections with particular characteristics (e.g., number of lanes, signal control, number of approaches). The analysis is a simple comparison of crash frequency with incidence of the particular roadway characteristics under consideration. Where the frequency greatly exceeds the incidence of the characteristic(s), the characteristic(s) could be identified as a risk factor. Figure 27 demonstrates a systemic comparison of crash frequency and severity to rural interstate mileage. The graph shows distribution of crashes and roadway mileage by traffic volume on

rural interstates. In this example, roadway segments carrying more than 15,000 vehicles per d (vpd) experience a higher level of single-vehicle roadway departure crashes than would be expected based on the mileage of these facilities. So, a treatment (countermeasure) that is "proven" to reduce single-vehicle roadway departure crashes could be applied on all rural interstates with average daily traffic values greater than 15,000 vpd.
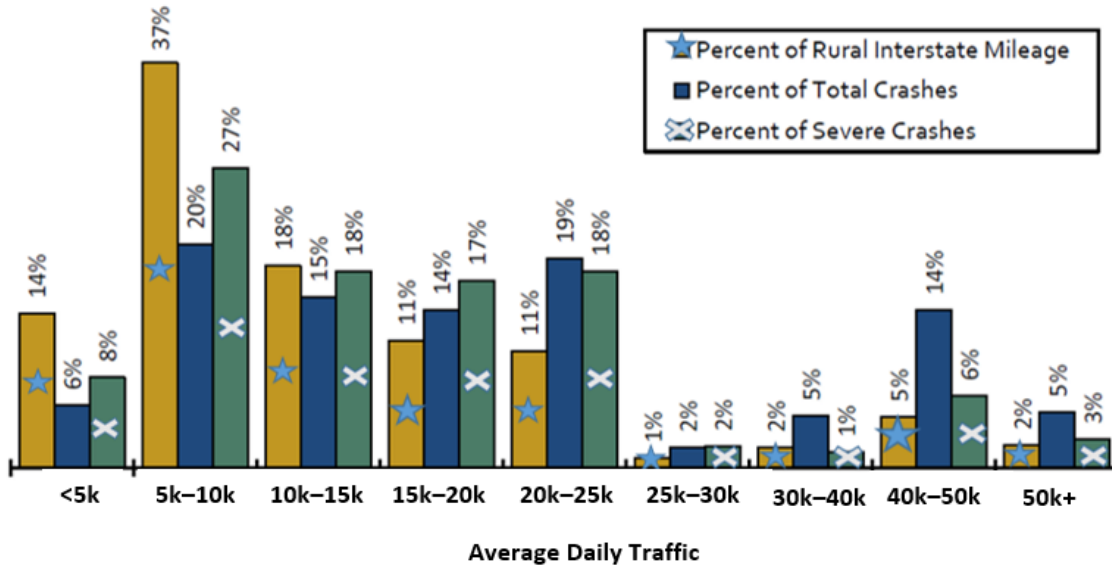
Source: FHWA.
k = thousand.

Figure 27. Graph. Single-vehicle roadway departure crashes on rural interstates.

The geometric, crash, and traffic volume data in the RID could be used to conduct a large-scale systemic safety analysis to identify risk factors and provide recommendations for common treatments for consideration by practitioners. Depending on the geographic extent of the analysis and the subsequent outcomes, it may be possible to identify results that can be considered typical and thus reduce the need for analyses at smaller geographic scales.

In summary, RID data could be used to determine if there are common risk factors that lead to common countermeasures that could be deployed without conducting extensive analyses.

Alignment Extraction Models

Data collection via instrumented vehicles has been used for asset management (e.g., for tracking pavement condition). The need for accurate road geometry has prompted vendors to derive and provide geometric data from the collected road GPS information. Of interest is how these data might be leveraged to populate the datasets required to quantitatively evaluate the safety of highways. However, the quality of the alignment extracted from these types of data is rarely evaluated. The RID includes alignment data that are the result of an automated postprocessing of the GPS data. Although the SHRP2 Safety Program defined minimum requirements for the extracted alignment, these requirements are not of high precision, and the resultant data may not be appropriate (i.e., not accurate enough) for certain types of analyses.[4] However, the original mobile-van data (not currently distributed with the RID) could be used as input data to generate more precise alignment data for the highways surveyed by the instrumented vehicles.

Analytical Possibilities Using the RID With Linkage to the NDS

Analytical possibilities for which the RID may be used in conjunction with the NDS data can be categorized into the following groups of analyses:[3]

- Crash and near-crash studies:
    - Near crash as a surrogate measure of safety.
    - Estimating number of crashes from predicted number of near crashes.

- Calibrating and developing simulation models:[29]
    - Interactive Highway Safety Design Model (IHSDM) Design Consistency Module (DCM).
    - IHSDM Driver/Vehicle Module (DVM).
    - IHSDM Traffic Analysis Module (TAM).
    - Expansion of the IHSDM DCM.
    - Expansion of the IHSDM DVM.
    - Crash/near-crash simulation.

Crash and Near-Crash Studies

Crashes are rare events, and databases (e.g., the RID) that contain roadway and traffic data associated with crashes are often inflated with values of 0 for crashes. This is the source of many limitations and problems in modeling crashes based on the roadway and traffic characteristics. The NDS data contain both crash events as well as near-crash events.[3] The number of near-crash events is significantly larger than the number of actual crashes, and this may provide an opportunity for researchers to use the number of near crashes as the dependent variable using the RID and NDS data in the following study types.[3]

Near-Crashes Events As a Surrogate Measure of Safety

Near-crash prediction models can be developed in the same ways crash prediction models are currently developed. This requires identifying the location of the NDS near-crash events on the RID.[3] Near-crash events similar to crash events can be assumed to have a Poisson distribution. By predicting near crashes for different alternatives of a project as a safety surrogate measure, the alternatives can be compared and ranked.

Estimating Number of Crashes From Predicted Number of Near Crashes

Near-crash prediction models can be developed as described in the previous section. Crashes can then be estimated by conducting a stochastic analysis based on relations between near crashes and crashes. A similar study was conducted to estimate the number of road-departure crashes based on the number of encroachments estimated from a driver/vehicle simulation program.[30]

Calibrating and Developing Simulation Models

Many existing simulation models can benefit from RID and NDS data for calibration.[3] There is also an opportunity for developing new types of simulation models related to highway safety using these databases. The project team has identified the following opportunities.

IHSDM DCM

The DCM estimates V_{85} of a free-flow vehicle along two-lane rural highways. The speed differential between a horizontal curve and its preceding tangent is identified as a safety surrogate measure for this type of facility. The RID, in conjunction with the NDS database, can be used to calibrate this model.[3] Full access to the NDS data is required for this study. Trips made on rural two-lane highways with available horizontal alignment data should be identified. Video and radar data of these trips are needed to identify a subset of trips that can be considered free-flow. Travel speeds on tangents and curves will be the main components of the data used in the calibration of the DCMs in conjunction with the speed-limit and roadside hazard rating. The project team believes the mobile-van Alignment data are of sufficient quality, from a curvature-precision perspective, to be used for this purpose.

IHSDM DVM

The DVM is a microsimulation model that simulates the interactions between the driver, vehicle, and road on rural two-lane highways in a single vehicle driving along the road. The NDS data needed for calibrating the DVM, similar to the data needed for the DCM, are a subset of trips made on rural two-lane highways with available horizontal alignment data while in free-flow traffic.[3] However, more data from both the RID and NDS database are needed for calibrating this simulation model compared to what are needed for calibrating the DCMs.[3] Also, it is not clear whether the mobile-van Alignment data are precise enough with regard to curvature characteristics. It is possible that, for this purpose, more precise alignment data will need to be extracted from the original mobile-van Location data.

IHSDM TAM

The TAM is a microsimulation traffic model developed for rural two-lane highways. The type of data proposed for use with the IHSDM DVM can also be used for validation of this model for unimpeded traffic flow.[29] For full traffic flow, including both impeded and unimpeded situations, the flow rate needs to be estimated.

Expansion of the IHSDM DCM

The safety surrogate measure of tangent-to-curve speed differential is currently applicable to only rural two-lane highways in the IHSDM.[29] The RID, in conjunction with the NDS data, can be used to expand the estimation of V_{85} to other facility types and to investigate whether the tangent-to-curve speed differential can also be used as a safety surrogate measure for other facility types.[3]

Expansion of the IHSDM DVM

The current DVM can only simulate the interactions between the driver, vehicle, and roadway on rural two-lane highways (for a single vehicle). The RID, in conjunction with the NDS data, provides an opportunity to expand this simulation to other types of facilities.[3] The two-lane urban facility type is the most promising because there is no lane-change action involved, and the main functions would be very similar to the current DVM.

Crash/Near-Crash Simulation

Current simulation models that imitate the interactions between the driver, vehicle, and roadway (e.g., the IHSDM DVM) stop when some simulated parameters pass certain thresholds that are representative of a failure of some sort.[29] As an example, the DVM simulation stops when any of the tires leave the road and the lane offset value passes a certain threshold. In real driving situations, this is identified as a near crash, or—much less frequently—a crash. In other words, the DVM, like many other simulation models, does not simulate near-crash or crash events. The RID, in conjunction with the NDS data, provides an opportunity to simulate different types of near-crash and crash events.[3] The simulation of single-vehicle departure near-crash and crash events associated with encroachment might be a good place to start. For this purpose, it is possible that more precise alignment data will need to be extracted from the original mobile-van Location data.

APPENDIX: DETAILS OF ADDITIONAL DATA PREPARATION FOR THE ANALYSES

Additional preparation of the linked data for conducting the analyses includes the following:

- Limiting the network to urban interstates and major arterials.
- Reflecting speed data on the network and building the case study geodatabase.
- Preparation of alignment data.
- Preparation of crash data.
- Preparation of speed data and putting all data together.

Each data preparation step is described in detail in this appendix.

LIMITING THE NETWORK TO URBAN INTERSTATES AND MAJOR ARTERIALS

The definition of free-flow conditions applied to the NPMRDS data in this study is from the FHWA report, *2014 Urban Congestion Trends: Improved Data for Operations Decision Making*.[2,17] Based on this definition, the network under study is limited to interstate highways and major arterials in urban areas. Development of the WA_Tops_ Urban_Int_Art feature class (i.e., urban interstates and arterials) is described in the section Process to Build Urban Interstate and Major Arterials Network. This feature class is used to filter the speed data for further analysis.

The initial plan was to use FunctionalClass_SR (RID) in the development of the WA_Tops_ Urban_Int_Art feature class.[1] However, further investigation showed that FunctionalClass_SR is missing significant lengths of urban interstates in the WA database. Although this was reported to the RID developer and corrected, the correction was completed after development of the Washington State urban interstates and arterials feature class (WA_Tops_Urban_Int_Art) for this study.

The project team initially considered using only the "oneway" attribute in Routes to target one-way highways representing divided highways. However, some problems were discovered with the conflation of the HPMS data into Expressways in the WA RID.[8] Therefore, the plan was changed to filter urban interstates and major arterials.

Process to Build the Urban Interstate and Major Arterials Network

The process for building the urban interstate and major arterial network is summarized as follows:

1. Add the Tops feature class to the geodatabase and name it WA_Tops.

2. Select records of WA_Tops with Urban Code < 99999 and F_System = 1 and 2 (i.e., urban interstates, principal arterials, and other expressways). A total of 720 records are selected.

3. Create a new feature class from the selection made in step 2 and name it WA_Tops_Urban_Int_Art.

REFLECTING SPEED DATA ON THE NETWORK AND BUILDING THE CASE STUDY GEODATABASE

This section describes the process for the following:

- Reflecting the speed data contained in the Links_Positive_Speeds and Links_Negative_Speeds feature classes on the WA_Tops_Urban_Int_Art.

- Building a subset of the Alignment for the case study analyses.

In addition to the RID feature classes, this process also uses the Alignment data from mobile vans (original alignment data) to relate the two directions of travel of the same highway. The last step is to create a new geodatabase for use in the case study analyses. The steps in the process can be summarized as follows:

1. Select features in the Routes feature class with attribute ONEWAY = 1.

2. Select features in Links_Positive_Speeds and Links_Negative_Speeds by location, with Routes having ONEWAY = 1 as the source layer (4,891 of 23,339 features and 4,829 of 23,245 features are selected, respectively).

3. Create new feature classes from the above features and name them Links_Positive_Speeds_1_W and Links_Negative_Speeds_1_W.

4. Select features in Links_Positive_Speeds_1_W and Links_Negative_Speeds_1_W by location with the source layer of WA_Tops_Urban_Int_Art and a 100-ft threshold distance. (This threshold—and others noted below—was arbitrarily selected as a reasonable distance by the project team.) Create new feature classes from the selections and name them Links_Positive_Speeds_1_W_Urban and Links_Negative_Speeds_1_W_Urban (2,572 and 2,029 features, respectively).

5. Select features of Links_Positive_Speeds_1_W_Urban and Links_Negative_Speeds_1_W_Urban by location. Use each as the source layer for the other, with a 100-ft threshold distance. Save the selections as new feature classes and name them Links_Positive_Speeds_1_W_Urban_Pair and Links_Negative_Speeds_1_W_Urban_Pair (1,815 and 1,829 features, respectively).

6. Import the Alignment feature class and name it WA_Alignment. Select features in WA_Alignment by location (with a 30-ft threshold distance) with Links_Positive_Speeds_1_W_Urban_Pair and Links_Negative_Speeds_1_W_Urban_Pair as source layers. Create new feature classes from the selections and name them WA_Alignment_Links_P_S_1_W_U_P and WA_Alignment_Links_N_S_1_W_U_P (1,071 and 1,049 features, respectively).

7. Make a copy of the WA_Alignment_Links_P_S_1_W_U_P feature class and name it WA_Alignment_Links_P_and_S_1_W_U_P. Then, append WA_Alignment_Links_N_S_1_W_U_P to it. This includes all alignment feature classes of interest (2,120 features).

8. Outside ArcGIS®, by using a database program, create a reduced version of the mobile-van Alignment table containing only Route, Fname, Visi_from, Visi_to, vis_dir, and SourceID.[17] The SourceID field in these data can be used to build relationships between these data and RID Alignment data.

9. Import mobile-van reduced Alignment data as a table and name it Fugro_Alignment_2012_2013.

10. Create a geodatabase named Final Case Study with only a few feature classes. The initial layers in this geodatabase are the following:

    o WA_Routes.
    o WA_Alignment.
    o WA_Alignment_Links_P_and_S_1_W_U_P.
    o Fugro_Alignment_2012_2013.

PREPARATION OF ALIGNMENT DATA

The following steps describe the final preparation of alignment data for the analyses:

1. Conduct Add Join of the fields of WA_Alignment_Links_P_and_S_1_W_U_P" to Fugro_Alignment_2012_2013. Use SourceID values, keeping only the matched rows. Save as a new table in the Final Case Study geodatabase (Final_Alignment).

2. Conduct Make Route Event Layer, selecting WA_Routes and ROUTEID as the Input Route Feature and Identifier on the Final_Alignment table and build the Final_Alignment feature class.

3. Import the Final_Alignment feature class to the geodatabase and name the new feature class Original_Alignment_Feature_Class.

4. Import the Original_Alignment_Feature_Class and name it Final_Alignment_Feature_Class. This is the layer in which the alignments for the analyses are kept; delete the remainder.

5. Sort the Final_Alignment_Feature_Class by Route, Fname, Visi_from, and Visi_to, respectively.

6. Identify the routes for analysis unit selection visually. Table 10 shows the selections, which are defined as the following:

    o The Route number is from the mobile-van data and matches the highway numbers in WA.

- o Section shows pieces of the routes that have alignment and speed data available for both directions of travel.

- o Dir 5 and Dir 6 show the two directions for sections of the same highway, as surveyed by instrumented vans (in the mobile-van data).

- o Visi_from and Visi_to are attributes from the mobile-van Alignment data to help visually select the two directions of travel that (with some approximation) represent the same section of highway.

Table 10. Analysis units.

Route	Section	Dir = 5; Visi_From	Dir = 5; Visi_to	Dir = 6; Visi_From	Dir = 6; Visi_to
3	1	276	7104	24	6880
3	2	9280	11940	9072	11760
3	3	14784	19192	14532	18984
5	1	25551	29015	28637	32157
5	2	59969	73087	60027	72759
5	3	86643	104979	87963	104739
5	4	106555	123803	107167	123799
5	5	124079	127303	124067	127307
5	6	146455	147735	146435	147963
5	7	151955	152163	152123	152791
5	8	162923	166783	163247	166839
5	9	179139	179875	179159	179879
5	10	182683	183759	182907	183771
5	11	183939	188219	183955	188255
5	12	212911	220621	212693	220412
16	1	656	4548	14132	17984
16	2	7056	10948	20544	24404
16	3	24760	26994	24644	26871
18	1	960	2200	980	2244
18	2	2908	4992	2936	5016
18	3	6548	8072	6576	8076
18	4	8460	13556	8464	13536
99	1	18540	19876	18366	19726
99	2	22434	23910	22362	23894
99	3	24474	25754	24454	25730
99	4	26154	27342	26134	27322
101	1	90	5660	-64	5527
167	1	17974	21370	17834	21374
167	2	24374	26386	24642	26634
405	1	2028	4776	1956	4708
405	2	4940	9660	4872	9660
410	1	560	2968	266	2706
410	2	3128	4288	2874	4030
512	1	81	8524	-389	8048
512	2	9856	11572	9360	11076
520	1	560	6900	396	6733
520	2	9104	11184	8909	10977
520	3	11576	12692	11369	12513
522	1	13592	17584	12992	17132
526	1	768	4128	776	4144

7. Some gaps are present for one direction in the data represented by table 10. To fill these gaps and include them in the alignment subset, make a copy of the table Fugro_Alignment_2012_2013 and name it Fugro_Alignment_2012_2013_Subset. Conduct a selection by attributes and keep only the items within the ranges shown in table 10. Then, save the table.

8. ResourceID is used to relate data in the mobile-van and RID alignments. However, this field is of type Double in the mobile-van alignment and of type TEXT in the RID alignment. Add a TEXT field to the Fugro_Alignment_2012_2013_Subset table and name it New_SourceID. Calculate the values by using the ResourceID (of type Double).

9. Join the Fugro_Alignment_2012_2013_Subset table to the WA_Alignment using ResourceID and New_SourceID, keeping only the matched items. Conduct a Select All and save the selection as a new feature class on the map named WA_Alignment_Subset. In the Catalog, import the feature class and name it WA_Alignment_Subset. Delete the old WA_Alignment_Subset from the map. The new feature class has 1,157 records.

10. Parts of Routes 5 and 16, Dir 6, have two sets of records. Therefore, portions of these routes appear in the WA_Alignment_Subset without Dir 5 data. Delete these sections manually from the feature class. The result has 1,125 features; this is the basis for the Alignment used in subsequent steps.

11. Create WA_Alignment_Subset_Dir_5, which is a subset of WA_Alignment_Subset with Dir = 5.

PREPARATION OF CRASH DATA

WA RID crash data for 2011 to 2013 are used in this study. For divided highways, crashes are conflated mostly on one direction of the highway. Therefore, to use WA RID crashes for any evaluation, both directions of the highway must be combined.

Preparation of crash data includes the following steps:

1. Import crash data feature classes (both LRS and Other) for the period 2011 to 2013 into the geodatabase. Combine all crash data in a single-feature class named Crash_2011_2013_Reduced. Make a selection by location on Crash_2011_2013_Reduced, intersecting WA_Alignment_Subset (by choosing the "are within the source layer features" option).

2. Create a new feature class of the selected crashes in Catalog and name it Crash_2011_2013_Reduced_Subset. This has 36,832 crashes. Crashes of the study network are among these crashes.

3. Create a text file from the Crash_2011_2013_Reduced table with all 36,832 crashes and corresponding RID RouteID and FrMeasure values showing the location of these

crashes on the RID network as well as columns that show the crash severities. Name this file Crash.dat.

4. Create a text file from WA_Alignment_Subset containing the sections of these alignments on both sides and their corresponding RID RouteID, FrMeasure, and ToMeasure values. Name this file Road.dat.

5. Develop a C programming code to identify crashes in Crash.dat that are related to sections contained in Road.dat. Identify the number of total and FI crashes for each section of highway in Road.dat.

PREPARATION OF SPEED DATA AND PUTTING ALL DATA TOGETHER

The RID WA_SpeedLimit_Reduced feature class contains many records with nonvalid speed-limit values. Delete the invalid records to produce a WA_SpeedLimit_Reduced_non_Zero feature class (11,741 of 268,550 records remain).

Conduct Dynamic Segmentation on the table produced from the WA_Alignment_Subset feature class as the Input Event Table and tables produced from the WA_SpeedLimit_Reduced_non_Zero, Links_Positive_Speeds_1_W_Urban_Pair, and Links_Negative_Speeds_1_W_Urban_Pair feature classes that contain speed limits and operating speeds as event tables. Use RouteID set as Measure Field, FrMeasure set as from Measure Field, and ToMeasure set as to Measure Field. The steps for conducting these dynamic segmentation processes are summarized as follows:

1. Open the attribute tables for all four feature classes (WA_Alignment_Subset, WA_SpeedLimit_Reduced_non_Zero, Links_Positive_Speeds_1_W_Urban_Pair, and Links_Negative_Speeds_1_W_Urban_Pair) one by one. Through the Export operation and by choosing File Geodatabase tables as type, create one table from each of these feature classes. Name these files WA_Alignment_Subset_Table, Links_Positive_Speeds, Links_Negative_Speeds, and WA_SpeedLimit_Reduced_non_Zero_table.

2. Conduct Overlay Route Events operations with INTERSECT selected as the OverlayType (from Linear Referencing Tool). Choose the WA_Alignment_Subset_Table as the Input Event Table and WA_SpeedLimit_Reduced_non_Zero_table as the Overlay Event Table to produce WA_Align_SL_Speed_Intersect.

3. Conduct two Overlay Route Events operations with INTERSECT selected as the OverlayType (from Linear Referencing Tool). Choose WA_Align_SL_Speed_Intersect as the Input Event Table and Links_Positive_Speeds and Links_Negative_Speeds as the Overlay Event Tables. Produce WA_Align_SL_P_Speed_Intersect and WA_Align_SL_N_Speed_Intersect tables. These two tables carry the speed limit and operating speeds for links of different directions of the chosen highways. A majority of links in WA_Align_SL_P_Speed_Intersect are Dir 5, and a majority of links in WA_Align_SL_N_Speed_Intersect are Dir 6.

The last two tables produced in this process (WA_Align_SL_P_Speed_Intersect and WA_Align_SL_N_Speed_Intersect) are used in the analyses. They contain all necessary information, including the following:

- Alignment information from the RID Alignment feature class.

- Speed-Limit information from the RID WA_SpeedLimit_Reduced feature class.

- V_{85} and V_{avg} from the NPMRDS (averaged along TMCs, dynamically segmented over Alignment records).

- Route, Fname, Visi_from, Visi_to, and Vis_dir from the mobile-van data. (Values of Visi_from and Visi_to are not dynamically segmented because these values are not the source of segmentation. As a result, some records have the same values but are actually related to different adjacent sections.)

In these files, the RouteID, FrMeasure, and ToMeasure columns identify the actual segments of highways, and vis_dir with values of 5 or 6 determines the directions of travel.

Note that, for analyses that require additional RID data, further dynamic segmentation can be conducted to add such data to the above.

Figure 20 shows the new geodatabase created after conducting the above data preparation for the analyses.

REFERENCES

1. CTRE. (2017). "SHRP 2 - Roadway Information Database." (website) Iowa State University, Ames, IA. Available online: http://www.ctre.iastate.edu/shrp2-rid/, last accessed December 20, 2017.

2. FHWA. (2017). "Operations Performance Measurement." (website) U.S. Department of Transportation, Federal Highway Administration, Washington, DC. Available online: https://ops.fhwa.dot.gov/perf_measurement/, last accessed December 20, 2017.

3. TRB. (2017). "InSight Data Access Website: SHRP 2 Naturalistic Driving Study (NDS)." (website) Virginia Tech Transportation Institute, Blacksburg, VA. Available online: https://insight.SHRP2nds.us/, last accessed December 20, 2017.

4. TRB. (2017). "SHRP 2." (website) The National Academies of Sciences, Engineering, and Medicine, Washington, DC. Available online: https://www.trb.org/StrategicHighwayResearchProgram2SHRP2/Pages/The-SHRP-2-Naturalistic-Driving-Study-472.aspx, last accessed December 20, 2017.

5. Esri®. (2018). "StreetMap Premium for ArcGIS." (website) Esri®, Redlands, CA. Available online: http://www.esri.com/data/streetmap, last accessed May 15, 2018.

6. Smadi, O., Hawkins, N., Hans, Z., Bektas, B.A., Knickerbocker, S., Nlenanya, I., Souleyrete, R., and Hallmark, S. (2015). *Naturalistic Driving Study: Development of the Roadway Information Database,* SHRP2 Report S2-S04A-RW-1, Transportation Research Board, Washington, DC. Available online: https://www.nap.edu/read/22261/chapter/1, last accessed May 15, 2018.

7. CTRE. (2018). "SHRP 2 (S04A) Roadway Information Database Development and Technical Coordination and Quality Assurance of the Mobile Data Collection Project (S04B)." (website) Iowa State University, Ames, IA. Available online: http://www.ctre.iastate.edu/research/detail.cfm?projectID=-333008524, last accessed May 15, 2018.

8. FHWA. (2018). "Highway Performance Monitoring System (HPMS)." (website) U.S. Department of Transportation, Federal Highway Administration, Washington, DC. Available online: https://www.fhwa.dot.gov/policyinformation/hpms.cfm, last accessed May 15, 2018.

9. HERE®. (2018). "HERE." (website) HERE, Amsterdam, Netherlands. Available online: https://www.here.com/en, last accessed May 15, 2018.

10. FHWA. (2018). "National Highway System." (website) U.S. Department of Transportation, Federal Highway Administration, Washington, DC. Available online: https://www.fhwa.dot.gov/planning/national_highway_system/, last accessed May 15, 2018.

11. Donnell, E.T., Hines, S.C., Mahoney, K.M., Porter, R.J., and McGee, H. (2009). *Speed Concepts: Informational Guide,* Report No. FHWA-SA-10-001, U.S. Department of

Transportation, Federal Highway Administration, Washington, DC. Available online: https://safety.fhwa.dot.gov/speedmgt/ref_mats/fhwasa10001/, last accessed December 20, 2017.

12. AASHTO. (2010). *Highway Safety Manual, 1st Edition*, American Association of State Highway and Transportation Officials, William Morrow, New York City, NY.

13. Campbell, K.L. (2012). "The SHRP 2 Naturalistic Driving Study: Addressing Driver Performance and Behavior in Traffic Safety." *TR News*, *282*, pp. 30–36, Transportation Research Board, Washington, DC. Available online: https://insight.shrp2nds.us/documents/shrp2_background.pdf, last accessed December 20, 2017.

14. TRB. (2018). "Roadway Data Elements Collected." (website) Transportation Research Board, Washington, DC. Available online: http://onlinepubs.trb.org/onlinepubs/SHRP2/RoadwayDataElements.pdf, last accessed April 20, 2018.

15. TRB. (2017). "Recorded Safety Webinars." (website) The National Academies of Sciences, Engineering, and Medicine, Washington, DC. Available online: http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/SafetyWebinars.aspx, last accessed December 20, 2017.

16. FHWA. (2012). *Manual on Uniform Traffic Control Devices for Streets and Highways,* U.S. Department of Transportation, Federal Highway Administration, Washington, DC.

17. Esri®. (2018). "About ArcGIS." (website) Esri®, Redlands, CA. Available online: https://www.esri.com/en-us/arcgis/about-arcgis/overview, last accessed May 15, 2018.

18. FHWA. (2015). *2014 Urban Congestion Trends: Improved Data for Operations Decision Making*, Report No. FHWA-HOP-15-006, U.S. Department of Transportation, Federal Highway Administration, Washington, DC. Available online: http://www.ops.fhwa.dot.gov/publications/fhwahop15006/fhwahop15006.pdf, last accessed December 20, 2017.

19. Finkelstein, J. (2017). "85th Percentile Explained: A Deeper Look at Calculating the 85th Percentile Speed." (website) Mike on Traffic, St. Louis Park, MN. Available online: http://www.mikeontraffic.com/85th-percentile-speed-explained/, last accessed December 20, 2017.

20. TxDOT. (2015). *Procedures for Establishing Speed Zones Manual*, Texas Department of Transportation, Austin, TX. Available online: http://onlinemanuals.txdot.gov/txdotmanuals/szn/determining_the_85th_percentile_speed.htm, last accessed December 20, 2017.

21. FHWA. (2017). "National Performance Management Research Data Set (NPMRDS) Technical Frequently Asked Questions." (website) U.S. Department of Transportation, Federal Highway Administration, Washington, DC. Available online:

https://ops.fhwa.dot.gov/freight/freight_analysis/perform_meas/vpds/npmrdsfaqs.htm, last accessed December 20, 2017.

22. Carter, D., Srinivisan, R., Gross, F., and Council, F. (2012). *Recommended Protocols for Developing Crash Modification Factors*, NCHRP Project 20-07, Task 314 Report, National Cooperative Highway Research Program, Transportation Research Board, Washington, DC.

23. Banihashemi, M. (2015). "Is Horizontal Curvature a Significant Factor of Safety in Rural Multilane Highways?" *Transportation Research Record*, *2515*, pp. 50–56, Transportation Research Board, Washington, DC. Available online: http://trrjournalonline.trb.org/doi/abs/10.3141/2515-07, last accessed December 20, 2017.

24. Srinivasan, R., Carter, D., and Bauer, K . (2014). *Safety Performance Function Decision Guide: SPF Calibration Versus SPF Development*, Report No. FHWA-SA-14-004, U.S. Department of Transportation, Federal Highway Administration, Washington, DC. Available online: https://safety.fhwa.dot.gov/rsdp/downloads/spf_decision_guide_final.pdf, last accessed December 20, 2017.

25. Fitzpatrick, K. et al. (2000). *Evaluation of Design Consistency Methods for Two-Lane Rural Highways, Executive Summary,* Report No. FHWA-RD-99-173, U.S. Department of Transportation, Federal Highway Administration, Washington, DC.

26. PennDOT. (2017). "Video Log." (website) Pennsylvania Department of Transportation, Harrisburg, PA. Available online: http://www.dot7.state.pa.us/VideoLog/Index.html, last accessed December 20, 2017.

27. AASHTO. (2017). "AASHTOWare Safety Analyst." (website) American Association of State Highway and Transportation Officials, Washington, DC. Available online: http://www.safetyanalyst.org, last accessed December 20, 2017.

28. Washington, S., Schalkwyk, I., Mitra, S., Meyer, M., Dumbaugh, E., and Zoll, M. (2006). *Incorporating Safety Into the Long-Range Planning Process*, NCHRP Report 546, Transportation Research Board, National Research Council, Washington, DC.

29. FHWA. (2017). "Interactive Highway Safety Design Model (IHSDM): Overview." (website) U.S. Department of Transportation, Federal Highway Administration, Washington, DC. Available online: http://www.fhwa.dot.gov/research/tfhrc/projects/safety/comprehensive/ihsdm/index.cfm, last accessed December 20, 2017.

30. Banihashemi, M. and Dimaiuta, M. (2011). "Use of the Output of the Driver Vehicle Module of the IHSDM as Surrogate Measure for Departure Crashes." Presented at the 3rd International Conference on Road Safety and Simulation of the Transportation Research Board, Indianapolis, IN. Available online: http://onlinepubs.trb.org/onlinepubs/conferences/2011/RSS/1/Banihashemi,M.pdf, last accessed December 20, 2017.