U.S. Department
of Transportation
National Highway
Traffic Safety
Administration

NHTSA

DOT HS 812 793

March 2020

Analysis of SHRP2 Speeding Data: Methods Used to Conduct the Research

DISCLAIMER

Suggested APA Format Citation:

Brown, J. L., & Richard, C. M. (2020, March). *Analysis of SHRP2 speeding data: Methods used to conduct the research* (Report No. DOT HS 812 793). National Highway Traffic Safety Administration.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS 812 793		

4. Title and Subtitle	5. Report Date
Analysis of SHRP2 Speeding Data: Methods Used to Conduct the Research	March 2020
	6. Performing Organization Code

7. Authors	8. Performing Organization Report No.
James L. Brown & Christian M. Richard	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Battelle Memorial Institute 505 King Avenue 110 Columbus, OH 43201-2696	11. Contract or Grant No. DTNH22-11-D-00229

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Office of Behavioral Safety Research National Highway Traffic Safety Administration 1200 New Jersey Avenue SE Washington, DC 20590	Final Report: September 2013 – March 2018
	14. Sponsoring Agency Code

15. Supplementary Notes

Dr. Randolph Atkins was the Contracting Officer's Representative on this project.

16. Abstract

This report describes the methodologies and datasets used to prepare a set of data reductions that supported analyses of speeding behavior as discussed in the Findings Report from the same study (Richard, Lee, Brown, & Landgraf, 2019). Importantly, it is intended to be a guide that provides helpful insights for researchers to use when navigating the process of Strategic Highway Research Program 2 (SHRP2) data acquisition and processing. The report focuses primarily on the methods used to (1) obtain and process trip time-series data, (2) prepare data reductions to extract free-flow episodes (FFE)—driving in free-flow conditions where drivers have the opportunity to speed—and (3) extract speeding episodes (SE). The workflow for preparing the data reductions consisted of three components: data acquisition, data management, and data processing. Data acquisition consisted of identifying variables of interest, selecting variables, obtaining a data sharing agreement, preparing and submitting data requests, and retrieving data extracted by the SHRP2 data contractor (VTTI). Data management included tracking the status of each trip during processing; retrieving, manipulating, and storing data in a relational database management system; database administration; data security; and data quality management. Data processing included developing and implementing data processing software tools that cleaned the data; parsed them into Trips, FFEs, and SEs; and produced data reductions suitable for analysis. Lessons learned during the conduct of the research are provided.

17. Key Words	18. Distribution Statement
driver, speeding, methods, SHRP2, naturalistic driving, data processing, data quality, safety, driver behavior	This document is available to the public through the National Technical Information Service, www.ntis.gov

19 Security Classif. (of this report)	20. Security Classif. (of this page)	21 No. of Pages	22. Price
Unclassified	Unclassified	110	

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

Table of Contents

List of Tables

List of Figures

Table of Acronyms

CRC ..... cyclical redundancy check
CSV ..... comma separated value
CTRE ..... Center for Transportation Research and Education
DATP ..... Data Access Technical Plan
DSA ..... Data Sharing Agreement
DUL ..... data use license
FFE ..... free flow episode
GIS ..... geographic information system
GPS ..... global positioning system
HPMS ..... Highway Performance Monitoring System
IRB ..... Internal Review Board
ISO ..... International Organization for Standards
LRS ..... linear referencing system
NDS ..... naturalistic driving study
OBD2 ..... Onboard Diagnostic 2
PII ..... personally identifying information
PSL ..... posted speed limit
RID ..... roadway information database
RDBMS ..... Relational Database Management System
SE ..... speeding episode
SHRP2 ..... Strategic Highway Research Program 2
SOW ..... statement of work
SQL ..... structured query language
VTTI ..... Virginia Tech Transportation Institute
WKB ..... "Well-Known Binary" (computer language format)

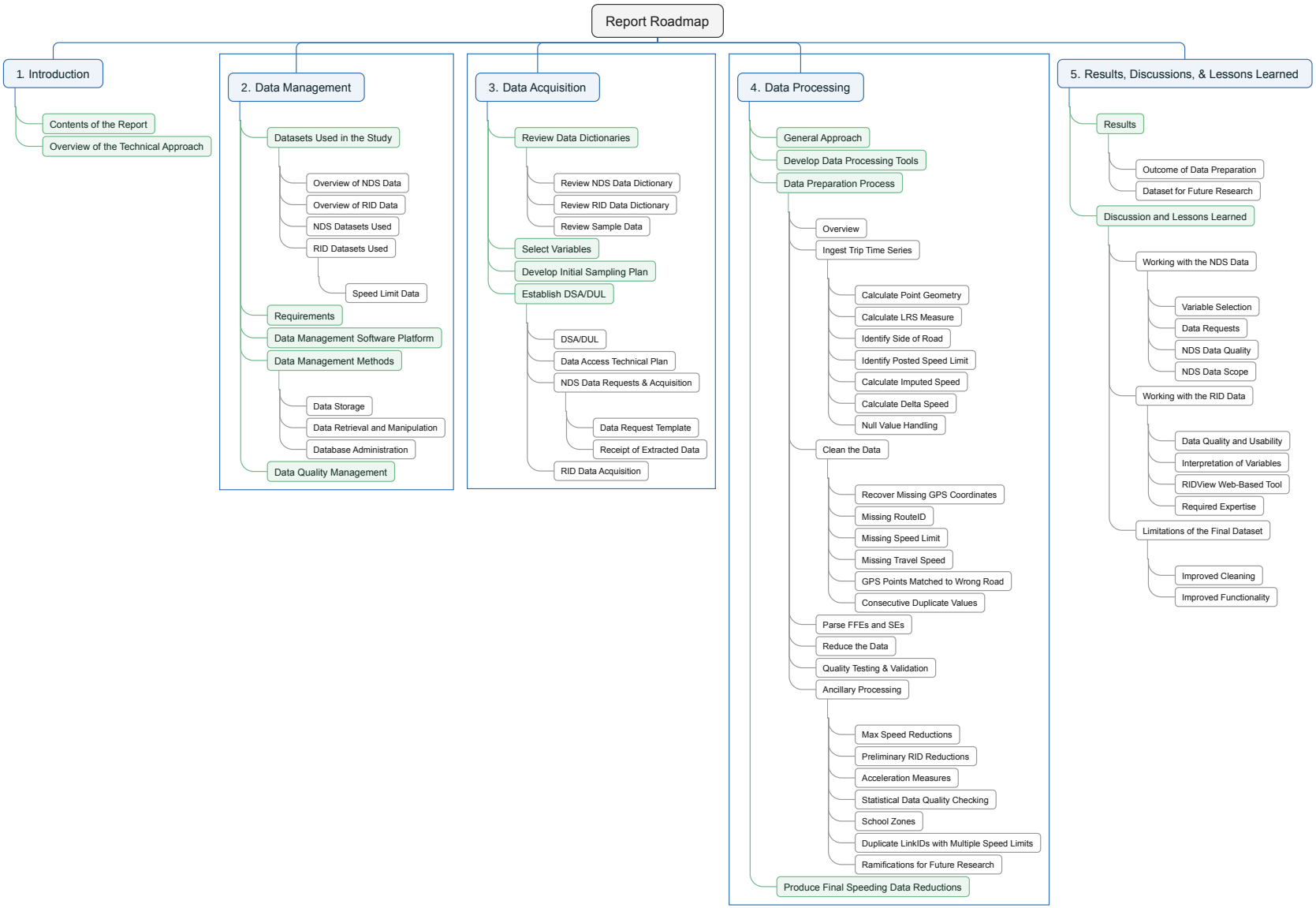

Figure 1. Report roadmap

Executive Summary

This report describes the methodologies and datasets used to prepare a set of data reductions that supported analyses of speeding behavior as discussed in the Findings Report from the same study (Richard, Lee, Brown, & Landgraf, 2019). It is also a guide that provides helpful insights for researchers to use when navigating the process of Strategic Highway Research Program 2 (SHRP2) data acquisition and processing. The report focuses primarily on the methods used to (1) obtain and process trip time-series data, (2) prepare data reductions to extract episodes of driving in free-flow conditions—where drivers have the opportunity to speed—and (3) extract speeding episodes.

The report discusses the following topics:

- Overview of the project technical approach, with focus on the data acquisition, data management, and data processing to extract data reductions suitable for analysis;
- Discussion of datasets used in the study and the process of identifying, prioritizing, and selecting useful variables for examining speeding behavior;
- Software and methods used to store, retrieve, and manipulate the data, and to manage the security and ensure the quality of the data;
- Methods, requirements, and considerations for requesting and acquiring data from the SHRP2 data contractor;
- Development of data processing software tools and processes for preparing the data;
- Processes for ingesting the SHPR2 data into the local database;
- Needs, challenges, and processes associated with cleaning the data;
- Processes for parsing the time series data into Trips, Free-Flow Episodes (FFE), and Speeding Episodes (SE);
- Preparing data reductions with descriptive statistics that characterize the driving behavior within each Trip, FFE, and SE;
- Data quality testing and validation processes; and
- Results of the data processing with a discussion of lessons learned during the conduct of the research.

In addition, helpful tips for researchers are provided throughout the document in the form of side-boxes. These side-boxes encapsulate relevant lessons learned and are intended to provide useful, supporting information that will aid researchers when requesting and consuming the SHPR2 data.

Chapter 1 – Introduction

This report describes the methodologies and datasets used to prepare a set of data reductions that supported analyses of speeding behavior as discussed in the Findings Report from the same study (Richard, Lee, Brown, & Landgraf, 2019). Importantly, it is intended to be a guide that provides helpful insights for researchers to use when navigating the process of Strategic Highway Research Program 2 (SHRP2) data acquisition and processing. Specifically, the report focuses primarily on the methods used to (1) obtain and process trip time-series data, (2) prepare data reductions to extract episodes of driving in free-flow conditions—where drivers have the opportunity to speed—and (3) extract speeding episodes. Methods related to developing sampling criteria and conducting data analyses are discussed in the Findings Report.

Contents of the Report

Chapter 1 introduces the technical approach and the data used to conduct the research. Chapters 2, 3, and 4 describe the three components that comprised the workflow and tools used to prepare the data for analysis. Chapter 5 discusses the results of the data processing and lessons learned about working with the SHPR2 data while conducting the research.

Three aids are provided to assist the reader:

1. Figure 1 provides a roadmap to help navigate the report and to identify the relationships between the various project activities and components.
2. Helpful tips for researchers are provided throughout the document by way of sidebars. These tips provide valuable information about what to expect when requesting, acquiring, and processing the data, and about the data itself. In addition, some sidebars provide definitions that can help the reader understand key concepts.
3. A glossary of terms is provided to help the reader understand some terms used regarding geographic information systems (GIS) and Relational Database Management Systems (RDBMS).

It should be noted that some discussions refer to other project documentation, particularly the Findings Report (Richard et al., 2019) and the Data Deliverable documentation. This was done to avoid burdening these reports with redundant, detailed information.

Overview of the Technical Approach

The SHRP2 Naturalistic Driving Study (NDS) provides a huge repository of speed data in the form of time-series of 1-Hz vehicle speed recordings within individual trips. A relatively small subset of these trips was selected to examine driver speeding behavior. A priority was to leverage the unprecedented number of participants in the SHRP2 dataset to obtain a broad sample of drivers for analysis. Thus, obtaining a large driver sample was prioritized over obtaining a large sample of trips for each driver.

The steps in the overall approach to the project are illustrated in Figure 2, with the focus of this report (i.e., obtain and process the data and extract speeding-related information) highlighted in dark blue. To begin, the sample of trips that were expected to represent a mix of speeding behaviors was identified. The data for this sample were obtained and processed, preparing a set

of trips suitable for examining multiple aspects of speeding. Key measures were then extracted from the trip data, including (1) periods in which drivers had an opportunity to speed, and (2) speeding episodes within those free-flow episodes. These data elements, as well as situational and driver-specific predictors of speeding, were examined using descriptive statistics and regression analyses. In addition, speeding episodes were used to identify different types of speeding and to develop a typology of speeders.

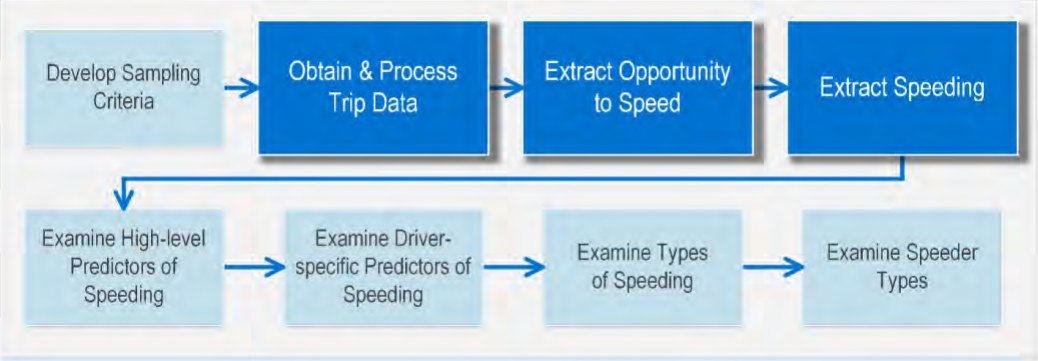

Figure 2. Overview of the steps in the research approach (data preparation methods highlighted)

The data were processed in stages. After the data from each data collection site were processed, the outcomes were reviewed to identify any potential changes to subsequent data requests that would facilitate better analyses, fill information gaps, or any other needs. This strategy allowed helped us to optimize our data requests and analysis needs with the available project resources.

Two-Phase Approach to the Project

The research was conducted in two phases, a test phase and an analysis phase. The objectives of the test phase were to:

- Gain sufficient understanding of the data to effectively and efficiently conduct the research;
- Identify processes, variables, requirements, and tools needed to conduct the research;
- Develop protocols and software tools needed to acquire, process, and analyze the data; and
- Test, debug, and refine the data processing tools and protocols and the analysis methods.

The test phase was conducted using the data from the Washington data collection site. Because Battelle was one of the six data collection contractors in the SHPR2 S07 data collection effort, we were intimately familiar with the Washington data collection area and the procedures and challenges associated with collecting data there. This understanding was invaluable when testing and validating the tools and processes for preparing and analyzing the data. We also used data from the Pennsylvania site to test and validate the data processing tools on a different site to ensure the tools were vetted at more than one site and driving culture. As data processing issues were found between sites, the tools were debugged and validated, and both sites were processed again to ensure that changes made to the processing code produced valid results at both sites.

The analysis phase was executed upon completion of the test phase after all processes, tools, and methods had been tested and qualified. The objectives of the analysis phase were to:

- Obtain a complete sample of data from each of the six SHPR2 data collection sites,
- Process the data using the protocols and tools developed in the test phase, and
- Analyze the complete dataset within and across all data collection sites.

In the analysis phase, the data for all six sites were processed to produce the final data reductions used in the analysis. The data from Washington and Pennsylvania were reprocessed to ensure that all data were consistently processed across all sites. The final output tables were converted to comma separated value flat files and forwarded to the statistician for analysis.

Overview of the Workflow

The overall workflow for preparing the data included three components: Data Management, Data Acquisition, and Data Processing.

- **Data Management:** Data management encompassed all aspects of data preparation, from storing the raw data, to maintaining the intermediate data generated in each step of data processing, to storage and maintenance of the final data reductions in a Relational Database Management System (RDBMS). Data management also included ensuring data quality, security, and privacy. As depicted in Figure 3, quality checking occurred both within and between each step in the data preparation process.
- **Data Acquisition:** Data acquisition entailed obtaining the NDS data for the drivers and trips identified in the sampling plan, along with roadway data from the Roadway Information Database (RID) developed as a companion dataset for the NDS.
- **Data Processing:** Data processing used a custom-developed software tool to import and clean the data, parse the cleaned time series into individual FFEs and SEs, and prepare data reductions that efficiently characterize these episodes.

Figure 3 is a conceptual illustration showing the three components that comprised the data processing workflow and how the data moved through each step in the process.

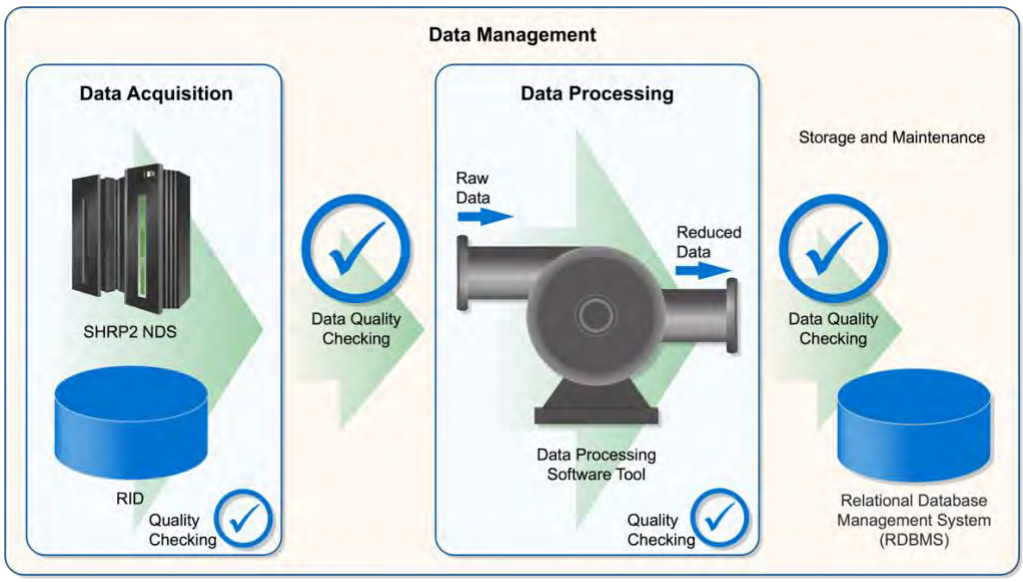

Figure 3. Three-component workflow for preparing the data for analysis

Central to the technical approach was the development of a software tool that accepted time series data files provided by the SHRP2 data contractor, the Virginia Tech Transportation Institute (VTTI), along with roadway data from the RID, and processed the data using automated methods. Automated processes were required because the large amount of data to be treated prevented us from using manual processes for cleaning or preparing the data, although some manual/visual techniques were used to validate representative samples of the data during various stages of the project.

The output of the tool consisted of three data reductions that provided descriptive statistics about the variables included in the time series, calculated at three levels: Trips, Free-Flow Episodes (FFEs) and Speeding Episodes (SEs). FFEs were defined as periods of travel at speeds greater than 5 mph below the posted speed limit, and SEs were defined as periods of travel at speeds higher than the speeding threshold of 10 mph above the posted speed limit. Slight dips below the free-flow and speeding thresholds were allowed to accommodate travel speeds near the threshold. See the Findings report (Richard , Lee, Brown, & Landgraf, 2019) for more details about the definitions of FFEs and SEs. Statistics for Trips, FFEs, and SEs were calculated identically to facilitate comparisons in the analyses.

The remainder of this report details the activities, processes, and tools used to conduct the research, describing each of these components within the workflow. In addition, Chapter 5 discusses the results and lessons learned during the conduct of the research.

Chapter 2 – Data Management

The SHRP2 Analysis of Speeding was a data-driven project. The amount of data to be imported, cleaned, processed, and incorporated into the various analyses planned for the project was quite large; rigorous data management practices and infrastructure were required to effectively and efficiently monitor and maintain the data during the daily activities of development, testing, and processing operations and to minimize error. This chapter discusses the:

- Datasets used in the study;
- Requirements for managing the data;
- GIS-based software platform used for managing the data;
- Methods used to track, maintain, and control access to the data; and
- Data quality management strategies and processes used to ensure the integrity of the data supporting the analyses.

Datasets Used in the Study

Two primary data sources were included in the study: the SHRP2 NDS data and the RID, which is a companion database of roadway information prepared by the Center for Transportation Research and Education (CTRE) at Iowa State University. These datasets are described below.

Overview of NDS Data

The NDS dataset provides hundreds of variables that broadly captured the activities, events, and driver characteristics associated with each trip recorded in the database. The SHPR2 NDS data used in this project consisted of a variety of datasets that included driver variables, trip variables, and vehicle information. Trip summaries identify aggregate measures that describe overall trips. Time series data provide a "breadcrumb" trail indicating the GPS locations of vehicle travel, along with many measures reflecting the vehicle kinematics during the trip. Some of these measures include GPS location, speed (two measures), acceleration, brake and accelerator pedal position, radar-based distance to vehicles ahead, and many more. Vehicle data provide information about the type of vehicle (passenger car, van, truck, etc.) and vehicle condition. Driver assessment variables include data from several surveys and tests administered to participants while they were waiting for the instrumentation to be installed in their vehicles. These tests include personality inventories (e.g., sensation seeking, risk acceptance, etc.), driver knowledge, physical state and mental health (e.g., medical history, clock drawing), and several others that provide important insight into driver characteristics that can influence behavior. The myriad variables examined in this this study provided a wealth of information for exploring the topic of speeding.

Overview of RID Data

The RID provides a rich, varied, and detailed data resource, with variables describing the roadway geometry and many characteristics (attributes) of the roadway network. For roadway information in the current project, we relied almost exclusively on the mobile van data in the RID Version 1.1 dataset and a custom speed limit feature class that CTRE derived from the RID

Signs feature class and other data sources. Mobile van data comprised data collected in the separate SHRP2 S04B data collection effort to capture detailed road data by driving an instrumented vehicle on highly-traveled roads from all data collection sites.

NDS Datasets Used

Table 1 lists the NDS datasets used in this project. The time series data constituted the primary dataset for developing the data reductions; however, other datasets were also used to support sample selection or were used directly in the analyses. Specifically, the Trip Summary dataset was used to identify drivers and trips of interest, which were used to create data requests for the time series data. Other types of data, such as driver demographic and driver assessment datasets, were used to conduct the analysis. See the Findings Report for details about those datasets.

Note that the rightmost column lists the method VTTI used to extract the data, which has a direct bearing on data acquisition costs. Extraction type is discussed following the table.

Table 1. NDS datasets used in the SHRP2 Analysis of Speeding project

Dataset	Description	Examples of Variables[1]	Extraction Type
Trip Summaries	Driving variables aggregated across entire trips	Trip duration, average speed, distance traveled in trip, elapsed time with headway between 1.0 and 1.5 seconds, etc.	Saved Insight Queries
Time Series Data	Variables associated with the vehicle's physical location, sampled at regular intervals. The maximum sample rate is unique for each variable; however, in the current study all variables were sub-sampled at one sample per second (1 Hz).	Instantaneous speed, acceleration, brake pedal state, etc.	Time Series Extraction
Vehicle Data[2]	Variables that provide descriptive information about the type and condition of the vehicle and about any integrated technologies on the vehicle	Vehicle classification (e.g., passenger car, van, truck, etc.), make, model, year, etc.	Saved Insight Queries
Driver Demographic Data	Data from a Questionnaire designed to investigate a variety of demographic, psychological, and health information about the participant	Driver age group, race, ethnicity, gender, Sensation Seeking Score, speed 10-20 mph over, speed 20+ mph over, etc.	Saved Insight Queries

[1] The examples listed are representative of the variables available in the dataset and do not necessarily imply that they were used in the current analyses.

[2] Vehicle classification was used during some exploratory analyses, but no vehicle variables were included in the final analyses.

Two methods were used to extract the data, depending on the type of data being requested. The method used to extract the data at VTTI has important cost implications for acquiring data.

- **Saved Insight Queries:** These datasets provided the same data that can be viewed using queries on the Insight Data Access Website. The results from Insight queries cannot be saved directly from within Insight; however, VTTI can extract data using the same search terms that can be generated on Insight. One example of such a data request might be to extract all variables from the Trip Summaries data where Trip duration > 10 minutes and Time Moving > 5 minutes. Saved Insight queries are relatively inexpensive.
- **Time Series Extraction:** Point-to-point vehicle location and associated variables. In the current project, all time-series data were sampled at one sample-per-second (1 Hz). These data, extracted from the NDS time series, are relatively expensive compared to saved Insight queries.

RID Datasets Used

Table 2 lists the RID feature classes used in the project with examples of the variables included in the feature class. It should be noted that due to limits in project resources, we did not examine the effects of most RID variables on speeding in the current analysis; however, several key variables in these feature classes were necessary for connecting the GPS location to the correct roadway and—ultimately—identifying the speed limit at each GPS point.

Table 2. RID (Version 1.1) feature classes used in the SHRP2 Analysis of Speeding project

Feature Class	Description	Examples of Variables
Speed Limit	Provides the posted speed limit data for identifying driver speeding at each GPS location	Posted Speed Limit
Links	Used to identify which road segment the vehicle is traveling on (via Link ID supplied in the time series). Also provides functional class information. Road segments in the Links feature class typically run from intersection to intersection.	Link ID, Functional Class
Routes	Primary feature class, which uniquely identifies road attributes and geometries at the roadway level. Road geometries contained in all other RID feature classes are a subset of the geometries in the Routes feature class. The Route ID variable is used to associate the road segments in each RID feature class with the roads in the Routes feature class.	Route ID

Speed Limit Data

Speed limit was a critical variable required for determining speeding behavior. At the time of the kickoff meeting and initial activities in this project, the RID did not contain speed limits in a form that could be readily used for identifying the speed limit of the road at each GPS location. Rather, the speed limit information was coded in the *Signs* point feature class of the RID. To obtain speed limit data in a form that could be easily incorporated into our dataset, we requested that CTRE develop a new line feature class that conflated the *Signs* point data onto the roadway network, providing a single line feature for each speed limit and direction of travel for each road. CTRE developed a set of feature classes, one for each data collection site, that provided the required speed limit data in this form. These feature classes were later incorporated into Version

2.0 of the RID as the *SpeedLimit_ReducedDataset* feature class within each data collection site's geodatabase.

Data sources for the *SpeedLimit_ReducedDataset* feature class included the Highway Performance Monitoring System (HPMS), State Department of Transportation data, the Environmental Systems Research Institute[3] Links feature class, and the S04B mobile van data collection effort. Our previous experience (Richard et al., 2013) with speed limit data from State and local jurisdictions found that data quality from these sources might be insufficient. Similarly, in past projects we found that the ESRI speed limit data had a relatively high error rate. Furthermore, an investigation of the ESRI data incorporated in the RID found that only 2.6 to 7.6 percent of links (depending on data collection site) included posted speed limit. To avoid a large data validation effort and keep to project schedule, we chose to focus the analysis on road segments that included mobile van data. These road segments were expected to have high accuracy, ground truth posted speeds, which we confirmed in a small-scale validation activity.

Relationship Between Datasets

As shown in Figure 4 below, for each record in the NDS time series data, key RID variables at the GPS-based location were extracted and merged with the variables in the time series record. This method provided driving data, speed limit, and roadway attributes needed for extracting subsequent RID variables at the GPS location in each record.

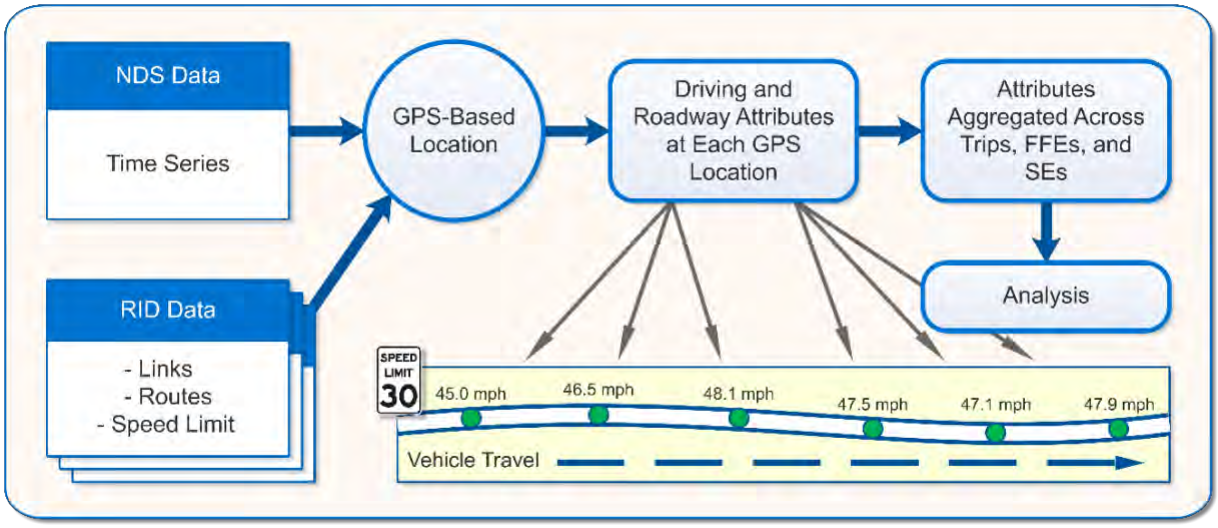

Figure 4. Linked data sources for extracting driver and roadway information at each GPS location

Requirements

A set of requirements was established to specify the characteristics of the data management software platform and associated processes. These requirements were developed to ensure that these tools and processes would be able to effectively manage these large and complex datasets

[3] ESRI is the maker of the ArcGIS geographic information systems software.

and efficiently implement the processes needed to accomplish the goals of the project. These requirements were as follows:

- All data storage and processing tools must be GIS enabled. They must be capable of storing data with GIS geometry information and performing or supporting all operations needed to create and manipulate geometries, extract relevant data associated with the vehicle location using a Linear Referencing System (LRS) and Dynamic Segmentation, and performing a wide array of GIS operations.
- The data management software must be capable of efficiently storing, retrieving, and manipulating millions of records in a data table or feature class. The data storage system must be scalable to allow expansion of the dataset as needed to accommodate the data without sacrificing performance.
- Data storage and retrieval must minimize the physical storage requirement (e.g., file size) and time to retrieve the data. Retrieval methods must minimize search and retrieval times.
- Data must be searchable using standard methods. Searches must be flexible enough to extract records with many types of data within and across trips, participants, and data collection sites.
- The underlying data structures, mechanisms, and operations of the data management platform must conform to well-established standards to minimize time validating processes and to ensure rigorous and defensible outcomes.
- The data management system must support both manual and automated processes for storing, retrieving, maintaining, and processing the data.
- The data management system must support regular maintenance operations such as backing up the data, optimizing data storage, recovering from database corruption, etc.
- The data storage system must be secure. At minimum, it must be capable of limiting access to the data to those who are listed on the Data Sharing Agreement/Data Use License. The data must be secured against both cyber-attacks and physical theft.
- The data management platform must support tools that can be used to track the status of each trip at each stage in the data processing, from pre-request through final disposition.
- Status must be searchable, and results must be actionable (e.g., process only those trips that are flagged as incomplete).

Data Management Software Platform

A robust, GIS-enabled infrastructure was required for storing, managing, and processing the large amount of data we would request in the project. We determined that the two most viable solutions for efficiently and effectively managing the data were: (1) ESRI ArcGIS Desktop to store, clean, and visualize the data; perform data quality evaluations; and develop the data reductions; or (2) a hybrid solution using an relational database management system (RDBMS) for most operations and third-party GIS tools to support data visualization, quality checking, and spatial analysis. We chose the second option because RDBMSs are designed to efficiently manage large datasets, and, from previous experience, we expected to achieve greater operational performance with an RDBMS. Also, an RDBMS provides a greater level of flexibility by supporting multiple GIS clients, including ArcGIS.

Our requirements for the RDBMS included the following.

- Mature, respected RDBMS technology
- Robust and comprehensive GIS capability
- Ability to integrate with ArcGIS and other third-party GIS clients
- High scalability
- High security
- Ease of maintenance and administration
- Clear, comprehensive documentation

We chose PostgreSQL as the RDBMS and a combination of PostGIS, ArcGIS, QGIS, Google Earth™, and Python supporting the GIS processing and visualization requirements. PostgreSQL is an enterprise-class, open-source RDBMS that is secure and extensible. GIS operations are supported via the PostGIS extension to PostgreSQL. PostGIS provides hundreds of International Organization of Standards (ISO)-compliant[4] GIS functions that can be called from Structured Query Language (SQL) queries, stored procedures, and triggers. Stored procedures can be programmed using several languages, including PL/pgSQL (similar to Oracle's PL/SQL language), Python, R, Java, Perl, Tcl, Ruby, C/C++, and others.

PostgreSQL is designed to be highly reliable and minimize the risk of data loss. PostgreSQL uses a variety of techniques, such as forced writes to disk from data caches, writing database changes to write-ahead logs before committing to the database, Cyclical Redundancy Checks, configurable checkpoints, and other techniques to avoid data loss and database corruption in the event of a power loss or hardware failure during operations.[5]

A detailed summary of the capabilities and features of PostgreSQL can be found at www.postgresql.org/about/.

Data Management Methods

This section discusses the methods used to manage the data, including data status tracking, data maintenance, and data security.

Data Status Tracking

A critical component of the data management strategy was keeping track of the status of each trip as it underwent each stage of data processing. Because the trips were processed one at a time, and faults could occur during any step of data processing, and because the data processing

[4] PostGIS conforms to or implements part or all of the following standards: ISO/IEC 13249-3:2016 (*Information technology – Database languages – SQL multimedia and application packages – Part 3: Spatial*), ISO 19162:2015 (*Geographic information – Well-known text representation of coordinate reference systems*), and ISO/OGC 19125 (*Simple Features Standards*).

[5] See Chapter 30 of the PostgreSQL documentation at www.postgresql.org/docs/current/static/wal.html for details about PostgreSQL reliability.

procedure was nearly entirely automated, it was critical to know the status of the data after each processing step was completed in order to troubleshoot and recover trips that could not be successfully processed. Figure 5 provides a model of the trip status tracking process, starting with identifying trips in the sampling plan and ending with delivering data to statisticians for analysis. This process is described following the table.

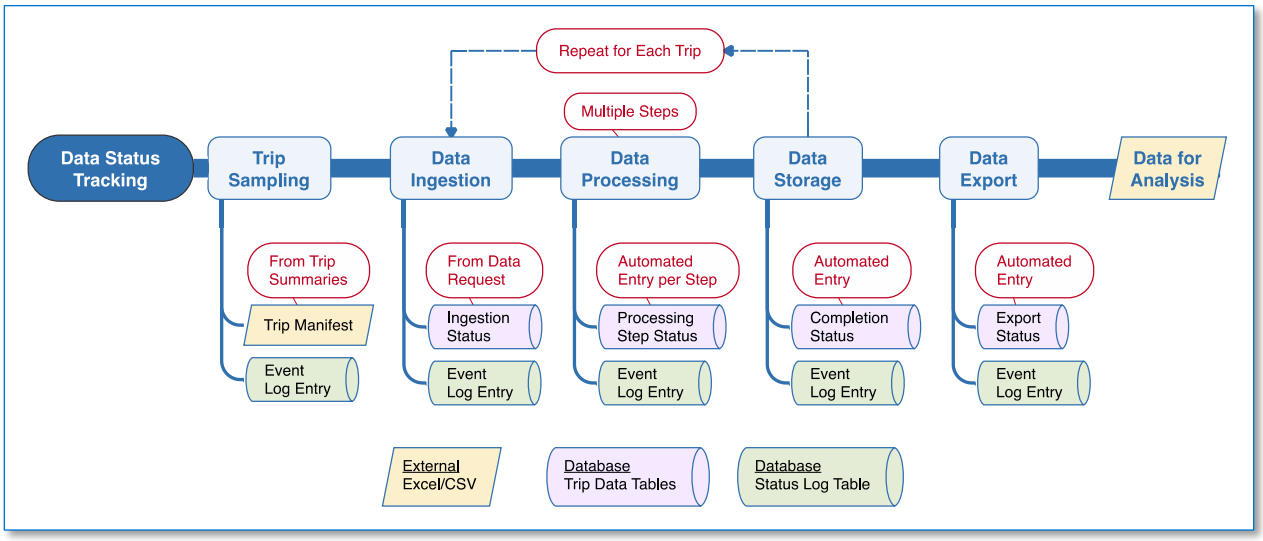

Figure 5. Data management – trip status tracking

Data status tracking began with the receipt of Trip Summary data requested from VTTI. The data included trips that were selected based on criteria developed in the sampling plan. Included in these data was an Excel spreadsheet listing the Trip IDs associated with the request. The spreadsheet data were imported into a master tracking table in the database, which formed the basis for all subsequent tracking operations. The time series data associated with each of these Trip IDs were then requested from VTTI.

Upon receiving the time-series, the data were saved to a file structure on a local server. The data processing Python script was executed to ingest the data from the local files into the database and process the trips, one trip at a time. Automated processes in the script assigned a unique status code in the status table for each step in the process, indicating which step had been processed and whether the data processing for the step was successful. Additional steps were not executed if a failing status was logged in the status table. In this way, we were able to identify the step at which a given trip failed processing to aid in troubleshooting and recovery of the data in that trip. Furthermore, the data processing script used the status code to determine which trips to process, allowing us to interrupt and restart processing if needed.

In addition to status tracking, a separate event log was created to record each data processing entry and any messages related to the event. This log was used during software development to debug the stored procedures used to process the data and to identify patterns of success or failure.

Data Maintenance

Data maintenance included database operations and strategies used to store, retrieve, manipulate, and backup the data. This section discusses these aspects of data maintenance.

Data Storage

The data were stored in a separate database schema assigned to each data collection site. This strategy was adopted for two reasons. First, it provided a way to organize the many tables generated during ingestion and data processing. Segregating the data tables by site ensured that cross contamination of the data between sites did not occur. Second, storing the data in separate database schemas facilitated the use of common functions to process the data. Each function included a site code as an input parameter, allowing the function to process the data within the scope of that schema. This methodology also ensured segregation of the data by site.

A general schema was used to hold common SHRP2 data fields. This schema included lookup tables for site code, day-of-week, time-bin, month, and so forth. In addition, the common schema held some functions for performing general routine GIS procedures, such as converting vehicle latitude and longitude to GIS-operable point geometries, snapping points to links, and preparing RID roadway data for use with PostgreSQL. This schema was generic enough that it could be reused for other projects involving SHRP2 time series data.

Another, separate schema was used to contain project-specific data and functions. These tables included a status message lookup table and intermediate working tables for data ingestion. All data processing functions were also held in this schema. Using common data processing functions ensured that the data were treated uniformly across all data collection sites.

A separate database schema was created to hold the data generated in the test phase of the project to prevent any test-phase data from erroneously being used in the analysis phase.

Data Retrieval and Manipulation

Because data processing was accomplished using database stored procedures (known as functions in PostgreSQL parlance) within PostgreSQL, the data processing operations were an integral part of data maintenance. Data storage, retrieval, and manipulation coincided with the three steps in the data preparation processes—ingestion, cleaning, and data reductions.

Data were retrieved in two ways: (1) within functions that extracted relevant variables during data processing and (2) exported for use by statisticians. In both cases, the data were retrieved non-destructively by using the data to create new tables or external data files rather than update or change existing tables. For example, data ingestion processes retrieved raw data to derive variables and prepare an initial source data table. Data processing functions retrieved the data and created new intermediate time series tables with cleaned trip, free-flow, and speeding time series—without altering the source table. The advantage to using a non-destructive retrieval strategy is that each step is traceable, with the original data from each step left intact; also, differences between steps are searchable. The disadvantage is that many tables that need to be organized and maintained are created, and the size of the database increases with each table. Fortunately, PostgreSQL was designed to scale to the needs of the data and could successfully accommodate the scope of the datasets we generated.

Data for use by statisticians were extracted using SQL queries that copied the table to CSV files.

Database Administration

The database administrator performed regular, routine database administration operations to minimize the risk of data loss and to ensure peak performance from the database operations. Regular, redundant backups were committed daily to prevent data loss in the event of hardware failure or database corruption. All backups were stored on secure, encrypted hard drives. In addition, the database tables were periodically maintained using PostgreSQL's "vacuum" facility to compact and re-structure the data. Vacuuming the tables as needed—both manually and using PostgreSQL's auto-vacuum feature—optimized the efficiency and performance of database operations and reduced the occurrence of database errors.

Data Security

An important aspect of data management is ensuring that access to the data is restricted to only those who have permission to view and use the data. The Data Access Agreement/Data Use License specifies—by name—the researchers who are authorized to access the data. Two strategies were employed to limit access to the data. First, the database was housed on a secure server, and PostgreSQL was configured to allow local access only. Specifically, connections to the local server were accepted by PostgreSQL, and all connections from other computers/servers were denied. This required the operator to physically work at the secured, local server.

The second layer of control was applied using role-based security measures natively built into the PostgreSQL RDBMS. The following capabilities, as configured in this project, were employed to secure the database against unauthorized access:

- Explicit permissions were granted only to personnel who needed access to database assets and only as the need arose. All others were denied access.
- A unique username and password[6] were required.
- By default, users cannot write to databases they did not create. This configuration was maintained.
- All database files are read/write protected by any account other than the PostgreSQL super-user (administrator) account. No users were granted super-user status, and the database administrator utilized the super-user account to administer the database files only as necessary.

Data Quality Management

Data quality was of paramount concern throughout all steps in data processing and analysis. Quality checks were performed at each stage of project development, from receiving and testing pilot data to final analysis. Our strategy for maximizing the quality of the data reductions was to minimize the possibility of falsely identifying speeding episodes where no speeding actually occurred (i.e., Type I error). For example, school zones have multiple speed limits depending on

[6] Although password protection is optional in PostgreSQL, the database was configured to accept only password-protected connections.

time of day, presence of children, or other criteria established by individual jurisdictions. Because there was uncertainty about whether the school zone speed was in force, driving on roads less than 30 mph was discarded to eliminate potential error of false positive speeding in school zones. Type II errors, in which no speeding is identified where speeding actually occurred, were not a problem because they simply ignored potential speeding episodes. Although it is likely that some valid speeding episodes associated with both error types were discarded, the large sample size afforded by the SHRP2 dataset ensured there were enough speeding episodes to satisfy the required statistical power in the analyses.

Data quality was tested throughout the project, beginning with checking the datasets received from VTTI to ensure that the trips we requested were received and that the datasets complied with the conditions specified in the data request. Processes were incorporated into the data import function to ensure that the variables ingested into the database were of the correct type and form.

Data processing quality was also verified during development of each function to ensure that each function produced accurate results. These functions identified records with missing data in key variables (travel speed, GPS location, road identifiers, etc.), out of range variables (e.g., high variance between GPS speed and Network speed—see page 31), incorrectly matched GPS points (see page 36), and so forth. Furthermore, site-specific data quality checks were performed to find any errors that may have been related to unique conditions at each data collection site (e.g. GPS availability/quality in urban versus rural areas). As the data processing functions were refined to address new data quality challenges found at one site, data from the previously tested sites were tested with the updated or new functions to ensure that the code produced high-quality outputs at all data collection sites.

Once the data from all sites were processed, statistical techniques were used to confirm data quality. Two types of errors were discovered for some SEs: (1) duplicate records with different posted speed limits (PSL) and (2) areas with high concentrations of speeding, which suggested the possibility that the PSL at those locations was wrong, producing false positives for speeding. To address the duplicate records issue, we excluded any SEs with these duplicate records to eliminate the possibility of introducing noise or error into the analyses. See the *Statistical Data Quality Checking* section for a discussion about how we addressed the potential for false speeding at those locations.

Addressing Data Quality Issues

Although several data quality challenges were encountered (as might be expected from large, naturalistic driving datasets), methods and processes for identifying and addressing each challenge were identified and implemented.

Chapter 3 – Data Acquisition

Data acquisition consisted of developing processes and tools for requesting and receiving the data used in the project, from planning and preparation to submitting requests and receiving data. Data acquisition included the following activities.

- Review SHRP2 data dictionaries and review available sample data
- Select, catalogue, and prioritize SHRP2 variables for target research questions
- Develop an initial detailed sampling plan and data analysis plan
- Establish data sharing agreement for SHRP2 data use
- Identify site-based trip sample
- Submit data requests
- Retrieve extracted data

The first four activities were preparatory elements that determined which variables to acquire and established the required Data Sharing Agreement with VTTI. The last three components implemented the operations by which the data were requested and acquired. First a sampling plan that identified the drivers and trips to be included in the dataset for each site was developed using trip summary data. Based on the sampling plan, a data request for time series data was submitted, and the subsequent extracted data were retrieved from VTTI. Figure 6 shows these steps performed to acquire the data, followed by details about each step.

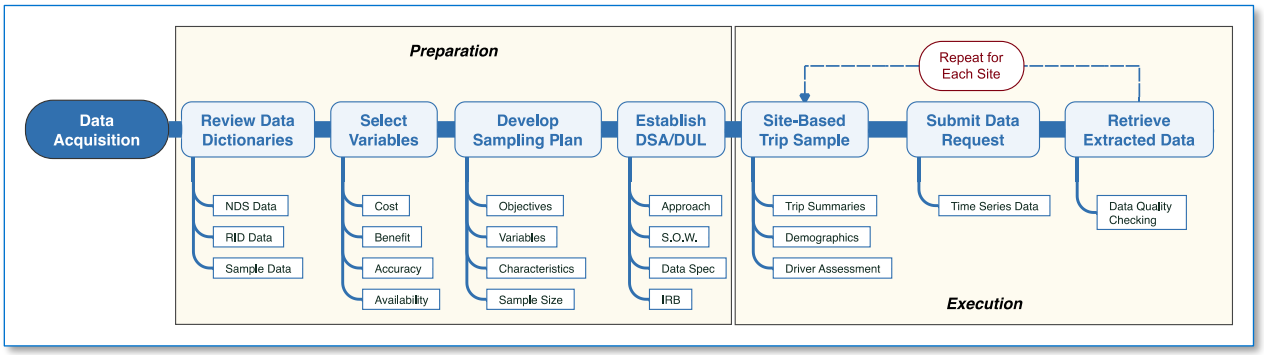

Figure 6. Steps in the data acquisition process

Review SHRP2 Data Dictionaries and Review Available Sample Data

The first step in identifying the datasets and variables to include in the analysis of speeding was to review the various data dictionaries and sample data to gain a detailed understanding of the variables available and how they were represented in the time series.

Review NDS data dictionaries

We reviewed the six primary data dictionaries that are available from the SHRP2 Insight Data Portal (TRB, 2018). These included time-series, trip summary, crash event, crash detail, driver assessment, and vehicle data dictionaries from the SHRP2 NDS data set. In addition to these, we reviewed GIS data from the RID. The source, name, and type, and number of variables within of each dictionary are listed in Table 3 below.

Accessing the Insight Portal

Before researchers can access data on the Insight portal, they must create a user account, which must be approved by the SHPR2 data contractor. This requirement is to ensure controls on who has access to the data in accordance with the participant informed consent agreement.

Table 3. Data Sources for Analysis

Source	Name	Type	Variables in Dictionary
SHRP2 NDS	Time-Series	Vehicle	76
SHRP2 NDS	Trip Summary	Vehicle	34
SHRP2 NDS	Crash Event (Video Reduction)	Crash	54
SHRP2 NDS	Crash Detail	Crash	331
SHRP2 NDS	Driver Assessment	Driver	506
SHRP2 RID	Roadway Information Database	GIS	36
Total			**1037**

Each dictionary contained information regarding individual variables' name, description, sample response, range, and other associated parameters. The 1,037 individual variables in the combined data dictionaries were used to populate a table for use in further analysis.

Review RID data dictionaries

The RID data dictionaries were also reviewed to identify the roadway variables available, how they are coded, what ranges of values are available, and which variables are most likely to influence speeding in some way. The data dictionary included one table for each GIS feature class and listed each variable in the feature class, along with units of measure, range of values, and so forth. Relevant variables from each feature class were listed and prioritized in terms of usefulness for identifying roadway factors associated with speeding behavior; however, all variables in the RID were available to us, as the RID license and data are purchased from CTRE and is separate from the NDS dataset.

Review Sample Data

In addition to reviewing the data dictionaries, we examined sample time series data, supplied by VTTI, to understand the form of the data-stream in the time series, how the variables interact, and what potential pitfalls we might encounter using the time series data. The sample data provided a clear understanding of the relationships between variables and some of the data processing challenges we would need to address.

Select, Catalogue and Prioritize SHRP2 Variables for Target Research Questions

Once we had a clear understanding of which variables were available and how they might be used to analyze speeding behavior, we underwent a rigorous, methodical procedure to prioritize the variables—and ultimately select the variables we would request for the project. This process began with cataloging the available variables to identify whether each variable would likely provide a potential contribution to the analyses. Potentially useful variables were prioritized by rating them with respect to four factors: cost, benefit, accuracy, and availability.[7] The variables were also rated in terms of expected utility, which was defined as the level to which the variable was expected to support the analysis. A final candidate set of variables was identified based on the combined scores of each variable. Appendix B contains a detailed description of the process we used to select, catalogue, and prioritize the variables.

Develop an Initial Detailed Sampling Plan and Data Analysis Plan

A sampling plan was critical to the success of the project so that the data included would be representative of the population of interest, ensuring that the results and conclusions from the analysis were meaningful. The goal of the sampling plan was to:

1. Define the specific objectives,
2. Specify the data elements and variables required to address the objectives,
3. Obtain a data set that is representative of a range of driving conditions, and
4. Obtain a sample that is of sufficient size to provide statistical power to detect significant differences of interest.

Sampling and Analysis Plan

The sampling and analysis plans impact decisions about other activities, such as variables selected, amount of data to request, structure of the reduced data, software development for data processing, etc. Early planning will help to facilitate other project activities.

The goal of the analysis plan was to identify the necessary statistical methods needed to conduct robust statistical analyses of the processed data. The sampling and analysis plans are described in detail in the Findings Report (Richard, Lee, Brown, & Landgraf, 2019).

Establish Data Sharing Agreement for SHRP2 Data Use

Data Sharing Agreement / Data Use License[8]

The Data Sharing Agreement is an agreement specifying the work to be performed by the SHRP2 data contractor (VTTI), the variables to be extracted, the personnel who will have access to the data, and associated costs. It is the primary mechanism for establishing the requirements, parameters, and terms associated with data extraction.

[7] Some variables were not available for some vehicles. Particularly, variables that provided data via the vehicle network were missing for older vehicles without an OBD2 port.

[8] At the time of project kickoff, VTTI used the Data Sharing Agreement as the mechanism for specifying the data to be extracted and conditions of its use. Subsequently, they have replaced the DSA with the Data Use License.

After receiving IRB approval, we worked with VTTI to develop a statement of work for providing the data required for the project. The SOW was prepared by VTTI, and it outlined the general activities and costs for the acquisition of the data. The SOW described the objectives for the data use, VTTI's general approach to extracting and providing the data, a description of the data that would be requested, deliverable data format, and costs associated with providing the data. Several rounds of discussion were required to balance the number of trips to be extracted against the number of variables to extract in order to keep costs within the project budget for acquiring data.

During the project, several amendments to the DSA were executed to include additional staff added to the project. The DSA required (1) IRB approval for the project by Battelle's IRB and (2) the signature and certificate of IRB training for each researcher who had access to the data. A separate amendment was executed to reflect a change in driver assessment variables to be extracted. The original DSA listed only specific driver variables; however, because of the exploratory nature of the project, we needed access to all driver assessment data to determine the most effective and useful driver variables to include in our analyses.

Data Access Technical Plan

In addition to the SOW, we developed a Data Access Technical Plan (DATP), which was our guiding document for making data requests. The DATP provided details about the data to be requested, the purpose of the data request, format of the delivered data, schedule for delivery, requirements for protecting participants' personally identifying information (PII), and a list of variables for extraction.

The DATP described our general approach for the two-phase structure of the project plan (i.e. Test phase and Analysis phase) and our approach to data requests that support each phase. The data extraction and processing workflow in Figure 7 was included in the DATP to guide the data request process.

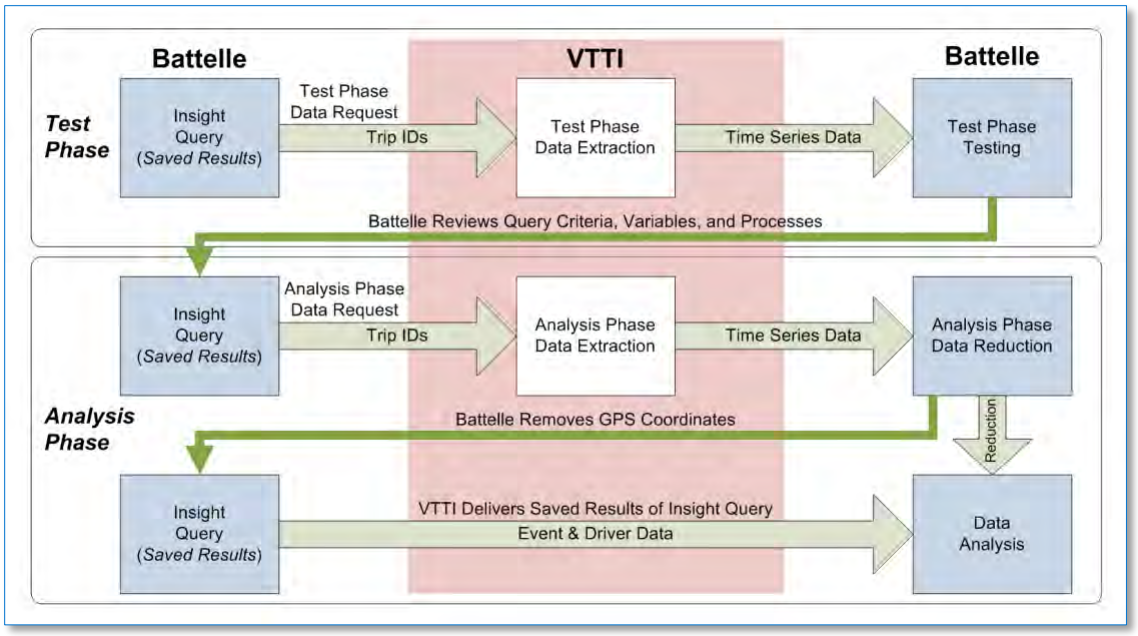

Figure 7. Data extraction and processing workflow

For the Test Phase, Battelle performed an "Insight"[9] (query language) query of the Trip Summaries to identify a sample of trips from the Seattle site. Insight queries were conducted on the SHRP2 Insight website. Once the query parameters were established and refined, we submitted the Trip IDs in a request for the related Time Series data. We used this data to test our software and processes. For the Analysis Phase, we updated our selection criteria and queries in the Insight portal for trips of interest, based on the findings in the Test Phase. We submitted the Trip IDs in subsequent Time Series data request for the larger sample of trips across all sites. In addition to time series requests, we performed Insight queries for relevant event and driver data and used them in the analysis. For each Insight query, Battelle submitted the query criteria to VTTI, who used them to provide the saved results of the Insight query. See the *NDS Data Requests and Acquisition* section below for more details about the Insight queries.

The DATP included a table listing the datasets that would be requested in the project. This table, shown in Figure 8, included the type of Insight data table that would provide the source data, the way the variables would be extracted, the scope of the data filters used to extract relevant data, the size of the data to be extracted, and the objectives for using the data.

[9] See https://insight.shrp2nds.us/

Dataset Number	Data Type	Variables	Query Filter	Data Quantity	Purpose
1	Trip Summaries	*Saved results of Insight query*	All Trips	—	Obtain Test Phase Trip IDs
2	Trip Time Series	Automatic extraction only (Variable list in the Appendix)	Battelle Provides: • Trip IDs • Variable list	40k–50k trips • Seattle site • 50 to 100 trips per driver	Pilot testing
3	Trip Summaries	*Saved results of Insight query*	All Trips	—	Obtain Analysis Phase Trip IDs
4	Trip Time Series	Automatic extraction only (Variable list in the Appendix)	Battelle Provides: • Trip IDs • Variable list	120k–150k trips • All remaining sites • 50 to 100 trips per driver	Analysis
5	Event Database	*Saved results of Insight query*	All available epochs	—	Analysis
6	Driver Datasets	*Saved results of Insight query*	All participants	—	Analysis

Saved results of Insight queries Request for Time Series data

Figure 8. Summary of data requests from the Data Access Technical Plan

The DATP also provided general specifications regarding how we would protect PII, data sampling rate and format (i.e., one sample per second, nearest neighbor sampling rectangularization),[10] data file format, and other key information needed to communicate our data requirements to VTTI.

NDS Data Requests and Acquisition

Two general types of data were required for the project and requested: saved Insight query data and time series data.

Saved Insight query data were results from the queries that any registered user of the Insight Portal can perform. Because Insight does not provide a way to save the results of queries, a formal request that provided query parameters that are available on Insight was prepared and submitted to VTTI.

CSV File Format

Some data tables contain variables with commas in the data field (e.g., crash event descriptions). Importing these files into applications as comma delimited files can be challenging because the commas in the data are erroneously interpreted as delimiters for new fields. Requesting data in tab delimited format for these tables can facilitate reading these files.

[10] See glossary

The output from these requests were largely in the form that can be viewed on Insight but were delivered to us as saved files that we could download and incorporate into our database. Because these were simple queries that filtered on variable values, the cost of acquiring these data were relatively minor. These data were used to extract trip summaries, event and baseline epochs, and driver demographics/driver assessment data.

Time series data included sequential GPS breadcrumbs that indicated the location of the vehicle at one-second intervals. This was the primary base dataset from which the data reductions used in the analysis were prepared. In addition to vehicle location, the time series requests included variables such as travel speed, acceleration, lane position, vehicle controls state, environmental variables, and distance to lead vehicle. See the Appendix C for a complete list of the variables we requested from the NDS data repository.

Data Request Template

To make sure the data request specified our requirements with enough detail, accuracy, and clarity, and to facilitate the multiple data requests we expected to submit, we developed a common template for requesting the data based on the DATP. The template contained detailed specifications related to each aspect of the DATP. With each data request, we completed and customized the template and submitted it to VTTI. The level of detail in the data specification incorporated in the template substantially removed any ambiguity about all aspects of the data requests.

This template included:

- **Description of the data requested:** Detailed description of the dataset being requested, including things like sample size (e.g., data collection site, number of trips, etc.). For the time-series requests, this section also referred to an accompanying Excel spreadsheet that listed the File IDs (trip identifiers) to be extracted.
- **Data quality filtering:** Specifications about pre-filtering of the data, if needed. For example, in the time series requests, we asked that VTTI not deliver any trips for which the GPS coordinates were not available.
- **Variables requested:** List of variables to be extracted
- **Data format specification:** Detailed specifications of the format in which we requested the data to be delivered. Elements that were specified here included things such as sample rate, file format (e.g., CSV), representation of NULL values, etc.
- **Data delivery specification:** Detailed specifications about the way we preferred the data be delivered (folder structure, preferred method of data transfer, etc.)

Receipt of Extracted Data

Once we submitted the data request, VTTI data analysts prepared the data extraction and delivered the data to us via file transfer over the Web. Initial datasets were delivered via VTTI's Scholar Portal, which was their data hosting platform. Partway through the project, the VTTI retired the Scholar Portal, and we received the remainder of the data using Box, Battelle's preferred secure cloud data hosting solution. Box provides configurable user permissions and secure file sharing.

Not all trips requested could be delivered to us. Some trips had data quality issues with source data at VTTI (GPS coordinates not available, problems with map matching, etc.). Also, to protect drivers' privacy, the sources and destinations of trips were redacted from the time-series; some trips were short enough that there was no driving outside of these 'PII Zones.' Finally, some trips were flagged as "no longer available.". Although these trips could not be delivered, the proportion of these trips was small. Table 4 shows the number of trips requested and received from each data collection site.

Expect and Plan for Data Loss

A small amount of data loss can be expected when requesting data. Approximately 4.5% of the trips we requested could not be delivered because of availability or data quality issues with source data at VTTI.

Table 4. Counts of trips requested and received

Site	Trips Requested	Trips Received	Percent Trips Undeliverable
Florida	67,163	63,804	5.00%
Indiana	21,300	20,205	5.14%
North Carolina	49,313	47,029	4.63%
New York	69,543	65,818	5.36%
Pennsylvania	16,344	15,856	2.99%
Washington	49,423	48,034	2.81%
Total	**273,086**	**260,746**	**4.52%**

RID Data Acquisition

The RID dataset consists of several roadway feature classes that describe numerous aspects of the roadway network it was driven in the study. Because there is no participant information encoded in the RID, acquisition of the dataset was as straightforward as requesting and purchasing the RID dataset. As described earlier, RID did not contain speed limit data in a form that could be readily used. Consequently, we requested a new speed limit line feature class to identify speed limit. Upon receiving the speed limit feature class from CTRE, we validated many locations in the Seattle area for which we knew ground truth speed limit to verify the accuracy of the data from CTRE. Within the validation sample, the speed limit data from CTRE appeared to be of high quality and reliable.

The RID data contains a separate file geodatabase for each data collection site. Our original intention was to combine feature classes from all six sites into single tables in our database. For example, the speed limit feature class from each data collection site would be combined into a single table with a new variable that identified the site. After examining the various feature classes across sites, however, we discovered that some data fields had different names or slight variations on the spelling of the data field, and some date of tables had different structures between locations. Also, we felt that keeping the feature classes separate in the database would improve processing performance. Consequently, we chose to leave the data tables in separate database schemas rather than combine them into a single table.

Another challenge for data processing was related to variability in data quality depending on the data source. The variables that were captured during the S04 mobile van data collection effort were expected to have very high accuracy, whereas the data from other sources, such as Environmental Systems Research Institute (ESRI), were known to have lesser quality, both in terms of accuracy and completeness. Furthermore, our experience with data from State and local jurisdictions suggested that the data quality from these sources could potentially have errors. In one study (Richard et al., 2013), we found that data from some jurisdictions had as much as 30 percent to 50 percent of the data with incorrect speed limits. Consequently, we chose to focus only on those variables that could be obtained from the mobile van data. We expected this would also guarantee that the variables would be consistent across sites, and with a few minor exceptions, this was the case.

Once we identified which feature classes we would use in the study, we imported the associated data tables into separate database schemas, one for each data collection site. This approach helped us to keep the data sets organized and prevent us from cross contaminating the data from across the sites. We also tested the measures coded in the feature classes and found that they differed slightly from those generated by PostGIS. To compensate for these errors, we re-calibrated the measures to ensure the GPS measures accurately aligned with the roadway measures. At this point, the RID feature classes were incorporated as data tables in the PostgreSQL database.

Chapter 4 – Data Processing

General Approach

The purpose of data processing was to prepare data reductions that would provide the foundational data for the analyses of speeding. In the project Test Phase, the time series data from one or two data collection sites were used to develop, test, and refine the data processing tools and processes for preparing the data reductions. The final output of data processing included a set of final data reductions suitable for conducting analyses of speeding.

This chapter discusses the following aspects of data processing.

- Develop data processing tools
- Data preparation process
- Clean the data
- Reduce the data
- Quality testing and validation
- Ancillary processing
- Produce final speeding data reductions

Figure 9 illustrates the inputs, processing steps, and outputs in the data processing workflow.

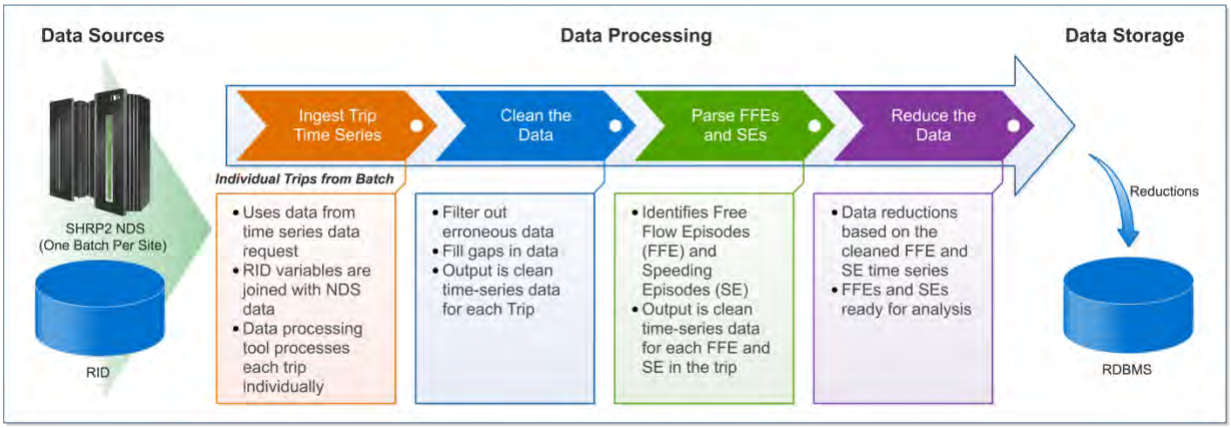

Figure 9. Data processing workflow

Develop Data Processing Tools

The data processing plan required tools that would (1) clean the data to remove or repair data quality issues that would affect the quality of the analysis; (2) parse the time series into Trip Episodes (Trips), Free Flow Episodes (FFEs), and Speeding Episodes (SEs); (3) prepare data reductions that describe each Trip, FFE, and SE associated with the respective Trip, FFE, and SE time series episodes; and 4) automate processing and status tracking.

Two options were considered for choosing the software platform to perform the data processing. In Option 1, processing would be done in ArcGIS using Python scripts. The data would be housed in a PostgreSQL repository, trips would be transferred to ArcGIS file geodatabases for processing, and the final data reductions and time series episodes would be transferred back to PostgreSQL for storage. In Option 2, the data would use native PostgreSQL and PostGIS functions to clean and process the data, using Python scripts to automate processing. Based on our experience using both products for processing GPS-based time series data, we considered the trade-offs associated with each option; Table 5 below summarizes some key advantages and disadvantages of each software platform. Note that this list contains only those features we considered to be most relevant to our requirements and is not intended to represent an exhaustive list of the pros and cons of each system. Also, note that processing speed is based on our experience running these tools on our specific hardware.

Table 5. Advantages and disadvantages of using ArcGIS versus PostgreSQL with PostGIS for data processing

Platform	Advantages	Disadvantages
ArcGIS	• Large library of functions for processing GIS data • Exceptional data visualization capability • Extensible • Python-enabled scripting capability	• Slower than PostgreSQL with PostGIS (longer processing time) • Can natively read from but not write to PostgreSQL tables[11] • Working with PostgreSQL tables in the ArcGIS catalog is clumsy
PostgreSQL with PostGIS	• Very fast and efficient processing • Large library of functions for processing GIS data • Data visualization and GIS processing capability through many third-party applications (e.g., QGIS, ArcGIS) • Extensible using a variety of programming languages • Python-enabled scripting capability • Aggregate and window functions to simplify data reduction	• SQL queries and functions are cumbersome to program and troubleshoot • Requires re-calibration of RID measures in the linear referencing system (LRS)

Given the large volume of data we would process, the speed and efficiency of PostgreSQL, and the ability to use ArcGIS and other third-party tools for data visualization and validation in conjunction with PostgreSQL, we chose to use PostgreSQL and PostGIS for data processing, with QGIS and ArcGIS for data visualization and validation, and Python for automating the data processing.

[11] The ability to bidirectionally interface ArcGIS with PostgreSQL depends on the ArcGIS license level. ArcGIS Server can both read and write PostgreSQL tables, but the ArcGIS Desktop Basic license level cannot write to PostgreSQL. The cost of procuring ArcGIS Server was out of scope for the project.

Tool development entailed multiple iterations of development and testing using the following steps.

1. Develop queries/functions/aggregates/triggers
2. Run pilot data and subset of the full data
3. Review the results
   - Procedural bugs
   - Problems based on data anomalies
4. Identify and implement fixes
5. Repeat steps 2-4

Once the data processing functions were developed, we prepared a Python script that automated the data processing procedure. Figure 10 shows the relationship between the components in the architecture of the data processing tool. For each trip, the script called each data processing function, tracked the status of processing at each step, logged status and error messages, and presented progress information. At this stage, we tested the script and examined the data reductions for a representative sample of trips. The calculations were verified by hand-calculating the reductions for a sample of speeding and free-flow episodes and comparing the results to the outputs from the data processing tool. We also examined the parsed time series data for FFEs and SEs to ensure that they were parsed according to their respective definitions.

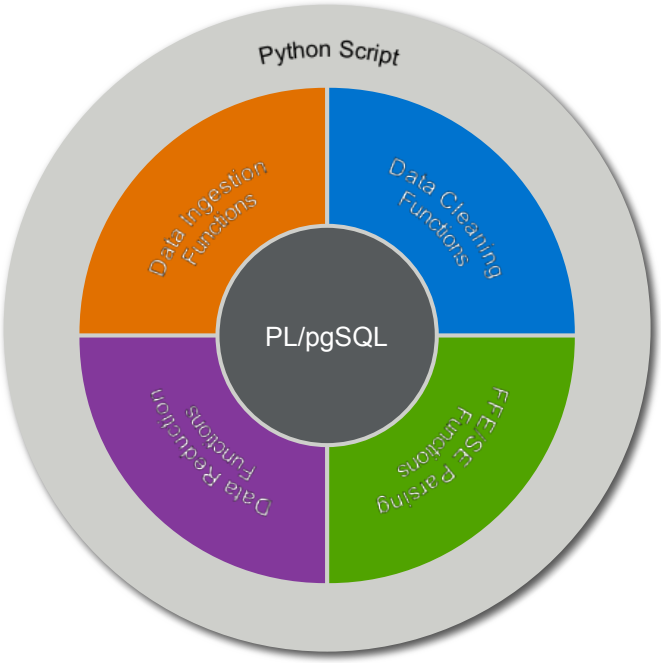

Figure 10. Data processing tool architecture

Data Preparation Process

Overview

The objective of data processing was to develop three time-series tables with Trip-, FFE-, and SE-level episodes, and three data reduction tables with one record per Trip, FFE, and SE in each respective table. The data processing was accomplished in the following steps, as illustrated in Figure 11.

- Ingest the trip time series
- Clean the data
- Parse the data into FFE, and SE time series episodes
- Calculate the data reductions
- Perform data quality testing and validation

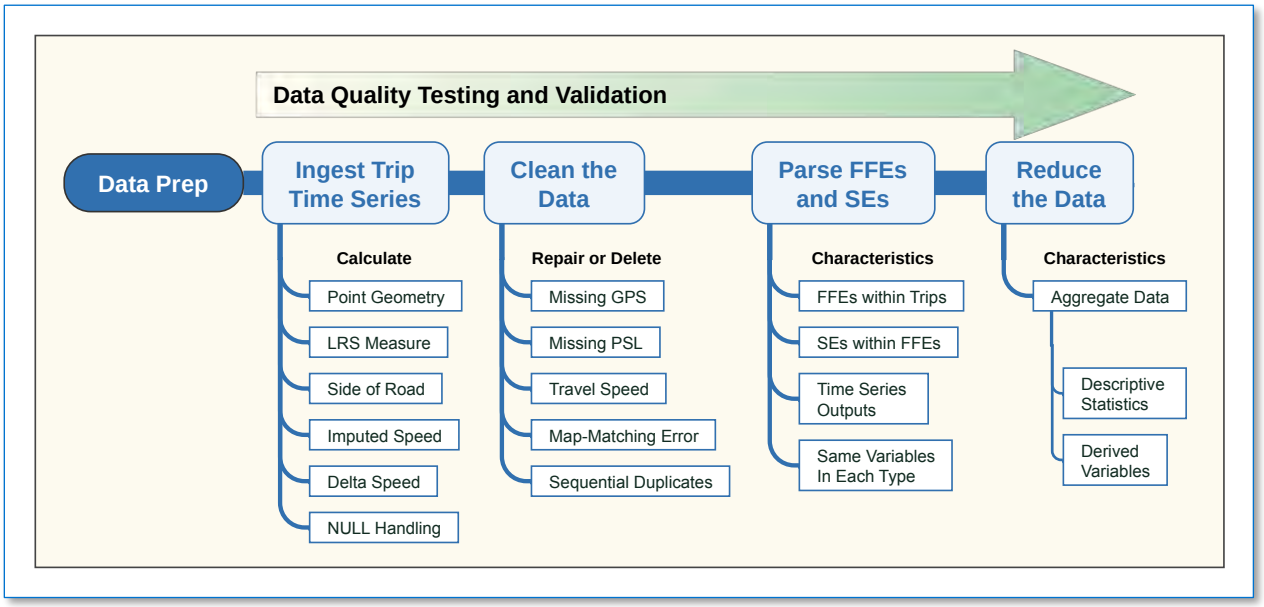

Figure 11. Data preparation process

Ingest Trip Time Series

The first step in data processing was to import the data from the source files VTTI delivered to us. The data were delivered in the form of CSV files. The function for importing the data performed some initial data processing, including calculating the point geometry from the GPS coordinates (i.e., latitude and longitude values), identifying the side of road the vehicle was driving on, and calculating speed-related derived variables. These included imputed speed and travel speed relative to posted speed limit (delta speed).

Travel Speed Units of Measure

One detail to consider is which units of measure to use for travel speed. Travel speeds in the NDS are reported in km/h, but speed limit is reported in mph in the RID. To be consistent with U. S. conventions for speed limit, we chose to convert the travel speed variables to mph.

Another aspect of data import was handling samples with variables that had no available data (i.e., NULL value). These are described below.

Calculate Point Geometry

In order to perform location-based GIS calculations, and to plot the data using GIS software, the GPS coordinates had to be converted to a format that GIS software could efficiently interpret and then snapped to the correct roadway. This procedure was performed using the LinkID variable, which was included in the time series data request. This variable indicates which road segment (link) the vehicle is driving upon at each coordinate and was extracted from the RID using a map-matching algorithm at VTTI. Per our data request, each GPS point was delivered to us with the latitude and longitude coordinates of the point, along with LinkID. The import function converted the latitude and longitude to a GIS geometry, in the "Well-Known Binary" (WKB) format, and snapped it to the closest location on the *Links* feature class having a LinkID common to the LinkID coded in the GPS sample. The RouteID from the Links feature class was then assigned to the GPS point for use in dynamic segmentation[12] processes throughout the remaining processing.

Some samples did not have values in the coordinate fields; for these samples, the geometry could not be calculated. There are two reasons the GPS coordinates were missing. First, VTTI purposely omitted the coordinates of a section of driving at the beginning and end of each trip to avoid exposing participants' personal identifying information (PII). Otherwise, the participant's home or work address could potentially be deduced from the source and destination of their trips. The time and/or distance of this "PII buffer" was randomly determined to further obfuscate the trip source and destination by varying the radius of exposure. Other variables were included in the PII buffer to identify provide context for the drive as the GPS location became available.

Some GPS coordinates were missing in some trips due to a variety of reasons. Occasionally, the GPS dropped out when traveling under bridges, in tunnels, or when other structures occluded or distorted the satellite signal. These types of errors were generally systematic, and the SHRP2 data were remarkably robust in overcoming data loss of this type. As an example of this robustness, Figure 12 illustrates the spatial accuracy of the GPS data locations while traveling in the I-90 tunnel on Mercer Island, Washington.

The yellow traces in the figure represent GPS locations. The traces appear to be solid lines outside of the tunnel because many individual points overlay one another. It can be seen that the as the vehicle enters into the tunnel from either direction, the vehicle location can be identified through roughly a third of the tunnel before there is not enough data to determine position. A small number of GPS samples can be seen through the entire tunnel.

12 See sidebar in the Calculate LRS Measure section below for a description of dynamic segmentation.

Figure 12. Example of GPS data while traveling through the I-90 Mercer Island Tunnel in the Washington site.

Some trips include no GPS coordinates at all, either because the entire trip was included in the PII buffer or there were data quality problems at VTTI. These trips were removed from the database, and data requests for subsequent time series data included a specification that trips without any GPS available not be delivered (see Table 4 on page 22 for the counts of trips requested versus received).

Calculate LRS Measure

The most efficient and accurate way to extract data from the RID at each GPS location uses dynamic segmentation based on the LRS defined by CTRE for the RID dataset and coded into the features in each feature class. The RID is designed to use dynamic segmentation for extracting information from the various feature classes. Simply put, it is an efficient method for extracting information at locations measured linearly along the roadway.

Dynamic Segmentation

Dynamic segmentation is a method of accurately extracting information from features based on their linear distance along the road, measured from the starting node of the road to the feature of interest. Multiple attributes from a variety of data sources can be associated with a location on the road (e.g., GPS location, RID characteristics, crash events). A strong foundational understanding of dynamic segmentation will facilitate efficient extraction of a variety of location-based data.

Once the point geometry was snapped to the *Routes* feature class, the linear measure along the roadway was calculated and assigned to the GPS point. Specifically, this measure is the linear distance in feet, measured along the roadway, from the start point of the route to the GPS point feature. The combination of WKB geometry and LRS linear measure provided a complete set of tools for supporting all relevant GIS operations, such as calculating distances between points, extracting RID information at a point, etc.

Identify Side of Road

An important function of the data import routine was to determine on which side of the road the vehicle was traveling. This was particularly important for extracting speed limits from the *SpeedLimit_ReducedDataset* feature class because speed limit is not guaranteed to be the same in both directions of travel on any road.

In most RID feature classes, side of road is defined by the direction of travel along the roadway. A GPS sequence (breadcrumb) is traveling on the right side of the road if the linear measure at each sequential point is increasing. That is, if the vehicle is moving in a direction from smallest linear measure toward highest linear measure, it is traveling on the right side of the road. Conversely, travel in the direction of decreasing measure is on the left side of the road. Most mobile van feature classes contain a variable called SideOfRoad that identifies left (coded as "L") or right (coded as "R") for each feature; there is a separate feature for each direction of travel on a particular road in most feature classes. During data import, we identified whether the measure at each GPS point was increasing or decreasing relative to the previous point, and the relevant 'R' or 'L' code was assigned to the point.

One exception to this coding scheme was found in the *SpeedLimit_ReducedDataset* feature class, which was the source of speed limit data for the project. This feature class does not provide the SideOfRoad variable directly. Rather, it indirectly provides the side of road based on a code that indicates the primary road orientation rather than direction of increasing/decreasing measure (see sidebar). Identifying the side of road traveled requires that the orientation of the GPS breadcrumbs compared with the orientation of the road be known.

Speed Limit Road Orientation

The Speed Limit feature class provides a code that indicates the general orientation of the road associated with the side of road. A west-to-east or north-to-south orientation is coded as '5' and an east-to-west or north-to-south orientation is coded as '6'. The Locations feature class can be joined with the Speed Limit feature class to extract side of road.

Although the orientation of many roads is relatively easy to identify, such as those as with north-south or east-west orientation, the orientation of other roads was more ambiguous. Roads that run north-east or north-west, for example, could be interpreted either as heading south to north or as east to west.[13] In the former case, a GPS breadcrumb traveling from south-west to north-east would use the data coded as traveling in the same direction as the general orientation of the road. If, on the other hand, the road is coded as generally oriented east to west, the GPS breadcrumb would be considered traveling against the general orientation of the road, leading to incorrect interpretation of direction of travel.

Another challenge to determining side of road from the orientation variable is that the local direction of travel at the section of roadway upon which the GPS is traveling is not guaranteed to be aligned along the general roadway orientation. A winding road that is generally heading north could head southward in small, localized sections, and GPS breadcrumbs across that section would appear to be opposing the general orientation rather than traveling with it.

[13] In general, roads were oriented from starting node in the south to ending node in the north, unless the road was perfectly east-west. However, there were roads that did not follow this convention.

To identify the side of road associated with each road segment in the *SpeedLimit_ReducedDataset* feature class, we used the Location feature class to tie the side of road to the road orientation. The Location feature class contains both the road orientation variable (Dir) and the SideOfRoad variable. To determine direction of travel, we extracted both variables from the Location feature class co-located with the corresponding Speed Limit feature and assigned SideOfRoad to the Speed Limit feature. Although this method was effective on most road segments, the process was inefficient because the Location feature class is large compared to other RID feature classes, which required substantially more processing time. Also, data quality checks revealed some inconsistencies and errors in this assignment, primarily because of multiple, co-located geometries creating duplicate records. Nevertheless, this method proved to be the most reliable for identifying the side of the road associated with direction of road orientation.

To improve the database performance, we added a column called SideOfRoad to the *SpeedLimit_ReducedDataset* feature class and populated it with values from the *Location* feature class. In this way, we only needed to extract the SideOfRoad from the *Location* feature class once for each road segment in the Speed Limit feature class rather than once for each GPS data point.

Identify Posted Speed Limit

The LRS measure and the SideOfRoad variables were used to extract the PSL from the *SpeedLimit_ReducedDataset* feature class at each GPS location using dynamic segmentation. The PSL was stored in the base time-series data table for later use in calculating delta speed, the difference between travel speed and PSL (see the discussion on page 32).

Calculate Imputed Speed

The SHRP2 time series data include two measures of travel speed: speed reported by the GPS and speed reported by the vehicle network via the onboard diagnostic (OBD2) port. When examining the data, we found that there could be as much as 5 mph difference (ignoring outliers) between GPS speed and network speed, and the difference in speed measures could not be corrected in any systematic or meaningful way. The difference between the two measures would vary over the life of the trip, and in some cases, the network speed was greater than GPS speed for part of the trip, and then later the GPS speed was greater than the network speed. If the actual travel speed was near the speeding threshold, one measure might be coded as speeding, while the other measure would be coded as not speeding.

Figure 13 shows the distribution of differences between the two travel speed measures for driving in the Seattle data collection site, for differences between -10 mph and 10 mph. As seen in the figure, most samples differed by only a few miles per hour; nevertheless, some differences not shown in the figure were quite large. In the largest outlier, the difference between speed measures was over 150 mph, which we attributed to GPS multipath error.

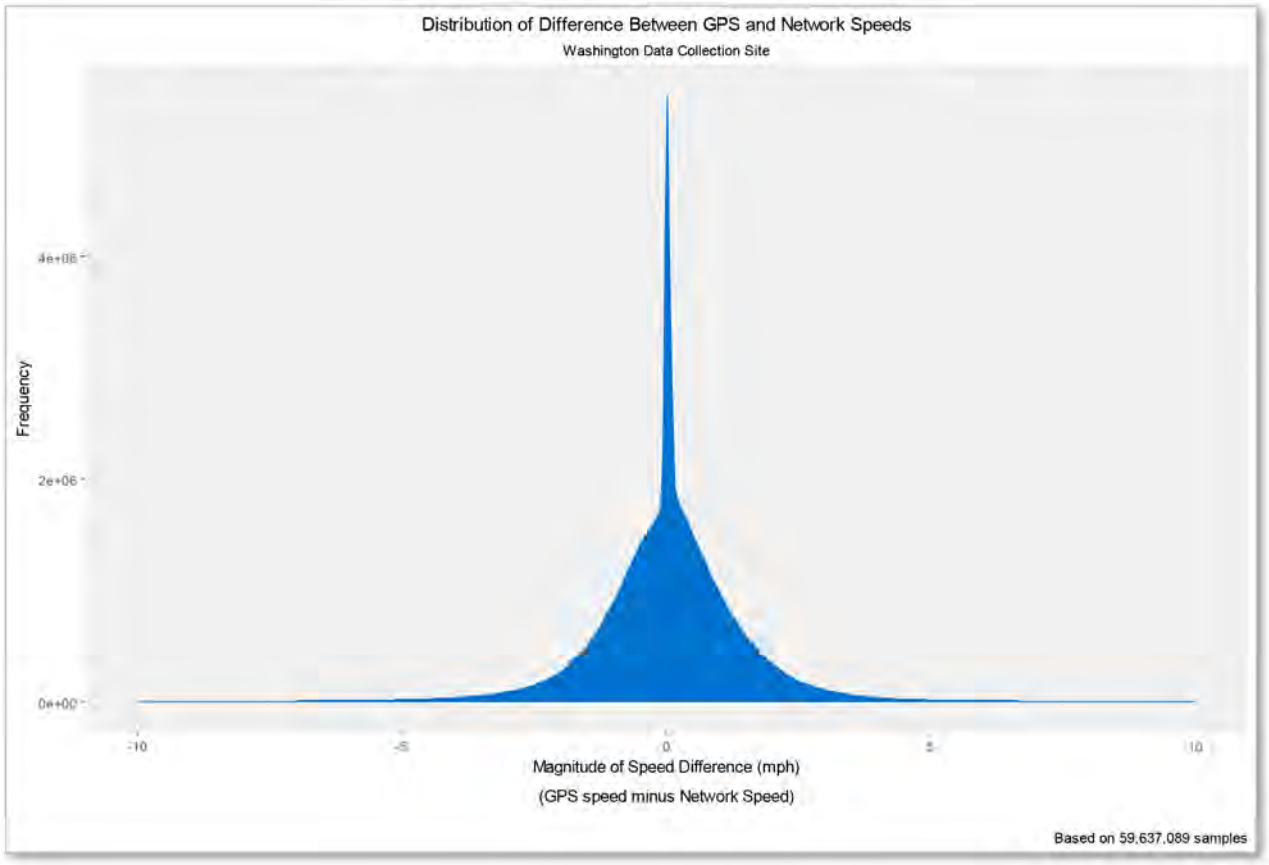

Figure 13. Distribution of error between GPS speed and network speed at the Seattle site

In addition to errors between travel speed measures, there were intermittent gaps in one or both measures, and in some cases no data were available for one of them (typically network speed). These discontinuities occurred in network speed due to intermittent problems with data collection equipment, and network speed was not available for older vehicles that did not have an OBD2 port. Similarly, GPS speed data losses occurred when the GPS signal was unavailable to the equipment or, occasionally, due to GPS equipment failure.

To account for these errors, we calculated an *imputed speed* using a strategy that selected the network speed whenever it was available and used the GPS speed whenever network speed was not available. If neither travel speed measure was available, the sample was assigned a value of NULL. We chose network speed as the primary travel speed measure because it was expected to be more accurate than GPS speed, which generally exhibits some lag, particularly when accelerating or decelerating. Also, network speed more accurately reflects the speed shown on the speedometer and therefore the reported speed drivers used to influence their decisions.

Calculate Delta Speed

Delta speed was an important variable that was used to identify travel speed relative to the posted speed limit. This variable was calculated as the difference between travel speed and posted speed limit shown in Equation 1.

$$v_{\Delta} = v_{imputed} - v_{PSL} \quad (1)$$

where:

v_{Δ} = delta speed
$v_{imputed}$ = imputed speed
v_{PSL} = posted speed limit

NULL Value Handling

Because of the way the vehicles in the study reported some variables via the vehicle's computer network, some data were not available for every 1 Hz sample in all variables. For example, some vehicles reported a value when there was a change in state (e.g., the vehicle reported valid wiper data only in samples when the motor was active or when the switch was turned on). Each sample that had no available data in a variable was imported to the database with a value of NULL. One challenge we had with importing these values was that NULL values in the initial CSV files were coded inconsistently across variables. Some variables were coded with nothing between the commas, while others were coded as an empty string (two consecutive double quotes) between the commas. Because PostgreSQL allows one definition of NULL during import, we could not directly copy the files into the database. To overcome this problem, we imported all data as strings into a temporary table and then copied them to the primary time-series table, casting each variable into its proper data type. In subsequent data requests, we requested that VTTI change their data extraction code for this project to provide all variables with nothing between commas to denote NULL values.

Clean the Data

The data required cleaning prior to use to ensure high quality, meaningful data reductions. By nature, naturalistic driving data are inherently messy because it is difficult or impossible to completely control all the conditions of data collection. Environmental conditions, data collection equipment malfunctions, variability in the way different vehicle makes and models (and even model years) report some variables from the CAN bus, and many other factors can cause intermittent discontinuities, noise, and/or inconsistencies in the data. Although the SHRP 2 data collection effort included real-time health checks, and every effort was made to ensure the highest quality data possible were captured both during and after data collection, there are unavoidable instances of data loss and/or noise in the data, as might be expected in naturalistic driving data. It should be noted that, while it may seem from the following discussion that the data were fraught with quality issues, they were generally full, robust, and of sufficient quality to produce high-quality analyses. Although specific vehicles or trips were characterized by noisy or missing data, by and large, we did not encounter data quality issues that could not be addressed.

A random sample of time series data from the Seattle site was used to identify what types of data quality issues needed to be cleaned and to test the algorithms as they were developed. Because

the SHRP2 Analysis of Speeding dataset is so large,[14] however, automated methods were used to detect and repair problems. With such a large dataset, it was not guaranteed that the random sample would include all possible data quality issues, nor was it possible to guarantee that all problems were perfectly detected and correctly addressed. Consequently, several rounds of development and testing were required to find and fix data quality issues to an acceptable degree. Data cleaning and quality testing was an ongoing process; data quality checks were performed after the data from each site were processed, and any new problems that were discovered were addressed. Once we were satisfied that data quality issues were addressed to an acceptable level and that the data cleaning algorithms were effective, all previously processed sites were re-processed to ensure that each site received the same cleaning and processing.

One important challenge we had to address early in the project was how to treat gaps in the data—one or more consecutive samples of a particular variable assigned NULL values—to prepare meaningful data reductions. Each SE and FFE was defined as a sequence of driving samples in which the vehicle traveled above the speeding threshold or free-flow threshold for a minimum length of time, with a minimum gap between episodes where the vehicle was traveling below the speeding or free-flow threshold (see Findings Report for more details). Discontinuities in the data, particularly samples missing GPS coordinates and travel speed measures, interfered with straightforward identification of SEs and FFEs; it was important to treat gaps in the data in a way that would compensate for these discontinuities in the data. For some variables, such as posted speed, it made sense to simply fill in the gaps where it was possible to identify posted speed with reasonable degree of confidence. Other variables, such as travel speed, were more challenging to fill because they were continually changing. Finally, a proportion of the data were missing the GPS coordinates; these could not be reliably estimated with any degree of accuracy.

Data cleaning and validation were conducted iteratively throughout the study. Quality checks were conducted after the data from each data collection site were processed to discover any new and different data errors that might be unique to each site and its characteristic driving environment. If a new error was found, a solution for fixing or filtering out the data was developed, and then all previous sites were re-processed to ensure that all sites were processed identically and errors were addressed uniformly across all sites.

Following are the steps we used to perform data cleaning to prepare the data for processing into Trips, FFEs, and SEs.

Recover Missing GPS Coordinates

The first step in data cleaning was to attempt to recover samples for which the geometry definition was not calculated during import because the latitude and longitude were not available for the sample. We attempted to estimate the location of these samples using the last known position and vehicle travel speed; however, differences between GPS and network speed, combined with latency in the GPS speed signal, resulted in obvious GPS location errors using this method. Furthermore, cumulative errors in estimated GPS location occurred when calculating position when multiple, consecutive points were missing coordinates. Eventually, we

[14] In this project, more than 356 million GPS time-series data samples were imported into the database. See Tables 6 and 7 for information about the volume of data processed at each data collection site.

decided to abandon GPS location estimation and simply discard samples that did not provide GPS location, since several important processes that depended upon location (extracting required RID variables, determining direction of travel and side of road, etc.) could not be performed reliably.

Missing RouteID

Additional cleaning was required in trips that had one or more samples in which the RouteID variable could not be extracted from the RID. RouteID is a critical variable defined in the *Routes* feature class of the RID. It uniquely identifies each road, usually (but not always) from end to end. All other road layers in the RID are commonly tied to the *Routes* feature class by the RouteID. It is used, along with GPS measure (i.e., the distance measured along the road from the roadway's starting node to the point at which the GPS point lies), to extract roadway characteristics at each GPS sample using dynamic segmentation.

As described in the Calculate Point Geometry section , the RouteID was extracted from the RID indirectly using the LinkID variable included in the time series data. LinkID was not available for some time series samples, however, because of problems with map matching at VTTI. Because extracting RouteID depended on knowing the LinkID, the RouteID could not be extracted for those GPS samples that did not include LinkID. To recover missing RouteIDs (i.e., fill the gap), we identified the RouteID samples just prior to and after the gap to make sure there was no change of road during the gap. If the values were the same, we filled the missing values with that RouteID. Otherwise, we left the missing values as NULL to avoid inadvertently coding the points with the wrong RouteID.

Missing Speed Limit

As discussed in the *Identify Posted Speed Limit* section above, PSL data were extracted from the *SpeedLimit_ReducedDataset* feature class using dynamic segmentation at each GPS location. Under certain conditions, the PSL data could not be extracted. First, the PSL was unavailable for samples that occurred in the PII buffer because GPS coordinates were redacted from the time series. These samples were removed from the Trip-level data, so no cleaning of the speed limit data was necessary for those missing PSL in the PII buffer. Second, we extracted PSL data only from mobile van data to ensure the highest and most consistent data quality. PSL was not included for samples not coded as associated with mobile van data; further cleaning of these data was not required. Finally, PSLs associated with samples that were missing GPS coordinates or RouteID could not be extracted because those variables were required to extract any RID data—including PSL.

To fill in gaps in PSL, we used a modal filter that assigned the most commonly occurring speed limit within in a moving window to each missing speed limit value in the time series. By trial-and-error, we found that a window comprising six samples before and six samples after the current sample yielded speed limit values that accurately filled the missing data. The filter calculated the modal value of samples within the same route on links with speed limits from the mobile van data. If the RouteID changed, the modal calculations from the previous route were cleared and new calculations were established. Speed limits were set to NULL for GPS locations on links with speed limits from sources other than the mobile van.

Missing Travel Speed

A small proportion of time-series samples contained neither network speed nor GPS speed measures. Because determining speeding and speed profile requires knowledge of the travel speed at each sample, it was not possible to fill these gaps in a meaningful way. For this initial investigation of speeding, we chose to discard samples with no travel speed, which, in effect, lowers the sample rate through the gap. Because the number of missing samples was small relative to the total number of samples in the dataset, this methodology was considered acceptable.

Once the gaps in PSL and travel speed were addressed, we recalculated the difference between travel speed and PSL for all samples in the dataset using the process described in *Calculate Delta Speed* above.

GPS Points Matched to the Wrong LinkIDs

One data quality check we employed was to determine if the LinkID provided by VTTI was correctly assigned to each time series sample. Figure 14 illustrates a trip in which LinkID was incorrectly assigned to the GPS points, leading to potentially assigning incorrect posted speed limit to the points.

To identify points with incorrect LinkIDs, we used the PostGIS ST_LineLocatePoint function to find the relative position of each point on the *Links* feature class. This function assigns a value between 0 and 1 representing the location of the closest point on the link (i.e., point projected onto the link) as a fraction of the total length of the link. A value of 0 or 1 indicates that either the point is located exactly on the beginning or end node of the link, or it is not located on the link at all. We considered GPS points to be mismatched when more than six consecutive samples were coded with 0 or 1 and the imputed speed indicated that the vehicle was not stopped. Small errors in GPS position—which are not unusual, particularly at slow speeds—can erroneously place the point outside a link, resulting in a 0 or 1. To differentiate between these and real link mismatches, we chose a sequence of at least six samples because the minimum

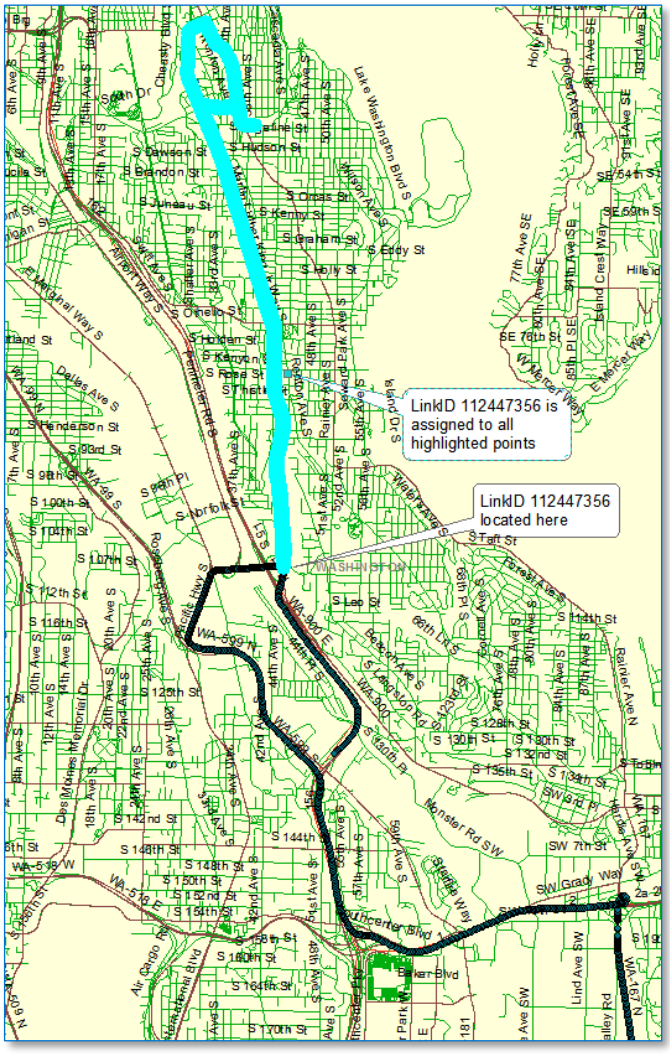

Figure 14. Illustration of GPS breadcrumbs with U-turn and incorrect LinkIDs

length of an SE was six seconds. This methodology found those GPS points for which the vehicle speed suggested it was moving for at least the time length of an SE while the relative location remained constant. Manual validation of a random sample of these cases suggested that the method was effective at locating mismatched points.

Using the methodology above, we discovered that some GPS points were not coded with the correct LinkID. As illustrated by Figure 14, this problem occurred when a vehicle exited a road segment (link) with LinkID *n* from one end (node) of the link, traveled on any number of links on the same or different roads, made a U-turn or loop, and re-entered the link at the same node from which it exited the link. Under these conditions, all the GPS points between leaving and reentering the link (i.e., during the U-turn) were assigned the same LinkID (i.e., LinkID *n*). This problem affected all GIS processes—including identifying speed limit—because the LinkID was used to identify RouteID, which in turn is a foundational variable used to extract all GIS road information. Consequently, GPS points with the wrong LinkID could potentially be assigned the wrong speed limit.

Implications of GPS-Road Mismatch

Although GPS points matched to the wrong LinkID ultimately did not affect the results in the current study, this error could impact future research. Researchers should exercise caution and check for mismatches in LinkID when identifying RID variables at GPS locations.

One option for addressing mismatched GPS points was to correct the erroneous points using map matching methods. Our experience with developing map matching algorithms, however, suggested that such an approach would be costly, time consuming, and out of scope for the project, particularly because the dataset was so large. Another approach was to process the data without correcting the mismatches and determine whether speeding occurred at those locations. This method was consistent with our general approach to error management throughout the project. Generally, if we could not apply cost-effective corrections to errors, we discarded data with Type 1 errors, in which the sample was incorrectly identified as speeding when no speeding occurred, but we ignored Type 2 errors, in which the sample was incorrectly identified as not speeding when speeding actually occurred. This methodology ensured we did not incorrectly identify speeding where none occurred, at the expense of potentially missing some speeding. We considered this an acceptable tradeoff because valid SEs were plentiful.

An examination of the SE time series data identified only one speeding episode across all six sites that occurred in GPS points that were matched to the wrong LinkID. We plotted this SE on a map and manually examined the location to determine the speed limits of the links located at each GPS point. A comparison of speed limits at each GPS location found that travel occurred on links with the same speed limit as the one erroneously assigned to the GPS points, and the delta speed values used to determine speeding were valid for the affected samples despite the mismatch. We could not, however, use these samples to identify other RID variables.

GPS coordinates, GPS speed, network speed, and/or longitudinal acceleration values with consecutive, identical values

One final data quality challenge we had to address was related to a phenomenon in which multiple consecutive samples in one or more variables contained exact duplicate values, down to

the maximum precision available for the variable. Generally, these variables included GPS coordinates, GPS speed, network speed, and/or acceleration values. Because these are continuous variables, and general noise in these types of signals make it extremely unlikely that they do not change at all, we suspected this was an artifact in the data. Because there was no way to recover the actual values, these records were discarded from the dataset. If most the trip was involved, the entire trip was discarded.

Parse the Data Into Trip, FFE, and SE Time Series Episodes

Once the data were cleaned, they were parsed into Trip, FFE, and SE time series episodes. A Trip episode consisted of the time series data for an entire trip, without any records from the PII buffer, and with additional helper variables for later producing the Trip, FFE, and SE data reductions. Examples of these helper variables included: elapsed time to the next PSL change, elapsed time from the previous PSL change, trip duration, speed variability (standard deviation of speed) across the entire trip, speed-based acceleration, change in steering angle, lateral position of lead vehicles, etc. Similarly, an FFE time series episode generally consisted of the same time series variables as the Trip episode, but parsed over the sequences of samples that met the criteria for free flow, and an SE time series episode was parsed over the sequence of samples that met the criteria for speeding. See the Findings Report for detailed definitions of Trip, FFE, and SE.

Two helper variables were used in the reductions to help inform the status of the posted speed limit before entering and after exiting a speeding episode. The first variable provided the elapsed time from the onset of the current PSL until the onset of the next PSL. Similarly, the second variable counted down the time from the onset of the current PSL until the onset of the next PSL. In this way, we could determine how long the driver had been driving at the current PSL before engaging in speeding, and how much time elapsed from the end of the speeding episode until the next PSL change. This information was used to identify speeding at PSL transitions. If the PSL changed from high-speed to low-speed, the driver might take some time decelerate to the current PSL. This condition would result in instantaneous speeding at the very first sample in the SE. If the PSL change from low-speed to high-speed, and the and driver was speeding on the low PSL road, this condition would be characterized by an instantaneous cessation of speeding. Posted speed limit transitions complicated the analysis and were not included in the current study. These variables made it possible for us to identify and discard speeding episodes that were not of interest in the current study. A benefit of this methodology is that these variables make it possible to analyze speed zone transitions in future analyses.

The Trip, FFE, and SE data were checked to ensure they were consistent with the respective definitions of Trips, FFEs, and SEs by manually examining a random sample of each. Each episode was examined for episode duration, speed with respect to free-flow or speeding threshold, and the duration and level of excursion below the threshold. The final output of this activity included three time-series tables with only those samples that were associated with cleaned Trips, FFEs, and SEs. Each table included identifiers that uniquely identified each Trip, FFE and SE. These unique ID numbers were used to link the SEs to FFEs and Trips in which

they were driven (i.e., the SE table contained FFE ID and Trip ID,[15] while the FFE table contained Trip ID).

Reduce the Data

Once the time series data were sufficiently cleaned and parsed in to Trip-, FFE-, and SE-level time-series episodes, data reductions were developed in preparation for analysis. Eighteen reductions were prepared—one each for Trips, FFEs, and SEs at each of the six data collection sites—using the Trip, FFE, and SE time series tables created in the previous step. The three reduction types contained the same variables to facilitate comparisons between speeding and non-speeding driving. Variables included descriptive statistics across each Trip, FFE, and SE for parameters such as travel speed, longitudinal and lateral acceleration/deceleration, environmental conditions (light level, wiper state as a surrogate for rain), vehicle state (cruise control, turn signal, brake and gas pedal position, etc.), lane and steering wheel positions, forward vehicle headway, etc. In addition, a series of variables that indicate the cumulative time driven above a range of both absolute and relative speed thresholds in 5 mph increments were developed. These "time above threshold" values were used in the analysis to identify speed profiles and the degree of speeding in each SE.

The reductions also included other variables, such as identifiers and timestamps, for keeping track of the association between the episodes in each table and the timing of these episodes. Each SE reduction included a field that indicated which FFE and Trip that contained the SE. Similarly, each FFE contained a field that indicated which Trip included the FFE. The timestamps at the beginning and end of each FFE and SE were included to identify elapsed time at which each FFE and SE occurred relative to the start of the Trip and the duration of each episode.

Quality Testing and Validation

Data quality checking occurred as an ongoing process throughout development of the data reductions. The first step was to test the various functions that calculated the individual variables. The majority of variables were calculated using aggregate window functions that are integral to PostgreSQL; these only required checking to ensure that variables were properly reduced across the correct episode (e.g., across an SE versus an FFE or Trip). All custom functions and aggregates we wrote in house were rigorously tested using multiple methods, including the use of dummy data designed to produce predictable outcomes and tests with subsamples of the NDS data to ensure they worked properly without side effects and within the proper range of values. Another technique used to test the outputs of the various functions and queries included calculating temporary flags that identified incorrect or impossible values and that could be queried or counted easily.

After the functions were validated using empirical means, we used statistical measures to check the data for potential anomalies that indicate possible errors.

[15] The Trip IDs generated in the Trip parsing were internally generated and are different from the Trip ID (or File ID) provided in the NDS data.

Ancillary Processing

During a preliminary analysis of the processed data, it became apparent that some additional information or cleaning would be helpful to better interpret the data in understanding speeding and classifying speeder types. Specifically, we needed a point-source reduction that provided some way to characterize speeding at a single point in each speeding episode. Also, we had not yet examined the effect of road characteristics on speeding. Finally, noise from the data acquisition system's accelerometers made it difficult to extract meaningful minimum and maximum acceleration. The following additional measures were calculated to address these shortcomings.

Max Speed Reductions

To represent speeding as a single point in the SE, we developed a new data file for each site that included the time-series record associated with the maximum speed in the speeding episode. The reduction also included the timestamps at the beginning and end of the SE to aid in assessing speeding profile by determining how long the driver took to reach maximum speed and to decelerate to non-speeding levels. To accommodate meaningful comparisons with non-speeding driving, we developed similar reductions of the FFE time series. These reductions provided the time-series record associated with the maximum speed in each FFE that did not include any speeding.

Preliminary RID Reductions

To test the extraction of RID variables at GPS locations and to conduct a preliminary examination of the effects of lane width, number of lanes, and other lane-related variables on speeding, we developed additional time series reductions. These reductions provided lane information at the point of highest speed for SEs and for FFEs that did not have any speeding to use for comparison. Like the maximum speed reductions, the lane reductions included the timestamps at the beginning of the SE and non-speeding FFEs. We also developed data reductions that reported roadway functional class at the beginning, middle, and end of each SE and FFE.

Considerations for analysis of roadway-related variables

The evaluation of road variables on speeding is more complicated than this preliminary examination suggests. The roadway environment leading up to the point of maximum speed is likely to be more important than the environment at the point of maximum speed.

Acceleration Measures

Because the longitudinal acceleration measure from the accelerometer sensors in the data acquisition equipment were inherently noisy, we found the measure produced unwanted side effects in our preliminary analyses. To augment the data from the accelerometers, we calculated a speed-based, imputed acceleration measure calculated by Equation 2, below.

$$a_{imputed} = \frac{v_t - v_{(t-\Delta t)}}{\Delta t} \quad (2)$$

where:

$a_{imputed}$ = delta speed

v_t = current speed
$v_{(t-\Delta t)}$ = last known speed
Δt = elapsed time between last known speed and current speed

Because imputed acceleration was based on vehicle speed, which was reported from the vehicle network in most cases, this method produced an accurate representation of acceleration and deceleration over the duration between samples. Although the sample rate was too slow to capture rapid decelerations, this measure proved adequate for the analyses we performed in this study.

Statistical Data Quality Checking

Because validation of the speed limit data was out of scope for the project, we used statistical methods to identify driving on road segments with potentially incorrect PSL. At locations that included consistent speeding, a heat map was developed using the ancillary Max Speed Reductions showing the number of seconds since the onset of the current SE versus the maximum speed relative to PSL. The points were color coded to indicate the level of max speed. Figure 15 shows one such location where there is a transition in PSL. The cluster of red-coded points at time zero indicates that speeding occurred at a change in PSL from low to high (suggesting drivers increased their speed before entering the higher speed zone. The second cluster of red-coded points between roughly 50 and 60 seconds into the SE, indicates there was a substantial level of speeding at a specific location. Although this condition could occur if some location-specific factor exists that encourages speeding, the hot-spot suggested that the PSL assigned to a short section of road could be incorrect. These locations were either checked on GIS maps and corrected where errors were found, or they were discarded from the final dataset as project resources became limited.

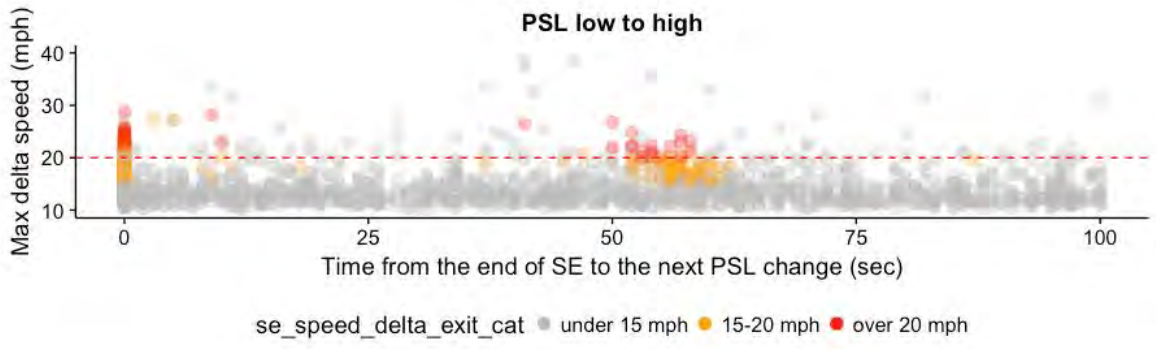

Figure 15. Illustration of a location with potential speed limit error

School Zones

We were concerned about how speed limits were coded for school zones. Because time of day in the time series was reduced into three-hour bins, and we had no data indicating the hours during which any given school zone speed limits were in force, it was not certain when drivers were speeding in the school zone or not. To avoid false positive indications of speeding, we ignored driving on roads with PSL below 35 mph.

Duplicate LinkIDs with Multiple Posted Speeds

Although the RID was largely robust and accurate, some locations contained more than one identical, co-located geometries with different PSLs. These often occurred on ramps or at some roads with designations by more than one jurisdiction (e.g., both State and interstate routes). During data processing, a separate record in the Trip, FFE, and SE was created with points on each geometry, causing duplicate records (except for posted speed). SEs with multiple duplicates were filtered out of the datasets by the statistician.

Ramifications of Ancillary Processing for Future Research

One advantage of our data processing software is that it incorporates a modular structure. The functionality described above in this section can be included in the data processing script, with minimal effort, for processing additional datasets. Indeed, functions for ancillary processing of all but the *Statistical Quality Checking* and *School Zone* filtering described above have already been developed and need only be called by data processing script.[16]

Produce Final Speeding Data Reductions

Once the test phase data from the Seattle site were developed, tested, and refined, and all of the data processing functions were finalized, the data from the Pennsylvania data collection site were acquired and processed using the same functions and process. Data quality checks were conducted to determine if any additional data challenges existed that were not identified in Seattle. Some additional data quality and processing issues were discovered, and the processing code was modified to address the problems. Once the data from Pennsylvania was successfully processed, the data from the other sites (New York, North Carolina, Indiana, and Florida) were requested. Data from each site were processed and checked for quality issues before processing the data for the next site. After all the other sites were processed, and no new data quality issues emerged, the Seattle site was re-processed using the final code to ensure that all changes to processing algorithms and code were consistently applied to the data from all sites. Finally, the SEs from each site were filtered to include only those with no PSL transitions within the SE.

[16] Although *Duplicate LinkIDs with Multiple Posted Speed Limits* filtering has not been fully developed (and was performed by the statistician in R), this functionality only requires modest changes in existing code to implement in routine processing.

Chapter 5 – Results, Discussion, and Lessons Learned

Results

Outcome of Data Preparation

Table 6 shows the number of GPS time series samples in each phase of data cleaning. These time series tables were used to calculate the data reductions that were used to perform the analyses. The Time Series (TS) column shows the number of raw GPS samples (sampled at one sample per second) we received in the data requests for each site. The Trip (Cleaned TS) column indicates the number of samples remaining in the dataset after removing samples from the PII buffer and other samples without GPS coordinates, PSL, etc., that could not be fixed. The FFE and SE columns show the number of time series samples in the FFEs and SEs identified at each data collection site.

Table 6. Counts of GPS samples in the various time series used to calculate the reductions.

Site	Time Series (TS)	Trip (Cleaned TS)	FFE	SE
Florida	97,390,351	68,036,677	44,130,585	6,929,022
Indiana	31,262,039	19,263,105	14,977,623	1,890,093
North Carolina	61,448,834	41,784,394	31,234,180	3,468,854
New York	76,698,482	58,078,009	39,564,349	4,302,895
Pennsylvania	23,703,828	19,114,532	15,802,964	2,855,013
Washington	65,745,334	43,532,076	30,645,512	2,088,767
Total	356,248,868	249,808,793	176,355,213	21,534,644

Table 7 shows the number of total and filtered Trips, FFEs, and SEs that were included in the data reductions for each site. Total SEs, FFEs, and Trips refers to the initial number of these episodes in the data reductions before filtering by the statistician (e.g., before removing those SEs with one or more PSL transitions). Filtered Trips, FFEs, and SEs constitute the dataset used to conduct the analyses after the statistician performed final filtering. See the Findings report for details about how the data were filtered.

Table 7. Counts of Trips, FFEs, and SEs in the final reductions.

Site	Trip (All)	FFE (All)	SE (All)	Trip (Filtered)	FFE (Filtered)	SE (Filtered)
Florida	62,762	360,159	190,455	61,183	236,432	26,965
Indiana	19,455	115,724	58,783	17,975	67,010	6,325
North Carolina	46,189	208,125	111,127	45,047	143,108	11,501
New York	64,773	365,538	142,073	63,924	261,941	14,044

Site	Trip (All)	FFE (All)	SE (All)	Trip (Filtered)	FFE (Filtered)	SE (Filtered)
Pennsylvania	15,443	96,390	75,438	15,061	57,734	9,476
Washington	45,567	160,620	81,554	43,571	117,222	9,870
Total	**254,189**	**1,306,556**	**659,430**	**246,761**	**883,447**	**78,181**

Dataset for Future Research

One of the goals of the study was to develop a dataset that can be used for conducting future research into speeding behavior. To that end, the database and associated data processing functions were designed to:

- Allow easy processing of additional time series data,
- Implement additional or improved data cleaning functions, and
- Facilitate additional types of analyses through development of new and different data reductions.

The existing structure of the data reduction was designed to accommodate a variety of analyses, and we expect that many types of analyses can be performed using these data reductions as they currently exist. Nevertheless, new and different data reductions can readily be performed by developing additional functions and reprocessing the data.

Processing additional data with the same outputs provided in the current speeding dataset requires minimal setup time. Data can be ingested into the database without any modifications to the processing script or functions, assuming the same variables are extracted as in the current time-series dataset, and data are delivered in a prescribed folder structure. Incorporating additional or improved cleaning or developing different data reduction datasets for another type of analyses requires developing the desired functions and modifying the shell Python script to call the new functions.

Discussion and Lessons Learned

This section briefly discusses issues we encountered and lessons learned while developing processes and tools, requesting data, and processing the data. This section is intended to help guide and inform decision-making, planning, and execution of projects for consumers of SHPR2 data and the RID.

Working With the NDS Data

Variable Selection

A deep and thorough understanding of the SHPR2 and RID datasets is critical to developing a robust analysis that can provide enough insight to answer key research questions without over-

burdening available project resources. There are hundreds of variables available in the NDS dataset, with more than 75 variables in the time-series dataset alone. With so many variables available, it is tempting to "throw the kitchen sink" at addressing the research questions; however, such an approach can be costly. One of the primary cost drivers for time-series data is the number of variables requested, and each additional time series variable can impact cost. Similarly, the rate at which the requested data are sampled can impact cost. The various time-series variables in the NDS were collected at different sampling rates, but they can be down-sampled to a lower sample rate to save cost—if doing so does not compromise the analyses. Furthermore, data sampled at a high sample rates can be burdensome to manage and process, requiring additional data storage and processing time compared with data delivered at a lower sampling rate. Of course, some analyses might require data sampled at high frequency, but substantial savings can be achieved by requesting data sampled at the lowest sample rate that can comfortably support the analysis.

Considerations for Variable Selection

Certain tradeoffs must be considered when selecting time series variables. Primary cost drivers for time series data extraction include the number of variables requested and the sample rate of the data extracted. Prioritizing variables and minimizing the sample rate can help optimize the amount of data that can be requested within the available budget.

Another consideration when requesting data is the intrinsic data quality of selected variables. Some variables are of inherently higher quality than others. For example, discussions with VTTI revealed that the head position variables were not very reliable; consequently, we eliminated those variables from consideration when selecting which data to use in subsequent tasks. Understanding what datasets are available, what the characteristics and limitations of each variable are, and even how the data were collected can help to identify those variables that will be the most cost-effective to use.

Data Requests

The SHRP2 NDS dataset provides a wealth of data that can be used to examine the problem of speeding. The volume of data available affords tremendous opportunity to address numerous research questions in a variety of ways; but it also creates unique challenges for the conduct of research projects. Researchers must balance variables required for analysis against budget resources. Extracting time series data can quickly become expensive if many variables are requested, and it takes time to extract data from such a large dataset. The following are recommendations for requesting data that can help to optimize the data received versus cost:

- Start planning early and plan sufficient time to work with VTTI to optimize cost vs. utility.
- The number of variables and the sample rate are primary cost drivers for acquiring time-series data. Careful selection of key variables can substantially improve the number of trips or cases that can be extracted when requesting time-series data.
- Some variables are more reliable than others, both in terms of quality and availability. Some variables, such as driver head position, may not be of sufficient quality to provide meaningful information. Availability of some variables is inconsistent between vehicle make, model, and/or year. The VTTI staff were knowledgeable and provided straightforward information about quality and availability of data during planning for data requests.

- Data from saved Insight queries can provide important insights without incurring significant costs. Although saved Insight queries are not free, they are relatively inexpensive to acquire.
- Some variables, such as longitudinal acceleration, can be noisy and may require filtering or smoothing to provide meaningful information.[17]
- When requesting data, it is important to provide as much clear, concise, and accurate detail about the data being requested, the form in which the data should be delivered, special considerations for the data, the data delivery mechanism, and any other important information related to the data request specification. Providing a detailed, formalized data request will reduce miscommunication, prevent errors, and minimize rework when requesting and receiving the data.

NDS Data Quality

Overall, the dataset provided ample data to perform our analyses. Nevertheless, naturalistic driving data, in general, are usually messy, and the NDS data is no exception. Although the level of difficulty associated with overcoming data quality challenges varied depending on the type of problems encountered, all data quality issues were readily identifiable in the data, and most challenges were addressed through cleaning (smoothing, interpolation in small gaps, etc.). Those data quality challenges that could not be fixed were addressed by selecting speeding thresholds that ensured false positive speeding errors did not occur. Additionally, because the dataset provided more than enough data, we were able to discard erroneous SEs—or even entire Trips—without compromising the sample size needed for confidence in the analyses.

NDS Data Scope

Working with such a large dataset requires substantial processing power and time to filter, clean, and process into data reductions. The choice of data processing software can have a significant impact on time required to develop data processing software and to process the data. The tradeoffs between software development efficiency and data processing efficiency should be considered when choosing data processing tools. For the current study, efficient data maintenance capabilities and processing efficiency were critical to the project, and the capabilities of the database approach we adopted outweighed the challenges associated with writing and debugging PL/pgSQL functions, which arguably can be more cumbersome that developing with procedural languages.

Working with the RID Data

The RID dataset provides a rich dataset for identifying roadway characteristics associated with location-based driving data. Because the RID does not contain driving data or Personally Identifying Information, the IRB and data privacy requirements do not apply. Consequently, acquiring the RID dataset is more straightforward than acquiring NDS data. It should be noted

[17] Findings about data quality of certain variables may change as VTTI and other researchers provide updated, cleaned, or improved datasets. We recommend discussing data quality concerns with VTTI when requesting data.

that the RID has been continually developed over the course of this project, and new reduced datasets have been included in the most recent version.

Data quality and usability

Overall, the RID provided a rich dataset for conducting our analysis. Quality of the data depends on the data source from which the variable was obtained. In past projects using data obtained directly from local and State jurisdictions, we found unacceptable levels of error, depending on the jurisdiction. The data captured in the SHRP2 S04B Mobile Van data collection effort, however, was generally very accurate and reliable, but not perfect (which is understandable given the size of the road network). Nevertheless, we were able identify sources of error and find ways to systematically or inferentially identify problem road segments.

Interpretation of Variables

It is important to understand how variables are coded in the NDS and RID datasets. In some cases, data that are technically not errors can be deceiving because of the way they are coded. One example in the RID[18] illustrates how such variable coding can affect the interpretation of the data. In the *Lanes* feature class, lane width captures only the right-most lane in the traveled way (O. Smadi, personal communication, February 9, 2017). If the lane includes parking, and there is no stripe separating parking from the traveled way, lane width is measured from the curb to the left lane marking, making the lane width artificially large. If cars are parked in the available space, the actual through way is much narrower than is coded in the RID.

Figure 16 below illustrates this phenomenon. In the figure, the right-most lane is coded in the RID as 9.158 ft. wide, and the width of the left lane (with the white vehicle traveling ahead) is unknown. Also, the center turn lane width is unknown. The oncoming lane is coded as 19.254 ft., which covers both the through way and room for parking. Looking only at the coded value without visual context, the oncoming lane seems wide. Because vehicles are usually parked there along the road during the day, the perceived width is much narrower than the coded lane width.

[18] Note that we used V1.1 of the RID. Subsequent versions of the RID have included an additional lane reduction that improves the coding of lane widths and provides additional information, such as average lane width.

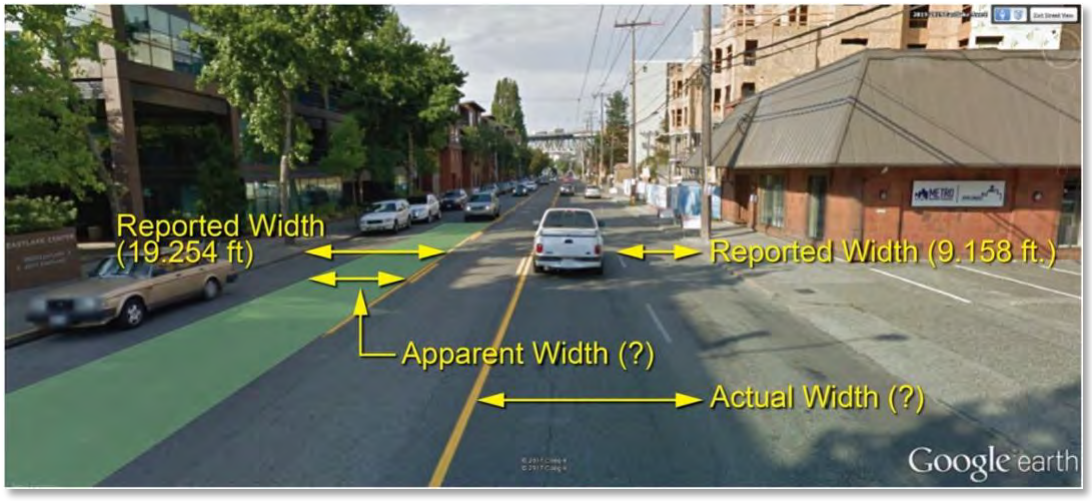

Figure 16. Example of reported versus perceived lane widths (RID V1.1)

Without this understanding of how the lane width is coded and perceived, interpretation of the data could lead to very different conclusions about the effect of lane width on speeding at that location.

RIDView Web-Based Tool

A useful tool provided by CTRE is the iVision RIDView web-based viewer (Smadi, 2015), available from CTRE. This tool provides a map with traces that indicates routes available in the RID dataset and video corresponding to a chosen route. This tool is invaluable for visually identifying or verifying roadway characteristics at specific locations. A user account is required to access the RIDView tool. Figure 17, shows a typical location from the Washington data collection site. In this example, a video frame shows the view of downtown Seattle from the mobile van video recording as the vehicle traveled southbound. An arrow on the corresponding map indicates the location and travel direction of the vehicle corresponding to the video frame. The table below the map and video frame provides information about the road segment begin traveled upon.

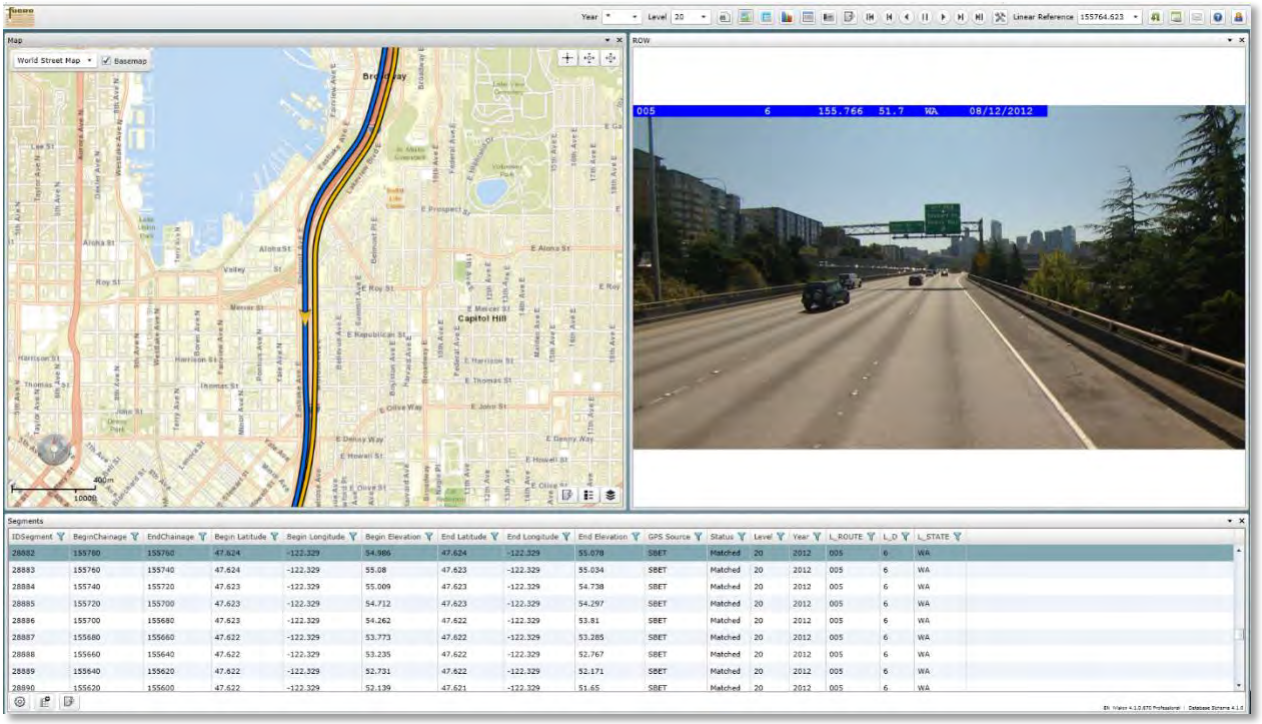

Figure 17. Web-based RID roadway visualization and information tool

Required expertise

The RID is delivered as a set of ESRI File Geodatabases, one for each data collection site. A working knowledge and expertise in general GIS principles, and of ESRI ArcGIS in particular, is essential for effectively working with the various datasets available in the RID. Furthermore, the RID feature classes support an LRS and require dynamic segmentation techniques to extract roadway information at a given location. A working knowledge and understanding of LRS and dynamic segmentation was required for us to extract posted speed limit and other roadway information from the RID. The most recent version of the RID (V2.1 as of this writing) includes a Dynamic Segmentation Tool (DST), which simplifies the process of dynamic segmentation and allow users to extract data without requiring a full understanding of LRS or dynamic segmentation.

Limitations of the Speeding Dataset

The data processing required to clean and prepare the data for analysis was a challenging process that required many iterations of development, testing, and refinement. Although the data was cleaned to a level that we considered to be sufficient for producing a high-quality analysis, several refinements to the algorithms and data processing function can be made to further improve the process and tools, add new functionality, and increase the number of SEs available in the data reductions. Following are some improvements that could help accomplish these goals.

Improved Cleaning

- Interpolation to fill small gaps in travel speed
- Improved imputed acceleration profile using signal processing methods for smoothing the accelerometer data
- Improved data quality detection and remedy for consecutive duplicates
- Improved data quality detection and remedy for multiple roads with same geometry in the RID

Improved Functionality

- Radar-based free-flow detection
- Speeding profile in episode
- Lane-change detection

References

ESRI. (2017). *GIS dictionary: dynamic segmentation* (Web site). Redlands, CA: Author. Retrieved from http://support.esri.com/en/other-resources/gis-dictionary/term/dynamic%20segmentation

International Standards Organization & International Electrotechnical Commission. (2016). *Information technology – Database languages – SQL multimedia and application packages – Part 3: Spatial.* ISO/IEC 13249-3:2016. Geneva: Authors.

International Standards Organization. (2015). *Geographic information – Well-known text representation of coordinate reference systems.* ISO 19162:2015. Geneva: Author.

International Standards Organization. (2004). *Geographic information – Simple feature access.* ISO 19125. Geneva: Author.

OpenStreetMap. (2017). *Database.* Retrieved from http://wiki.openstreetmap.org/wiki/Database

PostgreSQL. (2017). Featured users. Retrieved from https://www.postgresql.org/about/users/

Richard, C. M., Campbell, J. L., Lichty, M. G., Brown, J. L., Chrysler, S., . . . Reagle, G. (2013). *Motivations for speeding. Volume II: Findings Report* (DOT HS 811 818). Washington, DC: National Highway Traffic Safety Administration. Available at www.nhtsa.gov/sites/nhtsa.dot.gov/files/811818.pdf

Richard, C. M., Divekar, G., & Brown, J. L. (2014). *Motivations for speeding – Additional data analysis.* (Contract No. DTNH22-11-D-00229/0003). Washington, DC: National Highway Traffic Safety Administration.

Richard, C. M., Lee, J., Brown, J. L., & Landgraf, A. (2019). *Analysis of SHRP2 Speeding Data.* Washington, DC: National Highway Traffic Safety Administration.

Smadi, O. (January, 2015). *RIDView*. Ames, IA: Center for Transportation Research and Education, Iowa State University. Retrieved from http://RIDView.ctre.iastate.edu

Transportation Research Board. (2018). *Insight Data Access Website.* Retrieved from https://insight.shrp2nds.us

Appendix A. Glossary of Terms

This appendix provides a glossary of key terms related to Geographic Information Systems and Object-Relational Database Management Systems. It provides high-level, simplified definitions in the context of the Analysis of Speeding project to help the uninitiated reader understand the concepts discussed in this report.

ArcGIS File Geodatabase	Disk-based file structure that holds a collection of files that can store, search, and manage both geospatial and non-geospatial data. The ArcGIS file geodatabase file structure is a proprietary ESRI product
attribute	Additional information related to a spatial feature. Spatial features include geometry and optional attributes. Examples of attributes include instantaneous speed at a GPS location and lane width for a road feature.
CSV files	Plain text, flat-file data structure for storing tabular data. Comma Separated Value files are characterized by the following: one line for each record in a dataset, each record has the same sequence of fields, and fields are separated with commas or other delimiters. Although CSV files are highly portable and can be read by many applications and databases, there is no single, well defined data format required (beyond the basic structure defined above). Examples of variations include delimiter character (comma, tab, semicolon, etc.), presence or absence of quotes around text, type of quotes character (single versus double), and so forth.
Feature	A geographic object represented by a point, line, or area on a GIS map. Features can be naturally occurring or man-made. Examples of features include a GPS point, a road segment, a topological level contour, etc.
Feature class	Collection of GIS features of a common geometry type (i.e., points, lines, or polygons) and a common set of attributes. A feature class is typically stored in a single table in a geospatial database.
FFE	FFEs are Free-Flow Episodes, in which the driver has the opportunity to speed. FFEs include sequential driving above the free-flow threshold for at least 30 seconds, with at least 30 seconds between FFEs. Short gaps in the FFE, where speeding drops below the free-flow threshold for less than 30 seconds, are included in the FFE. The Free-Flow threshold was defined as 5 mph below the posted speed limit.

geometry	Spatial characteristics of a GIS feature. In the Analysis of Speeding project, point geometries are used to represent vehicle locations, and line geometries are used to represent road locations and alignments.
link	RID road segment that typically extends from one intersection to the next.
node	Point that is co-located with the location of either the start point or end point of a link, route, or other roadway line feature.
NULL values	Undefined values that represent missing data. Specifically, these occur when no data are provided in a field of a database record.
OBD2 port	Connector for attaching diagnostic or other computer equipment to the vehicle's onboard computer network.
PII buffer	Sequence of time series records at the beginning and end of a trip that do not include GPS latitude, longitude, or other information that could be used to identify the starting and ending locations of the trip, which could be used to identify the driver.
PostGIS	PostgreSQL extension that adds support for managing geographic objects in the PostgreSQL object-relational database. PostGIS provides more than 330 GIS functions that comply with the ISO 19125 and Open GIS Consortium (OGC) standards for storage and manipulation of point, line, polygon, multi-point, multi-line, and other geographic entities.
PostgreSQL extension	A modular add-on that provides additional functionality to the PostgreSQL Object-Relational Database.
rectangularization	Technique for aligning data sampled at different sample rates to a common, single sample rate. In the Analysis of Speeding project, the data were rectangularized using nearest neighbor rectangularization. In this technique, the GPS location, which was sampled at 1 sample per second, formed the "time base," and for all other variables, the sample with the nearest timestamp to the timestamp in the GPS location was selected, regardless of the variable's sample rate.
schema	Database structure that provides a way to organize and sequester data tables, functions, and other entities. In the Analysis of Speeding project, the database objects associated with each data collection site were stored in individual schemas for their respective site.
SE	SEs are Speeding Episodes that include sequential driving above the speeding threshold for at least six seconds, with at least six seconds between SEs. Short gaps in the SE, where speeding drops below the speeding threshold for less than six seconds, are

	included in the SE to account for speeding behavior near the threshold. The speeding threshold was defined as 10 mph above the posted speed limit.
stored procedure	Database functions developed to perform a pre-defined operation. In the Analysis of Speeding project, stored procedures were used to perform each step of data processing.
trip	Time series data that began when the driver started the ignition of the vehicle and ended when the driver cycled off the vehicle ignition.
Well-Known Binary (WKB)	Standard, portable representation of geographic geometries. WKB is a binary version of the Well-Known Text (WKT) representation of geometries, which is readable by humans. WKB provides a standards-compliant way to efficiently store and transfer geometric data for GIS objects and is defined in the ISO/IEC 13249-3:2016 (*Information technology – Database languages – SQL multimedia and application packages – Part 3: Spatial*) and ISO 19162:2015 (*Geographic information – Well-known text representation of coordinate reference systems*) standards.

Appendix B. Method for Selecting, Cataloguing, and Prioritizing SHRP2 Variables for Target Research Questions

Objective

The objective of this activity was to develop and apply a method of reviewing multiple data dictionaries, cataloging the variables described within each dictionary, and prioritizing variables for inclusion in a speeding database. The results of this activity were used to develop the final sampling and data analysis plans for the analyses of full data set.

Method

Variable Cataloging

Once we obtained an understanding of the variables in the NDS and RID datasets and listed all available variables, the variables were reviewed, and those that did not logically provide a potential contribution to the analysis (e.g., passenger weight, driver reaction to crash, etc.) were removed from further consideration. We eliminated 845 irrelevant variables from the NDS, which reduced our working set of variables to 192. Each remaining variable was classified under six factors corresponding to the specific research question the variable could address, the type of relationship to a research question, and the three possible variable types (crash/near-crash, speeding, and predictor). Note that the research question and variable type factors are not mutually exclusive; a variable may address more than one research question, likewise a variable may appear in more than one variable type factor. A description of the individual variable catalog factors is provided in Table B-1.

Table B-1. Variable catalog factors

Factor	Description
Research Question	Research question the variable could be used to analyze
Relationship	Relationship to a research question (direct, indirect, other)
Crash/Near-Crash Variable	Marked if the variable provides crash or near-crash-related data (binary coded)
Speeding Related Variable	Marked if the variable provides speeding-related data (binary coded)
Predictor Variable	Marked if the variable provides predictor data (binary coded)

Variables from all data dictionaries were entered into an Excel spreadsheet for analysis. The complete list of variables is provided in the *List of All Variables* section below.

Note that the SHRP2 NDS Trip Summary data were excluded from the data dictionary review. The Trip Summary data are intended to facilitate identifying data collection issues and mostly provide information regarding data quality (e.g., minimum and maximum observed values, measures of skew and kurtosis). As this data set only contains data derived from the time-series

data, no additional information for the analysis would be gained. Therefore, for this process of identifying variables' relationship to speeding-related factors, the SHRP2 NDS Trip Summaries were excluded.

Data Quality Variables

In addition to the variable catalog factors described in the following text, 27 of the 192 variables were identified as data quality or data management variables. These variables are required for joining and merging data sets, providing quality assurance checks, or performing similar functions within the analysis. For example, the variables *lane marking probability* and *number of satellites* provide information about the quality of the vehicle lane position variables and GPS data, respectively, but they are not used directly to answer research questions. Consequently, these were not included in the variable cataloging and prioritization process. The data quality variables are provided in the *Data Quality Variables* section below.

Variable Prioritization

A four-factor rating scale was used to rate each variable in the variables table. The four factors used were costs, benefits, accuracy, and availability. Variables were rated along each of the four factors using a scale of 1 to 3. The rating scale is provided in Table B-2. For this table, note that the availability ratings for the crash investigation variables reflect both the investigator's accessibility to the vehicle or other sources of information about the crash as well as the level of investigation (i.e., low-level crashes did not require as highly detailed investigations as more severe crashes).

Table B-2. Variable rating scale

Rating	Definition
Costs	3: Involves significant effort or time in processing or acquiring 2: Involves above-basic amount of effort/time in processing, yet less than required in Level 3 1: Involves a basic level of effort or time in processing or acquiring
Benefits	3: The variable is required to address at least one research question 2: The variable provides additional information beneficial to the research question 1: The variable does not provide information useful to a research question
Accuracy	3: High confidence in data accuracy 2: Lower confidence in data accuracy, may require additional processing/validation 1: Low confidence in data accuracy, will require additional processing/validation
Availability	3: Variable is available across all study vehicles; equipment generally reliable for variable 2: Variable may not be available across all study vehicles; variable may not always be available because of equipment reliability issues 1: Variable not available across all study vehicles

Priority Calculation

The four-factor rating scale was used to calculate a priority value for each variable. The priority value was calculated based on the formula:

$$Priority = Benefits + Accuracy + Availability - Cost$$

As the individual factors that composed the priority value had a range of 1 to 3, the minimum and maximum possible for priority was 0 and 8, respectively.

Utility Rating

Following the variable prioritization activity, a large number of candidate analysis variables were identified as suitable for the analysis (based on their non-zero priority rating). At this point, we do not know what the acquisition costs for the variables are; however, since most, if not all, of the variables in the data dictionaries should be available for "automatic extraction," it is possible that the full set can be obtained. Nevertheless, in the case that acquisition costs are a significant constraint, we further categorized the selected set of variables based on how necessary each is for the analysis and project objectives.

This utility rating process was performed independently from the variable cataloging and prioritization steps described above, yet with the knowledge and information gained from that process. The utility ratings associated with the categorization process had a range of 1 to 3, with 1 indicating that the variable indirectly supported the analysis, 2 indicating that the variable directly supported or was complementary to the analysis, and 3 indicating that the variable was required for the analysis. The utility rating scale is provided in Table B-3, and individual ratings from this activity were incorporated in the final variable selection table (Appendix C).

Table B-3. Utility rating scale

Level	Definition
3	The variable is required for the analysis
2	The variable is directly supporting or complementary to the analysis
1	The variable is indirectly supporting the analysis

Results

This section provides descriptive tables and charts that summarize the outcome of the different steps of the variable selection process for the selected variables. Of the initial 1,037 variables from all the data dictionaries, 845 variables were rated as being not useful for addressing our research questions (including trip summary variables), and 27 were data-quality variables. This resulted in a list of 165 priority variables. The summary tables and charts from the following Results sections only provide information about these 165 priority variables.

Variable Cataloging

This section summarizes information about the source and types of variables that were selected as part of the final variable list.

Figure B-1 shows the percentage of available variables by data dictionary. The majority of variables remaining in the analysis were from the crash event (35%) and time-series (31%) dictionaries, with the remainder coming from the RID, driver assessment, and crash event video reduction dictionaries.

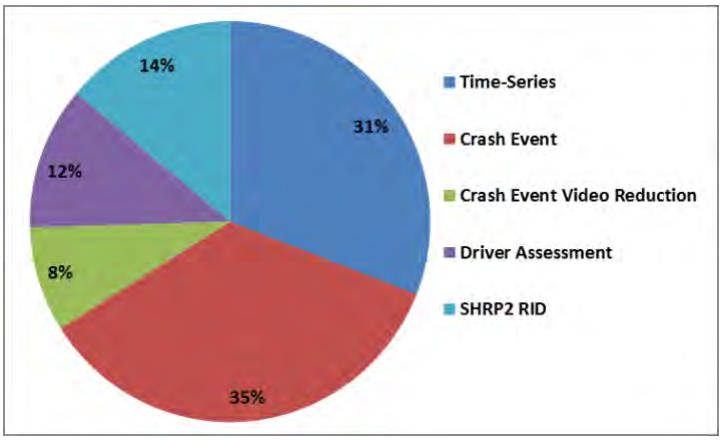

Figure B-1. Percentage of variables by dictionary

Of the 165 variables, 93 were identified from non-crash data sources (e.g., Time-Series data). A total of 72 were identified from crash data (e.g., Crash Event database). This distinction is made and explored in order to ensure adequate variable representation from both crash- and non-crash data sources in case of unanticipated delay or difficulty in obtaining a specific type of data. The number of variables from each type of data is provided in Table B-4.

Table B-4. Data sources for analysis

Source	Name	Variables in Dictionary	Variables in Analysis
SHRP2 NDS	Time-Series	76	51
SHRP2 NDS	Driver Assessment	506	19
SHRP2 RID	Roadway Information Database	36	23
All Non-Crash Data Sources	**Total**	**618**	**93**
SHRP2 NDS	Crash Event (Video Reduction)	54	14
SHRP2 NDS	Crash Detail	331	58
All Crash Data Sources	**Total**	**385**	**72**

In examination of the variables' relationships to research questions, the majority were marked as related to speed, demographics, and crash/near-crash involvement. Variables classification for the individual research question(s) they could address are provided in Table B-5. Note that

individual variables may be associated with multiple research question factors, thus the count does not sum to 165. Of the 165 variables' relationship to research questions, 68 (approximately 41%) were classified as direct, 54 (approximately 33%) were classified as indirect, and 43 (approximately 26%) were classified as other.

Table B-5. Number of variables marked per research question

Research Question	Variables Marked
Crash/Near-Crash	33
Demographics	40
Road Type	28
Speed	52
Time, Light, or Season	18

The variable classification by variable type (crash, near-crash, speeding-related, and predictor classification) is provided in Table B-6. Note that individual variables may be associated with multiple classifications (e.g., the ABS activation variable in the time-series data was classified as crash and near-crash related). Additionally, some variables associated with crashes were found outside of the crash event and crash event video reduction data (e.g., driver's air bag activation in the time-series data). Therefore, the information is this table is only provided as an overall descriptor of the data set.

Table B-6. Number of variables per variable classification

Variable Classification	Variables Marked
Crash/Near-Crash	91
Speeding-Related	39
Predictor	99

Variable Prioritization

Examination of the individual variable ratings assigned revealed that most costs associated with the 165 variables were classified as 1 (lowest cost), which reflects our overall strategy of focusing on variables that have already been coded and can be extracted using automated data processing routines. Seventy-two percent of variables were rated as such. Benefits ratings were more equally distributed, with approximately 44 percent and 39 percent being rated as 2 or 3, respectively. Accuracy ratings displayed a pattern similar to that of benefits, with approximately 51 percent and 48 percent being rated as 2 or 3, respectively. For availability ratings, a majority of variables were rated as 2 (approximately 67%) or 3 (approximately 26%). Seven percent of variables received an availability rating of 1. A summary of the variable ratings assigned across the four-factor rating scale is provided in Figure B-2.

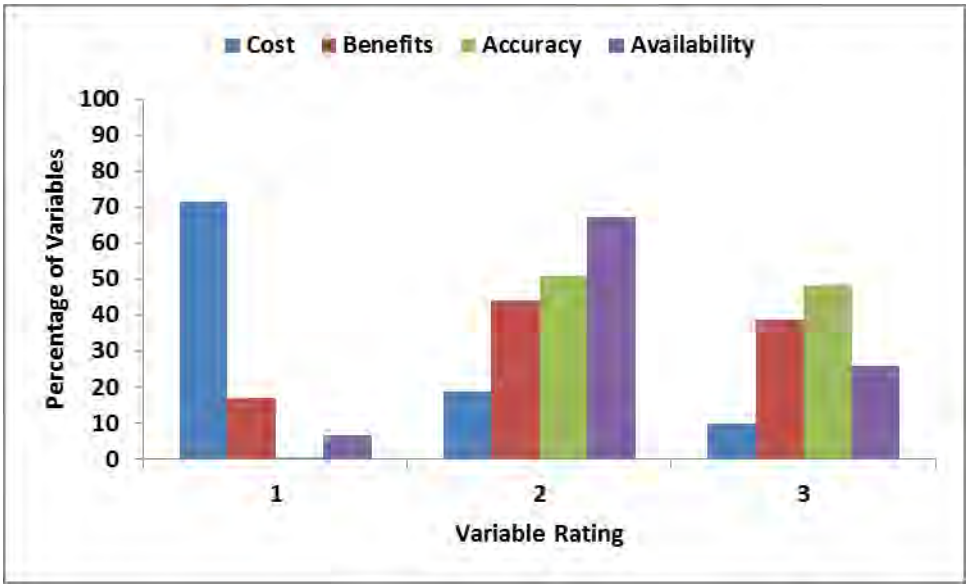

Figure B-2. Percentage of variable Ratings in Each Rating Category

The variable priority value was calculated for all 165 variables in the analysis. Although the potential range of priority values was from a minimum of 0 to a maximum of 8, no priority values between 0 and 2 were observed. This probably occurred because the variables that we assessed to be unrelated to the project objectives had already been removed before to the prioritization activity. The modal priority value was 6 (n = 41). The frequency of all variable priority values is provided in Figure B-3.

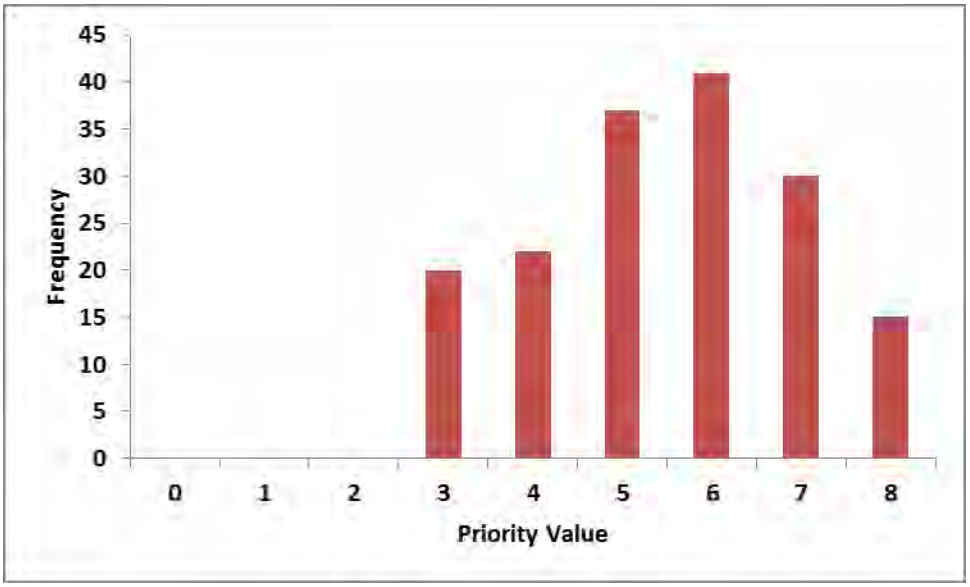

Figure B-3. Frequency of variable priority values

Utility Rating

Of the 165 priority variables examined, 46 variables (approximately 28%) were rated as 3 (required for the analysis). The list of required variables and their corresponding sources is shown in Table B-7. An additional 46 variables (28%) were assigned a utility rating of 2, which identified them as directly supporting or complementary to the analysis. The remaining 73 variables (≈44%), received utility ratings of 1, which identified them as indirectly supporting to the analysis. The *Final Candidate Variables* Section below provides the full list of priority variables, including their corresponding selection rating.

Table B-7. Variables with utility rating of 3 (required for the analysis) ordered by overall priority rating

Dictionary	Variable Name	Dictionary	Variable Name
Time-Series	Day	RID	Location End Long
Time-Series	Illuminance, Ambient	RID	Alignment Tangent
Time-Series	Month	RID	Alignment Curve Radius
Time-Series	Time	RID	Alignment Curve Length
Time-Series	Year	RID	Grade
Crash Event Video	Lighting	RID	Super-Elevation
Crash Event Video	Weather	RID	Lane Width
Time-Series	Latitude	RID	Lane Type
Time-Series	Longitude	Time-Series	Acceleration, x-axis
Time-Series	Speed, GPS	Time-Series	Acceleration, y-axis
Crash Event Video	Precipitating Event	Time-Series	Speed, Vehicle Network
Crash Event Video	Distraction 1	Driver Assessment	Demographic - Driver Education Level
Crash Event Video	Distraction 2	Driver Assessment	Demographic - Driver Marital Status
Crash Event Video	Distraction 3	Driver Assessment	Demographic - Driver Family Income
Crash Event Video	Traffic Density	Driver Assessment	Demographic - Driver Mileage Last year
Driver Assessment	Demographic - Driver Gender	Driver Assessment	Demographic - Driver Receive License
Driver Assessment	Demographic - Driver Birth Date	RID	Signs Message
Driver Assessment	Demographic - Driver Zip code	RID	Signs MUTCD Code

Dictionary	Variable Name	Dictionary	Variable Name
Driver Assessment	Driving History - Annual Mileage	RID	Intersection Number of Approaches
Driver Assessment	Driving History - Years of Driving	Time-Series	Radar, Range Rate Forward X Track n
RID	Location Begin Lat	Time-Series	Radar, Range Rate Forward Y Track n
RID	Location Begin Long	Time-Series	Radar, Range, Forward X Track n
RID	Location End Lat	Time-Series	Radar, Range, Forward Y Track n

Discussion and Conclusions

The objective of this activity was to develop and apply a method of reviewing data dictionaries, catalog the variables described within each dictionary, and prioritize variables for inclusion in a speeding database. The variable selection process was successful in producing a smaller data set and will be useful for omitting extraneous variables from planned data acquisition activities. This selection process resulted in a greatly reduced set of variables, with only those variables that are in support of the project objectives remaining.

The selection process still resulted in a large number of variables available for use in analyzing crash/near-crash events (91 variables identified), speeding-related events (39 variables identified), and predictors (99 variables identified). Beyond the initial selection process, analysis indicates that multiple variables are available to address each research question. Each research question has between 18 and 52 variables identified as relevant to the topic. This helps avoid the risk of problems with the analysis due to availability issues, mono-method biases, and other associated measurement/data threats. Analysis of the costs, benefits, accuracy, and availability ratings suggests that sufficient variability in the ratings was present to support the variable prioritization method that was developed and applied. This was demonstrated by the range and the distribution of variable priorities observed.

The selection process also had implications for data analysis. Specifically, it provided an approach that could be used to appropriately scale data requests to the available resources. We initially intended to request all 165 variables. Since we expected to rely almost entirely on automated variable extraction process at the SHRP2 data facility, we had no reason to believe that obtaining the full set was unfeasible. It should also be noted that almost half of these variables (77) were from the Driver Assessment and Crash-Event databases, which should have been trivial to obtain since they were standard survey-type responses rather than driving data. However, the initial data-acquisition cost estimates were higher than expected. Therefore, the priority rankings and tiered utility ratings were essential to scaling back to data request to ensure that the most important variables were included. At a practical level, the priority rankings and tiered utility ratings provided a simple basis for limiting the set of requested variables to a smaller number. For example, only 43 variables were assigned the highest utility rating.

Additional considerations, such as the potential for restrictions on data availability of personally identifying information also impacted what data we were able to request. A review of the 46 variables with a utility rating of 3 identified 2 variables that could be considered personally identifying information—latitude and longitude from the global positioning system (GPS). We were able to obtain these key variables by developing the data acquisition plan and study protocols in advance to maintain data confidentiality when working with these variables.

Variables Examined in the Data Dictionary Reviews

This section comprises lists of variables examined in the reviews of the data dictionaries and subsequent prioritization activities.

List of All Variables

This attachment provides all variables from the 5 different SHRP2 NDS, and the SHRP2 RID, data dictionaries identified in the reviews of the data dictionaries. The 1,037 variables listed in Table B-8 are organized by data dictionary and listed alphabetically.

Table B-8. All SHRP2 NDS and RID variables, by data dictionary

Dictionary	Name
Crash Event	Complete
Crash Event	Crash Details - Active Railroad Grade Crossing Near Site
Crash Event	Crash Details - Active Railroad Grade Crossing Near Site - [Other]
Crash Event	Crash Details - Activity Engaged In Prior to Crash
Crash Event	Crash Details - Activity Engaged In Prior to Crash - [Other]
Crash Event	Crash Details - Appendage Position
Crash Event	Crash Details - Appendage Position - [Other]
Crash Event	Crash Details - Average Daily Hours Worked Last Week
Crash Event	Crash Details - Average Daily Hours Worked Last Week - [Other]
Crash Event	Crash Details - Avoidance Actions
Crash Event	Crash Details - Avoidance Actions - [Other]
Crash Event	Crash Details - Brake Use
Crash Event	Crash Details - Brake Use - [Other]
Crash Event	Crash Details - Cargo Location
Crash Event	Crash Details - Cargo Location - [Other]
Crash Event	Crash Details - Cargo Presence
Crash Event	Crash Details - Cargo Shift
Crash Event	Crash Details - Cargo Shift Reason
Crash Event	Crash Details - Cargo Shift Reason - [Other]
Crash Event	Crash Details - Cell Phone Present Before Crash
Crash Event	Crash Details - Comfort Level With Cargo Load
Crash Event	Crash Details - Comfort Level With Passenger Load
Crash Event	Crash Details - Comfort Level With Vehicle

Dictionary	Name
Crash Event	Crash Details - Crash Type
Crash Event	Crash Details - Days Since Last Day Off
Crash Event	Crash Details - Driver Age
Crash Event	Crash Details - Driver Arguments In Last 12 Hours
Crash Event	Crash Details - Driver Arguments In Last 6 Hours
Crash Event	Crash Details - Driver Awareness
Crash Event	Crash Details - Driver Awareness - [Other]
Crash Event	Crash Details - Driver Education In Past
Crash Event	Crash Details - Driver Education Type
Crash Event	Crash Details - Driver Education Type - [Other]
Crash Event	Crash Details - Driver Gender
Crash Event	Crash Details - Driver Glance Location at Onset of Crash
Crash Event	Crash Details - Driver Glance Location at Onset of Crash - [Other]
Crash Event	Crash Details - Driver Glance Location Prior to Crash
Crash Event	Crash Details - Driver Glance Location Prior to Crash - [Other]
Crash Event	Crash Details - Driver Have Work Related Stress
Crash Event	Crash Details - Driver Have Work Related Stress - [Other]
Crash Event	Crash Details - Driver Height
Crash Event	Crash Details - Driver Hours Sleep Last 24 Hours
Crash Event	Crash Details - Driver Last Sleep Beginning Day
Crash Event	Crash Details - Driver Last Sleep Beginning Time
Crash Event	Crash Details - Driver Last Sleep End Day
Crash Event	Crash Details - Driver Last Sleep End Time
Crash Event	Crash Details - Driver Last Sleep Location
Crash Event	Crash Details - Driver Last Sleep Location - [Other]
Crash Event	Crash Details - Driver Narrative
Crash Event	Crash Details - Driver Personal Concerns
Crash Event	Crash Details - Driver Personal Concerns - [Other]
Crash Event	Crash Details - Driver Personal Concerns Immediately Prior to Crash
Crash Event	Crash Details - Driver Personal Concerns Immediately Prior to Crash - [Other]
Crash Event	Crash Details - Driver Requires Corrective Lenses
Crash Event	Crash Details - Driver Seat Belt Use
Crash Event	Crash Details - Driver Strenuous Household Activity
Crash Event	Crash Details - Driver Strenuous Household Activity Description
Crash Event	Crash Details - Driver Strenuous Recreational Activity
Crash Event	Crash Details - Driver Strenuous Recreational Activity Description
Crash Event	Crash Details - Driver Take Specific Education for Current Class of Vehicle
Crash Event	Crash Details - Driver Taken Medications In Last 24 Hours
Crash Event	Crash Details - Driver Using Cell Phone Prior to Crash

Dictionary	Name
Crash Event	Crash Details - Driver Visual Condition
Crash Event	Crash Details - Driver Visual Condition - [Other]
Crash Event	Crash Details - Driver Wear Hearing Aid
Crash Event	Crash Details - Driver Wearing Hearing Aid at Time of Crash
Crash Event	Crash Details - Driver Wearing Lenses at Time of Crash
Crash Event	Crash Details - Driver Wearing Lenses at Time of Crash - [Other]
Crash Event	Crash Details - Driver Wearing Prescription Sunglasses At Time of Crash
Crash Event	Crash Details - Driver Wearing Sunglasses at Time of Crash
Crash Event	Crash Details - Driver Weight
Crash Event	Crash Details - Driver's Ethnic Background
Crash Event	Crash Details - Driver's Race
Crash Event	Crash Details - Driver's Race - [Other]
Crash Event	Crash Details - Driver's Urgency
Crash Event	Crash Details - Eight Vehicle Lateral Movement
Crash Event	Crash Details - Eight Vehicle Lateral Movement - [Other]
Crash Event	Crash Details - EMS
Crash Event	Crash Details - EMS Agency - [Ambulance]
Crash Event	Crash Details - EMS Agency - [Fire department]
Crash Event	Crash Details - EMS Agency - [Other]
Crash Event	Crash Details - EMS Agency - [Police]
Crash Event	Crash Details - EMS Arrival Time
Crash Event	Crash Details - EMS Auto Response
Crash Event	Crash Details - EMS Contact
Crash Event	Crash Details - EMS Service
Crash Event	Crash Details - Endorsements on Driver License
Crash Event	Crash Details - Fifth Vehicle Lateral Movement
Crash Event	Crash Details - Fifth Vehicle Lateral Movement - [Other]
Crash Event	Crash Details - Fire Timing
Crash Event	Crash Details - Fire Timing - [Other]
Crash Event	Crash Details - Fourth Vehicle Lateral Movement
Crash Event	Crash Details - Fourth Vehicle Lateral Movement - [Other]
Crash Event	Crash Details - General Health
Crash Event	Crash Details - Headlight State
Crash Event	Crash Details - Health at time of Crash
Crash Event	Crash Details - Health at time of Crash - [Other]
Crash Event	Crash Details - Hospital Transport
Crash Event	Crash Details - How Driver Felt When Started Trip
Crash Event	Crash Details - How Driver Felt When Started Trip - [Other]
Crash Event	Crash Details - How Driver Normally Feels Upon Waking

Dictionary	Name
Crash Event	Crash Details - How Driver Normally Feels Upon Waking - [Other]
Crash Event	Crash Details - How Driver Tracked Source of Distraction to the Rear
Crash Event	Crash Details - How Driver Tracked Source of Distraction to the Rear - [Other]
Crash Event	Crash Details - Incident City
Crash Event	Crash Details - Incident Date
Crash Event	Crash Details - Incident State
Crash Event	Crash Details - Incident Time
Crash Event	Crash Details - Injury Occurrence
Crash Event	Crash Details - Injury Severity
Crash Event	Crash Details - Injury Specification
Crash Event	Crash Details - Interview Completion
Crash Event	Crash Details - Interview Completion Date
Crash Event	Crash Details - Jackknife
Crash Event	Crash Details - Jackknife Timing
Crash Event	Crash Details - Jackknife Timing - [Other]
Crash Event	Crash Details - License Restrictions
Crash Event	Crash Details - License State
Crash Event	Crash Details - License State - [Comment]
Crash Event	Crash Details - License State - [Foreign country (specify)]
Crash Event	Crash Details - License State - [Foreign country (specify)] [Comment]
Crash Event	Crash Details - License State - [Not licensed]
Crash Event	Crash Details - License State - [Not licensed] [Comment]
Crash Event	Crash Details - License State - [Unknown]
Crash Event	Crash Details - License State - [Unknown] [Comment]
Crash Event	Crash Details - License Validity
Crash Event	Crash Details - License Validity - [Other]
Crash Event	Crash Details - Line of Sight Clarity
Crash Event	Crash Details - Line of Sight Clarity - [Other]
Crash Event	Crash Details - Line of Sight Obstruction
Crash Event	Crash Details - Line of Sight Obstruction - [Other]
Crash Event	Crash Details - Longest Day Worked Last Week
Crash Event	Crash Details - Longest Day Worked Last Week - [Other]
Crash Event	Crash Details - Medical Attention
Crash Event	Crash Details - Nature of the Conversation
Crash Event	Crash Details - Nature of the Conversation - [Other]
Crash Event	Crash Details - Ninth Vehicle Lateral Movement
Crash Event	Crash Details - Ninth Vehicle Lateral Movement - [Other]
Crash Event	Crash Details - Normal Sleep Duration
Crash Event	Crash Details - Number of Lateral Movements

Dictionary	Name
Crash Event	Crash Details - Number of Passengers
Crash Event	Crash Details - Occupant Narrative
Crash Event	Crash Details - Other control Devices Near Site
Crash Event	Crash Details - Other control Devices Near Site - [Other]
Crash Event	Crash Details - Other Railroad Crossing Signs Near Site
Crash Event	Crash Details - Other Railroad Crossing Signs Near Site - [Other]
Crash Event	Crash Details - Other Signs Present
Crash Event	Crash Details - Other Signs Present - [Other]
Crash Event	Crash Details - Other Vehicle direction
Crash Event	Crash Details - Other Vehicle direction - [Other]
Crash Event	Crash Details - Other Vehicle Proximity
Crash Event	Crash Details - Other Vehicle Proximity -[Other]
Crash Event	Crash Details - Out Patient
Crash Event	Crash Details - Passenger 1 Age
Crash Event	Crash Details - Passenger 1 Gender
Crash Event	Crash Details - Passenger 1 Height
Crash Event	Crash Details - Passenger 1 Seat Belt Use
Crash Event	Crash Details - Passenger 1 Seating
Crash Event	Crash Details - Passenger 1 Weight
Crash Event	Crash Details - Passenger 2 Age
Crash Event	Crash Details - Passenger 2 Gender
Crash Event	Crash Details - Passenger 2 Height
Crash Event	Crash Details - Passenger 2 Seat Belt Use
Crash Event	Crash Details - Passenger 2 Seating
Crash Event	Crash Details - Passenger 2 Weight
Crash Event	Crash Details - Passenger 3 Age
Crash Event	Crash Details - Passenger 3 Gender
Crash Event	Crash Details - Passenger 3 Height
Crash Event	Crash Details - Passenger 3 Seat Belt Use
Crash Event	Crash Details - Passenger 3 Seating
Crash Event	Crash Details - Passenger 3 Weight
Crash Event	Crash Details - Passenger 4 Age
Crash Event	Crash Details - Passenger 4 Gender
Crash Event	Crash Details - Passenger 4 Height
Crash Event	Crash Details - Passenger 4 Seat Belt Use
Crash Event	Crash Details - Passenger 4 Seating
Crash Event	Crash Details - Passenger 4 Weight
Crash Event	Crash Details - Passenger 5 Age
Crash Event	Crash Details - Passenger 5 Gender

Dictionary	Name
Crash Event	Crash Details - Passenger 5 Height
Crash Event	Crash Details - Passenger 5 Seat Belt Use
Crash Event	Crash Details - Passenger 5 Seating
Crash Event	Crash Details - Passenger 5 Weight
Crash Event	Crash Details - Passenger 6 Age
Crash Event	Crash Details - Passenger 6 Gender
Crash Event	Crash Details - Passenger 6 Height
Crash Event	Crash Details - Passenger 6 Seat Belt Use
Crash Event	Crash Details - Passenger 6 Seating
Crash Event	Crash Details - Passenger 6 Weight
Crash Event	Crash Details - Passenger 7 Age
Crash Event	Crash Details - Passenger 7 Gender
Crash Event	Crash Details - Passenger 7 Height
Crash Event	Crash Details - Passenger 7 Seat Belt Use
Crash Event	Crash Details - Passenger 7 Seating
Crash Event	Crash Details - Passenger 7 Weight
Crash Event	Crash Details - Passenger Conversation
Crash Event	Crash Details - Passenger Conversation - [Other]
Crash Event	Crash Details - Passive Railroad Grade Crossing Near Site
Crash Event	Crash Details - Passive Railroad Grade Crossing Near Site - [Other]
Crash Event	Crash Details - Pre Crash Driver Activity
Crash Event	Crash Details - Pre Crash Driver Activity - [Other]
Crash Event	Crash Details - Pre-Crash Miles Driven
Crash Event	Crash Details - Pre-Crash Miles Driven - [Other]
Crash Event	Crash Details - Pre-Crash Vehicle Location
Crash Event	Crash Details - Pre-Crash Vehicle Location - [Other]
Crash Event	Crash Details - Pre-Crash Vehicle Movement
Crash Event	Crash Details - Pre-Crash Vehicle Movement - [Other]
Crash Event	Crash Details - Pre-Existing Conditions
Crash Event	Crash Details - Pre-Existing Conditions - [Other]
Crash Event	Crash Details - Pre-Existing Conditions Influence
Crash Event	Crash Details - Pre-Existing Conditions Influence - [Other]
Crash Event	Crash Details - Pre-Impact Fire
Crash Event	Crash Details - Pre-Impact Intent - [Accelerate]
Crash Event	Crash Details - Pre-Impact Intent - [Backup]
Crash Event	Crash Details - Pre-Impact Intent - [Change lanes to left]
Crash Event	Crash Details - Pre-Impact Intent - [Change lanes to right]
Crash Event	Crash Details - Pre-Impact Intent - [Merge]
Crash Event	Crash Details - Pre-Impact Intent - [Navigate curve]

Dictionary	Name
Crash Event	Crash Details - Pre-Impact Intent - [Other]
Crash Event	Crash Details - Pre-Impact Intent - [Slow down]
Crash Event	Crash Details - Pre-Impact Intent - [Stop]
Crash Event	Crash Details - Pre-Impact Intent - [Stopped, no movement intended]
Crash Event	Crash Details - Pre-Impact Intent - [Turn left]
Crash Event	Crash Details - Pre-Impact Intent - [Turn right]
Crash Event	Crash Details - Pre-Impact Intent - [Unknown]
Crash Event	Crash Details - Pre-Impact Intent - [U-turn]
Crash Event	Crash Details - Pre-Impact Intent -[Go straight]
Crash Event	Crash Details - Pre-Intersection Timing
Crash Event	Crash Details - Pre-Intersection Timing - [Other]
Crash Event	Crash Details - Regulatory Signs Near Site
Crash Event	Crash Details - Regulatory Signs Near Site - [Other]
Crash Event	Crash Details - Relationship Conversing Passenger
Crash Event	Crash Details - Relationship Conversing Passenger - [Other]
Crash Event	Crash Details - Roadway Familiarity 1
Crash Event	Crash Details - Roadway Familiarity 1 - [Other]
Crash Event	Crash Details - Roll Cause
Crash Event	Crash Details - Roll Cause - [Other]
Crash Event	Crash Details - Roll Direction
Crash Event	Crash Details - Roll Direction - [Other]
Crash Event	Crash Details - Roll Over
Crash Event	Crash Details - Roll Turns
Crash Event	Crash Details - Roll Turns - [Other]
Crash Event	Crash Details - School Zone Signs Near Site
Crash Event	Crash Details - School Zone Signs Near Site - [Other]
Crash Event	Crash Details - Second Vehicle Lateral Movement
Crash Event	Crash Details - Second Vehicle Lateral Movement - [Other]
Crash Event	Crash Details - Seventh Vehicle Lateral Movement
Crash Event	Crash Details - Seventh Vehicle Lateral Movement - [Other]
Crash Event	Crash Details - Shortest Day Worked Last Week
Crash Event	Crash Details - Shortest Day Worked Last Week - [Other]
Crash Event	Crash Details - Sign Presence
Crash Event	Crash Details - Sixth Vehicle Lateral Movement
Crash Event	Crash Details - Sixth Vehicle Lateral Movement - [Other]
Crash Event	Crash Details - Sleep Apnea Diagnosis
Crash Event	Crash Details - Sleep Apnea Treatment
Crash Event	Crash Details - Sleep Apnea Treatment - [Other]
Crash Event	Crash Details - Sleep or Work Schedule Rotate Last Week

Dictionary	Name
Crash Event	Crash Details - Specific Education Type
Crash Event	Crash Details - Specific Education Type - [Other]
Crash Event	Crash Details - Tenth Vehicle Lateral Movement
Crash Event	Crash Details - Tenth Vehicle Lateral Movement - [Other]
Crash Event	Crash Details - Third Vehicle Lateral Movement
Crash Event	Crash Details - Third Vehicle Lateral Movement - [Other]
Crash Event	Crash Details - Time Since Last Drivers Education
Crash Event	Crash Details - Times Driven Vehicle In Last 3 Months
Crash Event	Crash Details - Timing of Shift
Crash Event	Crash Details - Timing of Shift - [Other]
Crash Event	Crash Details - Total Hours Worked Last Week
Crash Event	Crash Details - Total Hours Worked Last Week - [Other]
Crash Event	Crash Details - Total Years Driving Experience
Crash Event	Crash Details - Traffic Control Device
Crash Event	Crash Details - Traffic Control Device - [Other]
Crash Event	Crash Details - Traffic Signal Operation
Crash Event	Crash Details - Traffic Signal Operation - [Other]
Crash Event	Crash Details - Travel Lane
Crash Event	Crash Details - Travel Lane - [Other]
Crash Event	Crash Details - Travel Speed
Crash Event	Crash Details - Trip Destination
Crash Event	Crash Details - Trip Destination - [Other]
Crash Event	Crash Details - Trip Origin
Crash Event	Crash Details - Trip Origin - [Other]
Crash Event	Crash Details - Trip Purpose
Crash Event	Crash Details - Trip Purpose - [Other]
Crash Event	Crash Details - Trip Start Time
Crash Event	Crash Details - Trip Start Time - [Other]
Crash Event	Crash Details - Trip Urgency
Crash Event	Crash Details - Turn Signal Use
Crash Event	Crash Details - Type of Cell Phone
Crash Event	Crash Details - Type of Cell Phone - [Other]
Crash Event	Crash Details - Urgency Reason
Crash Event	Crash Details - Urgency Reason - [Other]
Crash Event	Crash Details - Vehicle Avoidance Activation
Crash Event	Crash Details - Vehicle Avoidance Activation - [Other]
Crash Event	Crash Details - Vehicle Condition - [Brakes]
Crash Event	Crash Details - Vehicle Condition - [Brakes] [Comment]
Crash Event	Crash Details - Vehicle Condition - [Engine]

Dictionary	Name
Crash Event	Crash Details - Vehicle Condition - [Engine] [Comment]
Crash Event	Crash Details - Vehicle Condition - [Headlights]
Crash Event	Crash Details - Vehicle Condition - [Headlights] [Comments]
Crash Event	Crash Details - Vehicle Condition - [Steering]
Crash Event	Crash Details - Vehicle Condition - [Steering] [Comment]
Crash Event	Crash Details - Vehicle Condition - [Suspension]
Crash Event	Crash Details - Vehicle Condition - [Suspension] [Comment]
Crash Event	Crash Details - Vehicle Condition - [Tires]
Crash Event	Crash Details - Vehicle Condition - [Tires] [Comment]
Crash Event	Crash Details - Vehicle Condition - [Transmission]
Crash Event	Crash Details - Vehicle Condition - [Transmission] [Comment]
Crash Event	Crash Details - Vehicle Condition - [Wiring]
Crash Event	Crash Details - Vehicle Condition - [Wiring] [Comment]
Crash Event	Crash Details - Vehicle Fire
Crash Event	Crash Details - Vehicle Location at Crash
Crash Event	Crash Details - Vehicle Location at Crash - [Other]
Crash Event	Crash Details - Vehicle Speed Comparison
Crash Event	Crash Details - Vehicle Speed Comparison - [Other]
Crash Event	Crash Details - Warning Signs Near Site
Crash Event	Crash Details - Warning Signs Near Site - [Other]
Crash Event	Crash Details - Weather Condition
Crash Event	Crash Details - Weather Condition - [Other]
Crash Event	Crash Details - Weather Influence
Crash Event	Crash Details - Weather Influence - [Other]
Crash Event	Crash Details - Which Medications Driver Took In Last 24 Hours
Crash Event	Crash Details - Windshield Condition
Crash Event	Crash Details - Wiper Condition
Crash Event	Crash Details - Wiper State
Crash Event	Crash Details - Work / School Missed
Crash Event	Crash Details - Years Experience Driving Current Class of Vehicle
Crash Event	Crash Details -First Vehicle Lateral Movement
Crash Event	Crash Details -First Vehicle Lateral Movement - [Other]
Crash Event	Date Last Action
Crash Event	Date Started
Crash Event	ParticipantID
Crash Event Video	Alignment
Crash Event Video	Distraction 1
Crash Event Video	Distraction 1 End Sync
Crash Event Video	Distraction 1 Outcome

Dictionary	Name
Crash Event Video	Distraction 1 Start Sync
Crash Event Video	Distraction 2
Crash Event Video	Distraction 2 End Sync
Crash Event Video	Distraction 2 Outcome
Crash Event Video	Distraction 2 Start Sync
Crash Event Video	Distraction 3
Crash Event Video	Distraction 3 End Sync
Crash Event Video	Distraction 3 Outcome
Crash Event Video	Distraction 3 Start Sync
Crash Event Video	Driver Behavior 1
Crash Event Video	Driver Behavior 2
Crash Event Video	Driver Behavior 3
Crash Event Video	Driver Impairments
Crash Event Video	Driver Reaction
Crash Event Video	Driver Seatbelt Use
Crash Event Video	Event End
Crash Event Video	Event Nature
Crash Event Video	Event Severity
Crash Event Video	Event Start
Crash Event Video	Fault
Crash Event Video	Final Narrative
Crash Event Video	Hands on the Wheel
Crash Event Video	Incident Type
Crash Event Video	Infrastructure
Crash Event Video	Lighting
Crash Event Video	Locality
Crash Event Video	Maneuver Judgment
Crash Event Video	Motorist/Non-Motorist 2 Maneuver
Crash Event Video	Motorist/Non-Motorist 2 Reaction
Crash Event Video	Motorist/Non-Motorist 3 Maneuver
Crash Event Video	Motorist/Non-Motorist 3 Reaction
Crash Event Video	Motorist/Non-Motorist/Animal/Object 2 Location
Crash Event Video	Motorist/Non-Motorist/Animal/Object 2 Type
Crash Event Video	Motorist/Non-Motorist/Animal/Object 3 Location
Crash Event Video	Motorist/Non-Motorist/Animal/Object 3 Type
Crash Event Video	Number of Objects/Animals
Crash Event Video	Number of Other Motorists/Non-Motorists
Crash Event Video	Post-Maneuver Control
Crash Event Video	Precipitating Event

Dictionary	Name
Crash Event Video	Pre-Incident Maneuver
Crash Event Video	Relation to Junction
Crash Event Video	Subject Number
Crash Event Video	Surface Condition
Crash Event Video	Traffic Control
Crash Event Video	Traffic Density
Crash Event Video	Traffic Flow
Crash Event Video	Travel Lanes
Crash Event Video	Vehicle Contributing Factors
Crash Event Video	Visual Obstructions
Crash Event Video	Weather
Driver Assessment	ADHD Confidence Index
Driver Assessment	ADHD TScore Beta
Driver Assessment	ADHD TScore Commissions
Driver Assessment	ADHD TScore Dprime
Driver Assessment	ADHD TScore HitRT
Driver Assessment	ADHD TScore HitRTBlock
Driver Assessment	ADHD TScore HitRTIsi
Driver Assessment	ADHD TScore HitSE
Driver Assessment	ADHD TScore HitSEBlock
Driver Assessment	ADHD TScore HitSEIsi
Driver Assessment	ADHD TScore Omissions
Driver Assessment	ADHD TScore Perseverations
Driver Assessment	ADHD TScore VarSE
Driver Assessment	Age
Driver Assessment	Assessment Date and Time
Driver Assessment	Behavior - Wrong Lane at Intersection
Driver Assessment	Behavior - Brake Aggressively
Driver Assessment	Behavior - Disregards Speed Limits
Driver Assessment	Behavior - Driving Above Alcohol Limit
Driver Assessment	Behavior - Fail to Check Rearview Mirror
Driver Assessment	Behavior - Forget Where Car Is Parked
Driver Assessment	Behavior - Hit Something While Backing
Driver Assessment	Behavior - Miss Pedacyclist
Driver Assessment	Behavior - Miss Yield Signs
Driver Assessment	Behavior - No Recollection
Driver Assessment	Behavior - Pass a Turning Vehicle
Driver Assessment	Behavior - Pass on the Right
Driver Assessment	Behavior - Racing

Dictionary	Name
Driver Assessment	Behavior - Road Rage
Driver Assessment	Behavior - Run Red Light
Driver Assessment	Behavior - Tailgating
Driver Assessment	Behavior - Underestimate Speed of Oncoming Traffic
Driver Assessment	Behavior - Wrong Destination
Driver Assessment	Behavior - Wrong Gear
Driver Assessment	Behavior - Wrong Switch
Driver Assessment	Behavior - Wrong Way
Driver Assessment	Behavior -Miss Lead Vehicle
Driver Assessment	Behavior -Miss Pedestrians
Driver Assessment	Behavior -Roadway Aversion
Driver Assessment	Clock Drawing Test
Driver Assessment	Conner's Continuous Performance Test II - Summary Profile
Driver Assessment	Demographic - Driver Birth Date
Driver Assessment	Demographic - Driver Business Type for Study Vehicle
Driver Assessment	Demographic - Driver Business Use of Study Vehicle
Driver Assessment	Demographic - Driver Country of Birth
Driver Assessment	Demographic - Driver Education Level
Driver Assessment	Demographic - Driver Ethnicity
Driver Assessment	Demographic - Driver Family Income
Driver Assessment	Demographic - Driver Gender
Driver Assessment	Demographic - Driver Length of Vehicle Ownership
Driver Assessment	Demographic - Driver Living Status
Driver Assessment	Demographic - Driver Marital Status
Driver Assessment	Demographic - Driver Mileage Last year
Driver Assessment	Demographic - Driver Profession
Driver Assessment	Demographic - Driver Race
Driver Assessment	Demographic - Driver Receive License
Driver Assessment	Demographic - Driver Rental Status
Driver Assessment	Demographic - Driver Work Status
Driver Assessment	Demographic - Driver Zip code
Driver Assessment	Demographic - Eighth Additional Resident Age
Driver Assessment	Demographic - Eighth Additional Resident Drive
Driver Assessment	Demographic - Eighth Additional Resident Gender
Driver Assessment	Demographic - Fifth Additional Resident Age
Driver Assessment	Demographic - Fifth Additional Resident Drive
Driver Assessment	Demographic - Fifth Additional Resident Gender
Driver Assessment	Demographic - First Additional Resident Age
Driver Assessment	Demographic - First Additional Resident Drive

Dictionary	Name
Driver Assessment	Demographic - First Additional Resident Gender
Driver Assessment	Demographic - Fourth Additional Resident Age
Driver Assessment	Demographic - Fourth Additional Resident Drive
Driver Assessment	Demographic - Fourth Additional Resident Gender
Driver Assessment	Demographic - Household Population
Driver Assessment	Demographic - Length of Time at Residence
Driver Assessment	Demographic - Number of Vehicles in Household
Driver Assessment	Demographic - Second Additional Resident Age
Driver Assessment	Demographic - Second Additional Resident Drive
Driver Assessment	Demographic - Second Additional Resident Gender
Driver Assessment	Demographic - Seventh Additional Resident Age
Driver Assessment	Demographic - Seventh Additional Resident Drive
Driver Assessment	Demographic - Seventh Additional Resident Gender
Driver Assessment	Demographic - Sixth Additional Resident Age
Driver Assessment	Demographic - Sixth Additional Resident Drive
Driver Assessment	Demographic - Sixth Additional Resident Gender
Driver Assessment	Demographic - Third Additional Resident Age
Driver Assessment	Demographic - Third Additional Resident Drive
Driver Assessment	Demographic - Third Additional Resident Gender
Driver Assessment	Demographic - Vehicle 1
Driver Assessment	Demographic - Vehicle 2
Driver Assessment	Demographic - Vehicle 3
Driver Assessment	Demographic - Vehicle 4
Driver Assessment	Demographic - Vehicle 5
Driver Assessment	DHI - Button Pressed
Driver Assessment	DHI - Cumulative Total of Correct Answers String
Driver Assessment	DHI - Cumulative Total of Incorrect Answers String
Driver Assessment	DHI - Location of Test
Driver Assessment	DHI - Notes
Driver Assessment	DHI - Participant ID
Driver Assessment	DHI - Raw Walk Time
Driver Assessment	DHI - Test Date
Driver Assessment	DHI - Test Time
Driver Assessment	DHI - Total Time String
Driver Assessment	DHI - Useful Field of View Raw Score
Driver Assessment	DHI - Useful Field of View Word Score
Driver Assessment	DHI - Visual Search Summary Raw Score
Driver Assessment	DHI - Visual Search Test A Raw Score
Driver Assessment	DHI - Visual Search Word Score

Dictionary	Name
Driver Assessment	DHI - Visualizing Missing Information Raw Score
Driver Assessment	DHI - Visualizing Missing Information Word Score
Driver Assessment	DHI - Walk Time Word Score
Driver Assessment	DHI -Visual Search Test B Raw Score
Driver Assessment	Dominant Hand
Driver Assessment	Driver Knowledge - Bicycles
Driver Assessment	Driver Knowledge - Blind Spots
Driver Assessment	Driver Knowledge - City Driving
Driver Assessment	Driver Knowledge - Curve Signs
Driver Assessment	Driver Knowledge - Dimming Lights
Driver Assessment	Driver Knowledge - Drowsiness
Driver Assessment	Driver Knowledge - Emergency Vehicles
Driver Assessment	Driver Knowledge - Entering Expressways
Driver Assessment	Driver Knowledge - Fire Hydrants
Driver Assessment	Driver Knowledge - Green Arrows
Driver Assessment	Driver Knowledge - Light Changes
Driver Assessment	Driver Knowledge - Merge Signs
Driver Assessment	Driver Knowledge - Night Driving
Driver Assessment	Driver Knowledge - Police Officer
Driver Assessment	Driver Knowledge - Right of Way
Driver Assessment	Driver Knowledge - Run off Road
Driver Assessment	Driver Knowledge - Traffic Controls
Driver Assessment	Driver Knowledge - Yellow Lights
Driver Assessment	Driver Knowledge - Yellow Lines
Driver Assessment	Driving History - Annual Mileage
Driver Assessment	Driving History - Crash 1 Fault
Driver Assessment	Driving History - Crash 1 Severity
Driver Assessment	Driving History - Crash 2 Fault
Driver Assessment	Driving History - Crash 2 Severity
Driver Assessment	Driving History - Crash 3 Fault
Driver Assessment	Driving History - Crash 3 Severity
Driver Assessment	Driving History - Crash 4 Fault
Driver Assessment	Driving History - Crash 4 Severity
Driver Assessment	Driving History - Crash 5 Fault
Driver Assessment	Driving History - Crash 5 Severity
Driver Assessment	Driving History - Insurance
Driver Assessment	Driving History - Moving Violation Type
Driver Assessment	Driving History - Moving Violations
Driver Assessment	Driving History - Number of Crashes

Dictionary	Name
Driver Assessment	Driving History - Training
Driver Assessment	Driving History - Years of Driving
Driver Assessment	Gender
Driver Assessment	General TScore Beta
Driver Assessment	General TScore Commissions
Driver Assessment	General TScore Dprime
Driver Assessment	General TScore HitRT
Driver Assessment	General TScore HitRTBlock
Driver Assessment	General TScore HitRTIsi
Driver Assessment	General TScore HitSE
Driver Assessment	General TScore HitSEBlock
Driver Assessment	General TScore HitSEIsi
Driver Assessment	General TScore Omissions
Driver Assessment	General TScore Perseverations
Driver Assessment	General TScore VarSE
Driver Assessment	Integrated Systems - Cell Phone Directory Access
Driver Assessment	Integrated Systems - Cell Phone Integration
Driver Assessment	Integrated Systems - Cell Phone System Type
Driver Assessment	Integrated Systems - Factory Navigation
Driver Assessment	Integrated Systems - Have Nomadic MP3 Integration
Driver Assessment	Integrated Systems - Imbedded Navigation System Visual Display Location
Driver Assessment	Integrated Systems - Cell Phone Speech Recognition
Driver Assessment	Integrated Systems - Cell Phone Visual Display
Driver Assessment	Integrated Systems - Location for Cell Phone Controls
Driver Assessment	Integrated Systems - Music Control Inputs for Nomadic Devices
Driver Assessment	Integrated Systems - Nomadic MP3 Player Connection Type
Driver Assessment	Integrated Systems -Navigation System Activation
Driver Assessment	Left Hand Strength First Try
Driver Assessment	Left Hand Strength Second Try
Driver Assessment	Medical - Age-Related Conditions
Driver Assessment	Medical - Artificial Limbs
Driver Assessment	Medical - Brain Conditions
Driver Assessment	Medical - Cancer
Driver Assessment	Medical - Chronic Kidney Failure
Driver Assessment	Medical - Driving Vision Correction
Driver Assessment	Medical - Gave Up Driving
Driver Assessment	Medical - Hearing Conditions
Driver Assessment	Medical - Heart Conditions
Driver Assessment	Medical - Height

Dictionary	Name
Driver Assessment	Medical - Limited Flexibility
Driver Assessment	Medical - Metabolic Conditions
Driver Assessment	Medical - Multiple Medical Conditions
Driver Assessment	Medical - Multiple Medications
Driver Assessment	Medical - Muscle and Movement Disorders
Driver Assessment	Medical - Neck Size
Driver Assessment	Medical - Nervous System and Sleep Conditions
Driver Assessment	Medical - Other Brain Conditions
Driver Assessment	Medical - Other Hearing Conditions
Driver Assessment	Medical - Other Heart Conditions
Driver Assessment	Medical - Other Kidney Conditions
Driver Assessment	Medical - Other Medical Conditions
Driver Assessment	Medical - Other Metabolic Conditions
Driver Assessment	Medical - Other Musculoskeletal Disorders
Driver Assessment	Medical - Other Nervous System and Sleep Conditions
Driver Assessment	Medical - Other Psychiatric Conditions
Driver Assessment	Medical - Other Respiratory Conditions
Driver Assessment	Medical - Other Vascular Conditions
Driver Assessment	Medical - Other Vision Conditions
Driver Assessment	Medical - Paralysis
Driver Assessment	Medical - Prescribed Medications
Driver Assessment	Medical - Psychiatric Conditions
Driver Assessment	Medical - Respiratory Conditions
Driver Assessment	Medical - Severe Arthritis
Driver Assessment	Medical - Vascular Conditions
Driver Assessment	Medical - Vision Conditions
Driver Assessment	Medical - Vision Correction
Driver Assessment	Medical - Walking Aids
Driver Assessment	Medical - Weight
Driver Assessment	Neuro Confidence Index
Driver Assessment	Neuro TScore Beta
Driver Assessment	Neuro Tscore Commissions
Driver Assessment	Neuro TScore Dprime
Driver Assessment	Neuro TScore HitRT
Driver Assessment	Neuro TScore HitRTBlock
Driver Assessment	Neuro TScore HitRTIsi
Driver Assessment	Neuro TScore HitSE
Driver Assessment	Neuro TScore HitSEBlock
Driver Assessment	Neuro Tscore HitSEIsi

Dictionary	Name
Driver Assessment	Neuro TScore Omissions
Driver Assessment	Neuro TScore Perseverations
Driver Assessment	Neuro TScore VarSE
Driver Assessment	Optec - Color Score Fifth Circle
Driver Assessment	Optec - Color Score First Circle
Driver Assessment	Optec - Color Score Fourth Circle
Driver Assessment	Optec - Color Score Second Circle
Driver Assessment	Optec - Color Score Sixth Circle
Driver Assessment	Optec - Color Score Third Circle
Driver Assessment	Optec - Color Scoring Notes
Driver Assessment	Optec - Daytime Contrast Left Eye Row A
Driver Assessment	Optec - Daytime Contrast Left Eye Row B
Driver Assessment	Optec - Daytime Contrast Left Eye Row C
Driver Assessment	Optec - Daytime Contrast Left Eye Row D
Driver Assessment	Optec - Daytime Contrast Left Eye Row E
Driver Assessment	Optec - Daytime Contrast Right Eye Row A
Driver Assessment	Optec - Daytime Contrast Right Eye Row B
Driver Assessment	Optec - Daytime Contrast Right Eye Row C
Driver Assessment	Optec - Daytime Contrast Right Eye Row D
Driver Assessment	Optec - Daytime Contrast Right Eye Row E
Driver Assessment	Optec - Daytime Far Acuity Both Eyes
Driver Assessment	Optec - Daytime Near Acuity Both Eyes
Driver Assessment	Optec - Depth Perception
Driver Assessment	Optec - Nighttime Contrast Right Eye Row A
Driver Assessment	Optec - Nighttime Contrast Right Eye Row E
Driver Assessment	Optec - Nighttime Contrast Left Eye Row A
Driver Assessment	Optec - Nighttime Contrast Left Eye Row B
Driver Assessment	Optec - Nighttime Contrast Left Eye Row C
Driver Assessment	Optec - Nighttime Contrast Left Eye Row D
Driver Assessment	Optec - Nighttime Contrast Left Eye Row E
Driver Assessment	Optec - Nighttime Contrast Right Eye Row B
Driver Assessment	Optec - Nighttime Contrast Right Eye Row C
Driver Assessment	Optec - Nighttime Contrast Right Eye Row D
Driver Assessment	Optec - Nighttime Contrast With Glare Left Eye Row A
Driver Assessment	Optec - Nighttime Contrast With Glare Left Eye Row B
Driver Assessment	Optec - Nighttime Contrast With Glare Left Eye Row C
Driver Assessment	Optec - Nighttime Contrast With Glare Left Eye Row D
Driver Assessment	Optec - Nighttime Contrast With Glare Left Eye Row E
Driver Assessment	Optec - Nighttime Contrast With Glare Right Eye Row A

Dictionary	Name
Driver Assessment	Optec - Nighttime Contrast With Glare Right Eye Row B
Driver Assessment	Optec - Nighttime Contrast With Glare Right Eye Row C
Driver Assessment	Optec - Nighttime Contrast With Glare Right Eye Row D
Driver Assessment	Optec - Nighttime Contrast With Glare Right Eye Row E
Driver Assessment	Optec - Peripheral Vision Left Eye
Driver Assessment	Optec - Peripheral Vision Right Eye
Driver Assessment	Percent Beta
Driver Assessment	Percent Commissions
Driver Assessment	Percent Dprime
Driver Assessment	Percent HitRT
Driver Assessment	Percent HitRTBlock
Driver Assessment	Percent HitRTIsi
Driver Assessment	Percent HitSE
Driver Assessment	Percent HitSEBlock
Driver Assessment	Percent HitSEIsi
Driver Assessment	Percent Omissions
Driver Assessment	Percent Perseverations
Driver Assessment	Percent VarSE
Driver Assessment	Quick Screen - Difficulty Organizing
Driver Assessment	Quick Screen - Difficulty Waiting Turn
Driver Assessment	Quick Screen - Easily Distracted
Driver Assessment	Quick Screen - Feels Restless
Driver Assessment	Quick Screen - Loses Objects
Driver Assessment	Quick Screen -Difficulty Enjoying Leisure Activities
Driver Assessment	Raw Score Beta
Driver Assessment	Raw Score Commissions
Driver Assessment	Raw Score Dprime
Driver Assessment	Raw Score HitRT
Driver Assessment	Raw Score HitRTBlock
Driver Assessment	Raw Score HitRTIsi
Driver Assessment	Raw Score HitSE
Driver Assessment	Raw Score HitSEBlock
Driver Assessment	Raw Score HitSEIsi
Driver Assessment	Raw Score Omissions
Driver Assessment	Raw Score Perseverations
Driver Assessment	Raw Score VarSE
Driver Assessment	Response 1 – Response 360
Driver Assessment	Right Hand Strength First Try
Driver Assessment	Right Hand Strength Second Try

Dictionary	Name
Driver Assessment	Risk Perception - Bad Weather
Driver Assessment	Risk Perception - Checking Rearview Mirror
Driver Assessment	Risk Perception - Driving after taking Drugs or Alcohol
Driver Assessment	Risk Perception - Driving Sleepy
Driver Assessment	Risk Perception - Driving to Reduce Tension
Driver Assessment	Risk Perception - Driving While taking Drugs or Alcohol
Driver Assessment	Risk Perception - Eyes off Road
Driver Assessment	Risk Perception - Failure to Yield
Driver Assessment	Risk Perception - First off the Line
Driver Assessment	Risk Perception - Following Active Emergency Vehicles
Driver Assessment	Risk Perception - Illegal Turns
Driver Assessment	Risk Perception - In a Hurry
Driver Assessment	Risk Perception - Not Signaling
Driver Assessment	Risk Perception - Not Wearing Safety Belt
Driver Assessment	Risk Perception - Not Yielding to Pedestrians
Driver Assessment	Risk Perception - of Risk Aggressive Driving
Driver Assessment	Risk Perception - Passenger Interaction
Driver Assessment	Risk Perception - Passing on Right
Driver Assessment	Risk Perception - Risks for Fun
Driver Assessment	Risk Perception - Road Rage
Driver Assessment	Risk Perception - Rolling Stop
Driver Assessment	Risk Perception - Running Red Light
Driver Assessment	Risk Perception - Secondary Tasks
Driver Assessment	Risk Perception - Speeding for Thrill
Driver Assessment	Risk Perception - Speeding more than 20 mph Over Limit
Driver Assessment	Risk Perception - Tailgating
Driver Assessment	Risk Perception - Visual Obstructions
Driver Assessment	Risk Perception - Worn Tires
Driver Assessment	Risk Perception - Racing
Driver Assessment	Risk Perception - Running Stop Sign
Driver Assessment	Risk Perception - Speeding less than 20 mph Over Limit
Driver Assessment	Risk Perception - Yellow Light Acceleration
Driver Assessment	Risky - Accelerate At Yellow Light
Driver Assessment	Risky - Adjust CD Player
Driver Assessment	Risky - Change Lanes Suddenly
Driver Assessment	Risky - Drive After Drugs
Driver Assessment	Risky - Drive for Enjoyment
Driver Assessment	Risky - Drive Sleepy
Driver Assessment	Risky - Eyes Off Road To Passenger

Dictionary	Name
Driver Assessment	Risky - Fail to Yield
Driver Assessment	Risky - Failure to Adjust
Driver Assessment	Risky - First Off Line
Driver Assessment	Risky - Follow Emergency Vehicles
Driver Assessment	Risky - Make Illegal Turns
Driver Assessment	Risky - Merge Without Checking Rearview Mirror
Driver Assessment	Risky - Not Use Belt
Driver Assessment	Risky - Not Use Signal
Driver Assessment	Risky - Not Yield Pedestrians
Driver Assessment	Risky - Pass On Right
Driver Assessment	Risky - Pass When Visibility Obscured
Driver Assessment	Risky - Race Other Cars
Driver Assessment	Risky - Road Rage
Driver Assessment	Risky - Roll Through Stop Sign
Driver Assessment	Risky - Run Red Lights
Driver Assessment	Risky - Run Stop Signs
Driver Assessment	Risky - Secondary Tasks while Driving
Driver Assessment	Risky - Speed 10-20 mph Over
Driver Assessment	Risky - Speed 20+ mph Over
Driver Assessment	Risky - Speed For Thrill
Driver Assessment	Risky - Tailgate
Driver Assessment	Risky - Take Risks Because Of Hurry
Driver Assessment	Risky - Take Risks For Fun
Driver Assessment	Risky - Using Drugs while Driving
Driver Assessment	Risky - Worn Tires
Driver Assessment	Sensation Seeking - High Dive
Driver Assessment	Sensation Seeking - Alcohol at Party
Driver Assessment	Sensation Seeking - Body Odors
Driver Assessment	Sensation Seeking - Communication
Driver Assessment	Sensation Seeking - Contact with Swingers
Driver Assessment	Sensation Seeking - Dangerous Activities
Driver Assessment	Sensation Seeking - Date Personalities
Driver Assessment	Sensation Seeking - Exploring City
Driver Assessment	Sensation Seeking - Friend Personality
Driver Assessment	Sensation Seeking - Home Movies
Driver Assessment	Sensation Seeking - Illicit Drug Use
Driver Assessment	Sensation Seeking - Jet Set Lifestyle
Driver Assessment	Sensation Seeking - Learn to Fly
Driver Assessment	Sensation Seeking - Marijuana Use

Dictionary	Name
Driver Assessment	Sensation Seeking - Meeting New People
Driver Assessment	Sensation Seeking - Mountain Climbing
Driver Assessment	Sensation Seeking - New Experiences
Driver Assessment	Sensation Seeking - Parachuting
Driver Assessment	Sensation Seeking - Patience
Driver Assessment	Sensation Seeking - Perception of Art
Driver Assessment	Sensation Seeking - Predictable Movie Plot
Driver Assessment	Sensation Seeking - Recreational Drug Use
Driver Assessment	Sensation Seeking - Rewatching Movies
Driver Assessment	Sensation Seeking - Sailing
Driver Assessment	Sensation Seeking - Scuba Diving
Driver Assessment	Sensation Seeking - Sex in Movies
Driver Assessment	Sensation Seeking - Sexual Experience Before Marriage
Driver Assessment	Sensation Seeking - Skiing
Driver Assessment	Sensation Seeking - Social Drinking
Driver Assessment	Sensation Seeking - Social Sin
Driver Assessment	Sensation Seeking - Staying at Home
Driver Assessment	Sensation Seeking - Style of Dress
Driver Assessment	Sensation Seeking - Summary Metric
Driver Assessment	Sensation Seeking - Summary Metric
Driver Assessment	Sensation Seeking - Summary Metric
Driver Assessment	Sensation Seeking - Summary Metric
Driver Assessment	Sensation Seeking - Surfing
Driver Assessment	Sensation Seeking - Trip Planning
Driver Assessment	Sensation Seeking - Try New Foods
Driver Assessment	Sensation Seeking - Type of Parties
Driver Assessment	Sensation Seeking - Views on Homosexuality
Driver Assessment	Sensation Seeking - Water Skiing
Driver Assessment	Sensation Seeking - Witty Friends
Driver Assessment	Sensation Seeking -Unpredictable Friends
Driver Assessment	Sleep Questionnaire - Doze While Working the Day Shift
Driver Assessment	Sleep Questionnaire - Doze While Working the Night Shift
Driver Assessment	Sleep Questionnaire - Average Sleep Hours When Not Working
Driver Assessment	Sleep Questionnaire - Average Sleep Hours When Working
Driver Assessment	Sleep Questionnaire - Average Sleep Needed
Driver Assessment	Sleep Questionnaire - Awake Earlier Than Want
Driver Assessment	Sleep Questionnaire - Awakened By Children Last Month
Driver Assessment	Sleep Questionnaire - Awakenings At Night
Driver Assessment	Sleep Questionnaire - Bed Time When Working From the Home

Dictionary	Name
Driver Assessment	Sleep Questionnaire - Bed Time When Working Outside the Home
Driver Assessment	Sleep Questionnaire - Children At Home
Driver Assessment	Sleep Questionnaire - Children At Home Eleven To Thirteen
Driver Assessment	Sleep Questionnaire - Children At Home Fourteen To Eighteen
Driver Assessment	Sleep Questionnaire - Children At Home Less Than Two
Driver Assessment	Sleep Questionnaire - Children At Home Older Than Eighteen
Driver Assessment	Sleep Questionnaire - Children At Home Six To Ten
Driver Assessment	Sleep Questionnaire - Children At Home Three To Five
Driver Assessment	Sleep Questionnaire - Days Off Four Weeks Ago
Driver Assessment	Sleep Questionnaire - Days Off Last Week
Driver Assessment	Sleep Questionnaire - Days Off Three Weeks Ago
Driver Assessment	Sleep Questionnaire - Days Off Two Weeks Ago
Driver Assessment	Sleep Questionnaire - Doze In a Public Place
Driver Assessment	Sleep Questionnaire - Doze While In A Car Stopped Temporarily
Driver Assessment	Sleep Questionnaire - Doze While Lying Down
Driver Assessment	Sleep Questionnaire - Doze While Reading
Driver Assessment	Sleep Questionnaire - Doze While Talking to Someone
Driver Assessment	Sleep Questionnaire - Fatigued Upon Waking
Driver Assessment	Sleep Questionnaire - Fatigued While Awake
Driver Assessment	Sleep Questionnaire - Four Weeks Ago Typical
Driver Assessment	Sleep Questionnaire - Frequency Awake Between 20 and 24 Hours
Driver Assessment	Sleep Questionnaire - Frequency Awake Between 24 and 30 Hours
Driver Assessment	Sleep Questionnaire - Frequency Awake More Than 30 Hours
Driver Assessment	Sleep Questionnaire - Frequency Awakened By Children Last Month
Driver Assessment	Sleep Questionnaire - Frequency Night Shifts Last Year
Driver Assessment	Sleep Questionnaire - Frequency Work Start Before Five AM
Driver Assessment	Sleep Questionnaire - Functioning While Awake
Driver Assessment	Sleep Questionnaire - Greatest Number Continuous Hours Worked Last Month
Driver Assessment	Sleep Questionnaire - Hours Awakened By Children Last Month
Driver Assessment	Sleep Questionnaire - Hours Spent Sleeping the Past Week
Driver Assessment	Sleep Questionnaire - Hours Spent Sleeping the Week Four Weeks Ago
Driver Assessment	Sleep Questionnaire - Hours Spent Sleeping the Week Three Weeks Ago
Driver Assessment	Sleep Questionnaire - Hours Spent Sleeping the Week Two Weeks Ago
Driver Assessment	Sleep Questionnaire - Hours Spent Working the Past Week
Driver Assessment	Sleep Questionnaire - Hours Spent Working the Week Four Weeks Ago
Driver Assessment	Sleep Questionnaire - Hours Spent Working the Week Three Weeks Ago
Driver Assessment	Sleep Questionnaire - Hours Spent Working the Week Two Weeks Ago
Driver Assessment	Sleep Questionnaire - Last Week Alcohol Servings
Driver Assessment	Sleep Questionnaire - Last Week Caffeine Intake Pattern

Dictionary	Name
Driver Assessment	Sleep Questionnaire - Last Week Caffeine Servings
Driver Assessment	Sleep Questionnaire - Last Week Typical
Driver Assessment	Sleep Questionnaire - Nap Frequency
Driver Assessment	Sleep Questionnaire - Nap Length
Driver Assessment	Sleep Questionnaire - Nod Off Last Month
Driver Assessment	Sleep Questionnaire - Nod Off Last Year
Driver Assessment	Sleep Questionnaire - Nod Off While Driving
Driver Assessment	Sleep Questionnaire - Nod Off While Driving Frequency
Driver Assessment	Sleep Questionnaire - Occupation
Driver Assessment	Sleep Questionnaire - Quality Of Sleep
Driver Assessment	Sleep Questionnaire - Quit Breathing During Sleep
Driver Assessment	Sleep Questionnaire - Quit Breathing During Sleep Frequency
Driver Assessment	Sleep Questionnaire - Sleep Aid Type
Driver Assessment	Sleep Questionnaire - Sleep Duration
Driver Assessment	Sleep Questionnaire - Sleep In Recliner or Sitting
Driver Assessment	Sleep Questionnaire - Sleep Schedule
Driver Assessment	Sleep Questionnaire - Sleep Status When Working From the Home
Driver Assessment	Sleep Questionnaire - Sleep Status When Working Outside the Home
Driver Assessment	Sleep Questionnaire - Sleeper Type
Driver Assessment	Sleep Questionnaire - Sleepiness While Awake
Driver Assessment	Sleep Questionnaire - Sleepy During Daytime
Driver Assessment	Sleep Questionnaire - Snoring
Driver Assessment	Sleep Questionnaire - Snoring Bother Others
Driver Assessment	Sleep Questionnaire - Snoring Frequency
Driver Assessment	Sleep Questionnaire - Snoring Loudness
Driver Assessment	Sleep Questionnaire - Three Weeks Ago Typical
Driver Assessment	Sleep Questionnaire - Time To Fall Asleep
Driver Assessment	Sleep Questionnaire - Tobacco Use
Driver Assessment	Sleep Questionnaire - Tobacco Use Frequency
Driver Assessment	Sleep Questionnaire - Two Weeks Ago Typical
Driver Assessment	Sleep Questionnaire - Typical Week Alcohol Servings
Driver Assessment	Sleep Questionnaire - Typical Week Caffeine Intake Pattern
Driver Assessment	Sleep Questionnaire - Typical Week Caffeine Servings
Driver Assessment	Sleep Questionnaire - Use Sleep Aids Last Month
Driver Assessment	Sleep Questionnaire - Use Sleep Aids Typical Month
Driver Assessment	Sleep Questionnaire - Wake Time When Working From the Home
Driver Assessment	Sleep Questionnaire - Wake Time When Working Outside the Home
Driver Assessment	Sleep Questionnaire - Well Being While Awake
Driver Assessment	Sleep Questionnaire - Why Four Weeks Ago Not Typical

Dictionary	Name
Driver Assessment	Sleep Questionnaire - Why Last Week Not Typical
Driver Assessment	Sleep Questionnaire - Why Three Weeks Ago Not Typical
Driver Assessment	Sleep Questionnaire - Why Two Weeks Ago Not Typical
Driver Assessment	Sleep Questionnaire - Work Shifts Four Weeks Ago
Driver Assessment	Sleep Questionnaire - Work Shifts Last Week
Driver Assessment	Sleep Questionnaire - Work Shifts Three Weeks Ago
Driver Assessment	Sleep Questionnaire - Work Shifts Two Weeks Ago
Driver Assessment	Sleep Questionnaire - Work Status
Driver Assessment	Sleep Questionnaire -Doze While Watching TV
Driver Assessment	Trial 1 Response – Trial 360 Response
RID	Alignment Curve Direction
RID	Alignment Curve Length
RID	Alignment Curve PC Lat
RID	Alignment Curve PC Long
RID	Alignment Curve PT Lat
RID	Alignment Curve PT Long
RID	Alignment Curve Radius
RID	Alignment Tangent
RID	Cross-Slope
RID	Grade
RID	Intersection Control Type
RID	Intersection Location Lat
RID	Intersection Location Long
RID	Intersection Number of Approaches
RID	Lane Number
RID	Lane Type
RID	Lane Width
RID	Lighting
RID	Location Begin Elevation
RID	Location Begin Lat
RID	Location Begin Long
RID	Location End Elevation
RID	Location End Lat
RID	Location End Long
RID	Median Presence
RID	Median Type
RID	Rumble Strip Location
RID	Rumble Strip Presence
RID	Shoulder Type

Dictionary	Name
RID	Shoulder Width if Paved
RID	Signs Lat
RID	Signs Long
RID	Signs Message
RID	Signs MUTCD Code
RID	Signs Number of Signs on Post
RID	Super-Elevation
Time-Series	ABS Activation
Time-Series	Acceleration, x-axis
Time-Series	Acceleration, x-axis fast
Time-Series	Acceleration, y-axis
Time-Series	Acceleration, y-axis fast
Time-Series	Acceleration, z-axis
Time-Series	Acceleration, z-axis fast
Time-Series	Airbag, Driver
Time-Series	Alcohol
Time-Series	Cruise Control
Time-Series	Day
Time-Series	Dilution of Precision, Position
Time-Series	Distance
Time-Series	Driver Button Flag
Time-Series	Electronic Stability Control
Time-Series	Elevation, GPS
Time-Series	Engine RPM
Time-Series	Head Confidence
Time-Series	Head Position X
Time-Series	Head Position X Baseline
Time-Series	Head Position Y
Time-Series	Head Position Y Baseline
Time-Series	Head Position Z
Time-Series	Head Position Z Baseline
Time-Series	Head Rotation X
Time-Series	Head Rotation X Baseline
Time-Series	Head Rotation Y
Time-Series	Head Rotation Y Baseline
Time-Series	Head Rotation Z
Time-Series	Head Rotation Z Baseline
Time-Series	Heading, GPS
Time-Series	Headlight Setting

Dictionary	Name
Time-Series	Illuminance, Ambient
Time-Series	Lane Marking, Distance, Left
Time-Series	Lane Marking, Distance, Right
Time-Series	Lane Marking, Probability, Right
Time-Series	Lane Marking, Type, Left
Time-Series	Lane Marking, Type, Right
Time-Series	Lane Markings, Probability, Left
Time-Series	Lane Position Offset
Time-Series	Lane Width
Time-Series	Latitude
Time-Series	Longitude
Time-Series	Month
Time-Series	Number of Satellites
Time-Series	Pedal, Accelerator Position
Time-Series	Pedal, Brake
Time-Series	Pitch Rate, y-axis
Time-Series	Pitch Rate, y-axis fast
Time-Series	PRNDL
Time-Series	Radar, Range Rate Forward X Track n
Time-Series	Radar, Range Rate Forward Y Track n
Time-Series	Radar, Range, Forward X Track n
Time-Series	Radar, Range, Forward Y Track n
Time-Series	Radar, Target Identification
Time-Series	Roll Rate, x-axis
Time-Series	Roll Rate, x-axis fast
Time-Series	Seatbelt, Driver
Time-Series	Speed, GPS
Time-Series	Speed, Vehicle Network
Time-Series	Steering Wheel Position
Time-Series	Temperature, Interior
Time-Series	Time
Time-Series	Timestamp
Time-Series	Traction Control
Time-Series	Turn Signal
Time-Series	Video Dashboard and Steering Wheel View
Time-Series	Video Frame
Time-Series	Video, Driver and Left Side View
Time-Series	Video, Forward Roadway
Time-Series	Video, Occupancy Snapshot

Dictionary	Name
Time-Series	Video, Rear View
Time-Series	Wiper Setting
Time-Series	Yaw Rate, z-axis
Time-Series	Yaw Rate, z-axis fast
Time-Series	Year

Data Quality Variables

A total of 27 variables were classified as data quality variables. These variables, and their associated dictionaries, are provided in Table B-9.

Table B-9. Data quality variables.

Dictionary	Variable Name
Time-Series	Dilution of Precision, Position
Time-Series	Head Confidence
Time-Series	Heading, GPS
Time-Series	Lane Marking, Probability, Right
Time-Series	Lane Marking, Probability, Left
Time-Series	Number of Satellites
Time-Series	Radar, Target Identification
Time-Series	Timestamp
Time-Series	Video Frame
Crash Event	Participant ID
Crash Event	Complete
Crash Event	Date Last Action
Crash Event	Date Started
Crash Event	Crash Details - Interview Completion Date
Crash Event	Crash Details - Interview Completion
Crash Event Video	Event Start
Crash Event Video	Event End
Crash Event Video	Subject Number
RID	Alignment Curve PC Lat
RID	Alignment Curve PC Long
RID	Alignment Curve PT Lat
RID	Alignment Curve PT Long
RID	Alignment Curve Direction
RID	Signs Lat
RID	Signs Long
RID	Intersection Location Lat

Dictionary	Variable Name
RID	Intersection Location Long

Final Candidate Variables

This section lists the variables with a score greater than zero after prioritization. These variables were considered to be the most likely candidates for providing interesting results, and they constituted the starting point for finalizing the variable set for extraction during negotiations with VTTI in the data acquisition process.

The variable rating and utility rating scale in Table B-10 below provides the criteria used for to rate the variables in the prioritization activities.

Table B-10. Variable rating and utility rating scale.

Level	Definition
Costs	**3**: Involves significant effort or time in processing or acquiring **2**: Involves above-basic amount of effort/time in processing, yet less than required in Level 3 **1**: Involves a basic level of effort or time in processing or acquiring
Benefits	**3**: The variable is required to address at least one research question **2**: The variable provides additional information beneficial to the research question **1**: The variable does not provide information useful to a research question
Accuracy	**3**: High confidence in data accuracy **2**: Lower confidence in data accuracy, may require additional processing/validation **1**: Low confidence in data accuracy, will require additional processing/validation
Availability	**3**: Variable is available across all study vehicles; equipment generally reliable for variable **2**: Variable may not be available across all study vehicles; variable may not always be available because of equipment reliability issues **1**: Variable not available across all study vehicles
Utility	**3:** The variable is required for the analysis **2:** The variable is directly supporting or complementary to the analysis **1:** The variable is indirectly supporting the analysis

Ratings for each non-zero variable are provided in Table B-11. Dictionary provides the data set (data dictionary) the variable (listed under variable name) is described within. Research Question provides the research question topic(s) that the variable can address. Relationship provides the relationship between the variable and the research question topic. Crash, near-crash, speeding, and predictor are binary descriptors of the variable type; values of "1" indicate an association with that descriptor while zero values have been suppressed for ease of reading the table. The costs, benefits, accuracy, availability, priority, and utility ratings are also provided and are calculated as described in the body of the report.

Table B-11. Non-zero priority variables, by data source.

Dictionary	Variable Name	Research Question	Relationship	Crash	Near-Crash	Speeding	Predictor	Costs	Benefits	Accuracy	Availability	Priority	Utility
Time-Series	ABS Activation	Crash/Near-Crash	Indirect	1	1			1	3	3	1	**6**	1
Time-Series	Acceleration, x-axis	Crash/Near-Crash, Speed	Indirect	1	1			2	3	2	3	**6**	3
Time-Series	Acceleration, y-axis	Crash/Near-Crash	Indirect	1	1			2	3	2	3	**6**	3
Time-Series	Acceleration, z-axis	Crash/Near-Crash	Indirect	1	1			2	3	2	3	**6**	1
Time-Series	Airbag, Driver	Crash/Near-Crash	Direct	1				1	3	3	1	**6**	1
Time-Series	Alcohol	Crash/Near-Crash	Other			1	1	2	1	1	3	**3**	1
Time-Series	Cruise Control	Speed	Other			1	1	1	1	3	1	**4**	2
Time-Series	Day	Time/Light/ Season	Direct				1	1	3	3	3	**8**	3
Time-Series	Distance	Speed	Direct			1		1	2	3	1	**5**	2
Time-Series	Driver Button Flag	Crash/Near-Crash	Other	1	1			2	1	2	2	**3**	1
Time-Series	Electronic Stability Control	Crash/Near-Crash	Indirect	1	1			1	3	3	1	**6**	1
Time-Series	Elevation, GPS	Road Type	Other				1	2	2	2	3	**5**	1
Time-Series	Head Position X	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Head Position X Baseline	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Head Position Y	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Head Position Y Baseline	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Head Position Z	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Head Position Z Baseline	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Head Rotation X	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Head Rotation X Baseline	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Head Rotation Y	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Head Rotation Y Baseline	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Head Rotation Z	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Head Rotation Z Baseline	Crash/Near-Crash	Other				1	3	2	2	2	**3**	1
Time-Series	Headlight Setting	Time/Light/ Season	Indirect				1	1	2	3	3	**7**	2

Dictionary	Variable Name	Research Question	Relationship	Crash	Near-Crash	Speeding	Predictor	Costs	Benefits	Accuracy	Availability	Priority	Utility
Time-Series	Illuminance, Ambient	Time/Light/ Season	Direct				1	1	3	3	3	**8**	3
Time-Series	Lane Marking, Distance, Left	Crash/Near-Crash	Indirect	1	1			2	2	2	3	**5**	1
Time-Series	Lane Marking, Distance, Right	Crash/Near-Crash	Indirect	1	1			2	2	2	3	**5**	1
Time-Series	Lane Position Offset	Crash/Near-Crash	Indirect	1	1			2	2	2	3	**5**	1
Time-Series	Lane Width	Crash/Near-Crash	Indirect	1	1			2	2	2	3	**5**	1
Time-Series	Latitude	Road Type, Speed	Direct				1	1	3	2	3	**7**	3
Time-Series	Longitude	Road Type, Speed	Direct				1	1	3	2	3	**7**	3
Time-Series	Month	Time/Light/ Season	Direct				1	1	3	3	3	**8**	3
Time-Series	Pedal, Accelerator Position	Speed	Other			1		2	1	3	1	**3**	1
Time-Series	Pedal, Brake	Speed	Other			1		2	1	3	3	**5**	1
Time-Series	Pitch Rate, y-axis	Crash/Near-Crash, Speed	Indirect	1	1			2	2	2	3	**5**	2
Time-Series	Radar, Range Rate Forward X Track n	Speed	Direct	1	1	1		3	3	2	2	**4**	3
Time-Series	Radar, Range Rate Forward Y Track n	Speed	Direct	1	1	1		3	3	2	2	**4**	3
Time-Series	Radar, Range, Forward X Track n	Speed	Direct	1	1	1		3	3	2	2	**4**	3
Time-Series	Radar, Range, Forward Y Track n	Speed	Direct	1	1	1		3	3	2	2	**4**	3
Time-Series	Roll Rate, x-axis	Crash/Near-Crash	Indirect	1	1			2	2	2	3	**5**	2
Time-Series	Seatbelt, Driver	Demographics	Other				1	1	2	3	1	**5**	1
Time-Series	Speed, GPS	Speed	Direct			1		1	3	2	3	**7**	3
Time-Series	Speed, Vehicle Network	Speed	Direct			1		1	3	3	1	**6**	3
Time-Series	Steering Wheel Position	Road Type	Other				1	2	1	3	1	**3**	1
Time-Series	Time	Time/Light/ Season	Direct				1	1	3	3	3	**8**	3
Time-Series	Traction Control	Crash/Near-Crash	Indirect	1	1			1	3	3	1	**6**	1
Time-Series	Turn Signal	Demographics	Other				1	1	1	3	3	**6**	1
Time-Series	Wiper Setting	Time/Light/ Season	Other				1	1	2	3	1	**5**	2
Time-Series	Yaw Rate, z-axis	Crash/Near-Crash	Indirect	1	1			2	2	2	3	**5**	2
Time-Series	Year	Time/Light/ Season	Direct				1	1	3	3	3	**8**	3

Dictionary	Variable Name	Research Question	Relationship	Crash	Near-Crash	Speeding	Predictor	Costs	Benefits	Accuracy	Availability	Priority	Utility
Crash Event	Crash Details - Incident City	Road Type	Direct	1	1			1	3	3	3	**8**	2
Crash Event	Crash Details - Incident State	Road Type	Direct	1	1			1	3	3	3	**8**	2
Crash Event	Crash Details - Incident Date	Time/Light/ Season	Direct	1	1			1	3	3	3	**8**	2
Crash Event	Crash Details - Incident Time	Time/Light/ Season	Direct	1	1			1	3	3	3	**8**	2
Crash Event	Crash Details - Driver Seat Belt Use	Demographics	Other	1	1			1	2	2	2	**5**	2
Crash Event	Crash Details - Driver Age	Demographics	Direct	1	1			1	3	3	3	**8**	2
Crash Event	Crash Details - Driver Gender	Demographics	Direct	1	1		1	1	3	3	3	**8**	2
Crash Event	Crash Details - Number of Passengers	Demographics	Other	1	1		1	1	1	2	2	**4**	1
Crash Event	Crash Details - Travel Lane	Road Type	Indirect	1	1		1	1	2	3	2	**6**	1
Crash Event	Crash Details - Travel Lane - [Other]	Speed	Indirect	1	1		1	2	2	3	2	**5**	1
Crash Event	Crash Details - Travel Speed	Speed	Direct	1	1	1		1	3	2	2	**6**	2
Crash Event	Crash Details - Pre-Impact Intent -[Go straight]	Speed	Direct	1	1	1		1	2	2	2	**5**	1
Crash Event	Crash Details - Pre-Impact Intent - [Accelerate]	Speed	Direct	1	1	1		1	2	2	2	**5**	1
Crash Event	Crash Details - Pre-Impact Intent - [Change lanes to right]	Speed	Direct	1	1	1		1	2	2	2	**5**	1
Crash Event	Crash Details - Pre-Impact Intent - [Change lanes to left]	Speed	Direct	1	1	1		1	2	2	2	**5**	1
Crash Event	Crash Details - Pre-Impact Intent - [Merge]	Speed	Direct	1	1	1		1	2	2	2	**5**	1
Crash Event	Crash Details - Pre-Impact Intent - [Navigate curve]	Speed	Direct	1	1	1		1	2	2	2	**5**	1
Crash Event	Crash Details - Pre-Impact Intent - [Stopped, no movement intended]	Speed	Direct	1	1	1		1	2	2	2	**5**	1
Crash Event	Crash Details - Pre-Impact Intent - [Unknown]	Speed	Direct	1	1	1		1	2	2	2	**5**	1
Crash Event	Crash Details - Pre-Impact Intent - [Other]	Speed	Direct	1	1	1		2	2	2	2	**4**	1
Crash Event	Crash Details - Weather Condition	Time/Light/ Season	Direct	1	1		1	1	2	3	2	**6**	2

Dictionary	Variable Name	Research Question	Relationship	Crash	Near-Crash	Speeding	Predictor	Costs	Benefits	Accuracy	Availability	Priority	Utility
Crash Event	Crash Details - Weather Condition - [Other]	Time/Light/ Season	Direct	1	1		1	2	2	3	2	**5**	2
Crash Event	Crash Details - Weather Influence	Crash/Near-Crash	Direct	1	1		1	1	2	2	2	**5**	2
Crash Event	Crash Details - Weather Influence - [Other]	Crash/Near-Crash	Direct	1	1		1	2	2	2	2	**4**	2
Crash Event	Crash Details - Pre-Crash Vehicle Movement	Speed	Indirect	1	1			1	2	2	2	**5**	1
Crash Event	Crash Details - Pre-Crash Vehicle Movement - [Other]	Speed	Indirect	1	1			2	2	2	2	**4**	1
Crash Event	Crash Details - Activity Engaged In Prior to Crash	Speed	Indirect	1	1	1		1	3	2	2	**6**	2
Crash Event	Crash Details - Activity Engaged In Prior to Crash - [Other]	Speed	Indirect	1	1	1		2	3	2	2	**5**	2
Crash Event	Crash Details - Passenger Conversation	Speed	Indirect	1	1	1		1	3	2	2	**6**	2
Crash Event	Crash Details - Passenger Conversation - [Other]	Speed	Indirect	1	1	1		2	3	2	2	**5**	2
Crash Event	Crash Details - Driver Using Cell Phone Prior to Crash	Speed	Indirect	1	1	1		1	2	2	2	**5**	2
Crash Event	Crash Details - Driver Glance Location Prior to Crash	Speed	Indirect	1	1	1		1	2	2	2	**5**	2
Crash Event	Crash Details - Driver Glance Location Prior to Crash - [Other]	Speed	Indirect	1	1	1		2	2	2	2	**4**	2
Crash Event	Crash Details - Pre Crash Driver Activity	Speed	Indirect	1	1	1		1	3	2	2	**6**	2
Crash Event	Crash Details - Pre Crash Driver Activity - [Other]	Speed	Indirect	1	1	1		2	3	2	2	**5**	2
Crash Event	Crash Details - Driver Awareness	Speed	Indirect	1	1	1		1	2	2	2	**5**	2
Crash Event	Crash Details - Driver Awareness - [Other]	Speed	Indirect	1	1	1		2	2	2	2	**4**	2
Crash Event	Crash Details - Vehicle Speed Comparison	Speed	Indirect	1	1	1		1	2	2	2	**5**	1
Crash Event	Crash Details - Vehicle Speed Comparison - [Other]	Speed	Indirect	1	1	1		2	2	2	2	**4**	1
Crash Event	Crash Details - Trip Purpose	Demographics	Other	1	1		1	1	1	2	2	**4**	1

Dictionary	Variable Name	Research Question	Relationship	Crash	Near-Crash	Speeding	Predictor	Costs	Benefits	Accuracy	Availability	Priority	Utility
Crash Event	Crash Details - Trip Purpose - [Other]	Demographics	Other	1	1		1	2	1	2	2	**3**	1
Crash Event	Crash Details - Trip Urgency	Demographics	Indirect	1	1		1	1	1	2	2	**4**	2
Crash Event	Crash Details - Pre-Crash Miles Driven	Demographics	Other	1	1		1	1	1	2	2	**4**	1
Crash Event	Crash Details - Pre-Crash Miles Driven - [Other]	Demographics	Other	1	1		1	2	1	2	2	**3**	1
Crash Event	Crash Details - Driver's Urgency	Demographics	Other	1	1		1	1	1	2	2	**4**	2
Crash Event	Crash Details - Urgency Reason	Demographics	Other	1	1		1	1	1	2	2	**4**	1
Crash Event	Crash Details - Urgency Reason - [Other]	Demographics	Other	1	1		1	2	1	2	2	**3**	1
Crash Event	Crash Details - Roadway Familiarity 1	Demographics	Other	1	1		1	1	1	2	2	**4**	1
Crash Event	Crash Details - Roadway Familiarity 1 - [Other]	Demographics	Other	1	1		1	2	1	2	2	**3**	1
Crash Event	Crash Details - Total Years Driving Experience	Demographics	Other	1	1		1	1	1	2	2	**4**	1
Crash Event	Crash Details - Years Experience Driving Current Class of Vehicle	Demographics	Other	1	1		1	1	1	2	2	**4**	1
Crash Event	Crash Details - Times Driven Vehicle In Last 3 Months	Demographics	Other	1	1		1	1	1	2	2	**4**	1
Crash Event	Crash Details - Comfort Level With Vehicle	Demographics	Other	1	1		1	1	1	2	2	**4**	1
Crash Event	Crash Details - Windshield Condition	Time/Light/ Season	Indirect	1	1		1	1	3	2	2	**6**	2
Crash Event	Crash Details - Wiper Condition	Time/Light/ Season	Indirect	1	1		1	1	3	2	2	**6**	2
Crash Event	Crash Details - Wiper State	Time/Light/ Season	Indirect	1	1		1	1	3	2	2	**6**	2
Crash Event	Crash Details - Headlight State	Time/Light/ Season	Indirect	1	1		1	1	3	2	2	**6**	1
Crash Event	Crash Details - Crash Type	Crash/Near-Crash	Direct	1	1			1	3	3	2	**7**	2
Crash Event Video	Event Severity	Crash/Near-Crash	Direct	1	1			1	2	3	3	**7**	1
Crash Event Video	Event Nature	Speed	Direct	1	1			1	3	3	3	**8**	2
Crash Event Video	Incident Type	Speed	Direct	1	1			1	3	3	3	**8**	2

Dictionary	Variable Name	Research Question	Relationship	Crash	Near-Crash	Speeding	Predictor	Costs	Benefits	Accuracy	Availability	Priority	Utility
Crash Event Video	Pre-Incident Maneuver	Speed	Indirect	1	1			1	2	3	3	**7**	1
Crash Event Video	Maneuver Judgment	Speed	Indirect	1	1			1	1	2	3	**5**	1
Crash Event Video	Precipitating Event	Speed	Direct	1	1	1		1	3	2	3	**7**	3
Crash Event Video	Distraction 1	Speed	Direct	1	1	1		1	3	2	3	**7**	3
Crash Event Video	Distraction 2	Speed	Direct	1	1	1		1	3	2	3	**7**	3
Crash Event Video	Distraction 3	Speed	Direct	1	1	1		1	3	2	3	**7**	3
Crash Event Video	Surface Condition	Crash/Near-Crash	Direct	1	1		1	1	2	2	3	**6**	1
Crash Event Video	Traffic Density	Speed	Direct	1	1	1		1	2	3	3	**7**	3
Crash Event Video	Lighting	Time/Light/ Season	Direct	1	1		1	1	3	3	3	**8**	3
Crash Event Video	Weather	Time/Light/ Season	Direct	1	1		1	1	3	3	3	**8**	3
Crash Event Video	Driver Seatbelt Use	Demographics	Other	1	1		1	1	1	3	3	**6**	1
Driver Assessment	Demographic - Driver Gender	Demographics	Direct				1	1	3	3	2	**7**	3
Driver Assessment	Demographic - Driver Birth Date	Demographics	Direct				1	1	3	3	2	**7**	3
Driver Assessment	Demographic - Driver Ethnicity	Demographics	Indirect				1	1	2	3	2	**6**	1
Driver Assessment	Demographic - Driver Race	Demographics	Indirect				1	1	2	3	2	**6**	1
Driver Assessment	Demographic - Driver Country of Birth	Demographics	Indirect				1	1	2	3	2	**6**	1
Driver Assessment	Demographic - Driver Education Level	Demographics	Indirect				1	1	2	3	2	**6**	3
Driver Assessment	Demographic - Driver Marital Status	Demographics	Indirect				1	1	2	3	2	**6**	3
Driver Assessment	Demographic - Driver Family Income	Demographics	Indirect				1	1	2	3	2	**6**	3
Driver Assessment	Demographic - Driver Zip code	Demographics	Direct				1	1	3	3	2	**7**	3
Driver Assessment	Demographic - Driver Mileage Last year	Demographics	Indirect				1	1	2	3	2	**6**	3
Driver Assessment	Demographic - Driver Business Use of Study Vehicle	Demographics	Indirect				1	1	2	3	2	**6**	1
Driver Assessment	Demographic - Driver Length of Vehicle Ownership	Demographics	Indirect				1	1	2	3	2	**6**	1
Driver Assessment	Demographic - Driver Receive License	Demographics	Indirect				1	1	2	3	2	**6**	3

Dictionary	Variable Name	Research Question	Relationship	Crash	Near-Crash	Speeding	Predictor	Costs	Benefits	Accuracy	Availability	Priority	Utility
Driver Assessment	Driving History - Annual Mileage	Demographics	Indirect				1	1	3	3	2	**7**	3
Driver Assessment	Driving History - Years of Driving	Demographics	Indirect				1	1	3	3	2	**7**	3
Driver Assessment	Driving History - Training	Demographics	Indirect				1	1	2	3	2	**6**	1
Driver Assessment	Driving History - Moving Violations	Demographics	Indirect				1	1	2	3	2	**6**	1
Driver Assessment	Driving History - Moving Violation Type	Demographics	Indirect				1	1	2	3	2	**6**	1
Driver Assessment	Driving History - Number of Crashes	Demographics	Indirect				1	1	2	3	2	**6**	1
RID	Location Begin Lat	Road Type	Direct				1	1	3	3	2	**7**	3
RID	Location Begin Long	Road Type	Direct				1	1	3	3	2	**7**	3
RID	Location End Lat	Road Type	Direct				1	1	3	3	2	**7**	3
RID	Location End Long	Road Type	Direct				1	1	3	3	2	**7**	3
RID	Alignment Tangent	Road Type	Direct				1	1	3	3	2	**7**	3
RID	Alignment Curve Radius	Road Type	Direct				1	1	3	3	2	**7**	3
RID	Alignment Curve Length	Road Type	Direct				1	1	3	3	2	**7**	3
RID	Grade	Road Type	Direct				1	1	3	3	2	**7**	3
RID	Super-Elevation	Road Type	Direct				1	1	3	3	2	**7**	3
RID	Lane Number	Road Type	Direct				1	1	3	3	2	**7**	2
RID	Lane Width	Road Type	Direct				1	1	3	3	2	**7**	3
RID	Lane Type	Road Type	Direct				1	1	3	3	2	**7**	3
RID	Shoulder Type	Road Type	Direct				1	1	2	3	2	**6**	2
RID	Shoulder Width if Paved	Road Type	Direct				1	1	2	3	2	**6**	2
RID	Signs Message	Road Type	Other				1	1	2	3	2	**6**	3
RID	Signs MUTCD Code	Road Type	Other				1	1	2	3	2	**6**	3
RID	Signs Number of Signs on Post	Speed	Other				1	1	2	3	2	**6**	1
RID	Intersection Number of Approaches	Road Type	Direct				1	1	2	3	2	**6**	3
RID	Lighting	Time/Light/ Season	Direct				1	1	3	3	2	**7**	2
RID	Median Presence	Road Type, Speed	Direct				1	1	1	3	2	**5**	2
RID	Median Type	Road Type, Speed	Direct				1	1	1	3	2	**5**	2
RID	Rumble Strip Presence	Road Type	Other				1	1	1	3	2	**5**	2
RID	Rumble Strip Location	Road Type	Other				1	1	1	3	2	**5**	2

Appendix C. SHRP2 Variables Used in the Study

This appendix lists the SHRP2 variables requested from the SHRP2 data contractor and used in the study. Table C-1 includes the following fields:

- **Row Number:** Unique row number for referencing items in the table
- **Insight Field Name:** Variable name as listed in the Insight data dictionary.
- **VTTI Field Name:** Variable name as received from the data requests.
- **Data Table Type:** To protect sensitive driving data, the participant identification number was not included in the time series but was provided in a separate Dataset Key file. Trips were associated with drivers using the Trip ID variable. This column identifies which data table type included the variable.

Table C-1. SHRP2 variables used in the study

Row Number	Insight Field Name	VTTI Field Name	Data Table Type
1	Subject_ID	anonymousParticipantID	Dataset Key
2	Trip ID	displayTripId	Dataset Key, Time Series
3	—	system.time_stamp	Time Series
4	Timestamp	vtti.timestamp	Time Series
5	Trip ID	vtti.file_id	Time Series
6	LinkID	vtti.link_id	Time Series
7	Latitude	vtti.latitude	Time Series
8	Longitude	vtti.longitude	Time Series
9	Acceleration, x-axis	vtti.accel_x	Time Series
10	Acceleration, y-axis	vtti.accel_y	Time Series
11	Heading, GPS	vtti.heading_gps	Time Series
12	Illuminance, Ambient	vtti.light_level	Time Series
13	Lane Marking Probability, Left	vtti.left_marker_probability	Time Series
14	Lane Marking Probability, Right	vtti.right_marker_probability	Time Series
15	Lane Position Offset	vtti.lane_distance_off_center	Time Series
16	Pedal, Accelerator Position	vtti.pedal_gas_position	Time Series
17	Pedal, Brake	vtti.pedal_brake_state	Time Series
18	Speed, GPS	vtti.speed_gps	Time Series
19	Speed, Vehicle Network	vtti.speed_network	Time Series
20	Steering Wheel Position	vtti.steering_wheel_position	Time Series
21	Turn Signal	vtti.turn_signal	Time Series
22	Wiper Setting	vtti.wiper	Time Series
23	Radar, Range, Forward X Track 0	TRACK1_X_POS_PROCESSED	Time Series
24	Radar, Range, Forward X Track 1	TRACK2_X_POS_PROCESSED	Time Series
25	Radar, Range, Forward X Track 2	TRACK3_X_POS_PROCESSED	Time Series
26	Radar, Range, Forward Y Track 0	TRACK1_Y_POS_PROCESSED	Time Series
27	Radar, Range, Forward Y Track 1	TRACK2_Y_POS_PROCESSED	Time Series
28	Radar, Range, Forward Y Track 2	TRACK3_Y_POS_PROCESSED	Time Series

DOT HS 812 793
March 2020

U.S. Department
of Transportation

**National Highway
Traffic Safety
Administration**

14238-012120-v4a