Public Safety Policy and Research Center
Office of Financial Management

TRIP Data Handbook

TRAFFIC RECORDS INTEGRATION PROGRAM

About TRIP

The Washington State Traffic Records Integration Program (TRIP) is a data integration program housed within the Office of Financial Management's Public Safety and Research Policy Center and is funded by a grant through the Washington Traffic Safety Commission. TRIP works with various state agencies to collect and integrate data related to motor-vehicle crashes. The purpose of the TRIP program is to develop and maintain a data repository for public health and safety research to further the goals of the Washinton's Target Zero plan to achieve zero fatalities or serious injuries on Washington roadways.

Contact Us

Phone	**Fax**	**Address**	**Email**
360-902-0599	360-586-1988	1500 Jefferson Street SE, Suite 1145 Olympia, WA 98501	trip@ofm.wa.gov

Acknowledgements

This program would not be possible without the help and support of the Washington Traffic Safety Commission (WTSC) and our data contributors – the Washington State Department of Transportation (WSDOT), Washington Department of Licensing (DOL), Washington State Patrol (WSP), Washington Administrative Office of the Courts (AOC), and Washington Department of Health (DOH).

This document and the program and processes would not have been possible without the help and support of OFM's Education Research and Data Center (ERDC). Many of these policies and procedures have been directly adapted from ERDC efforts.

Version History

Date	Version	Author(s)	Revision Note
05/09/2025	1.0	Vasiliki Georgoulas-Sherry, PhD, Trevor Annis, MS, & Ian Kinder-Pyle, MS	Document creation
05/27/2026	2.0	Dr. Vasiliki Georgoulas-Sherry, Ian Kinder-Pyle, MS, and Trevor Annis, MS	Update WSDOT language, WSP breathalyzer data, Episode linkage

Table of Contents

Introduction

Purpose

This document describes the data available through the TRIP repository and the appropriate use of that data for different purposes. This Handbook provides a high-level description of each data source including details on the state agency that manages the data collection, substantive changes in data collection over time, an overview of the type of data available, any relevant policy or program changes, a description of how the data could be used, and impacts of external events (e.g., COVID-19 pandemic) on the data. This is done so that researchers, analysts, and data requesters can understand the data lineage and the breadth and depth of the information available through the TRIP repository data system.

While not a comprehensive look at any individual source data set, this Handbook is a body of knowledge that can help support an informed understanding of the data that TRIP has available and how it can be used for public safety and public health research.

For more information on program history including mission and vision, please review the TRIP Program Manual. For more information on data governance structure and processes, please review the TRIP Data Governance Manual.

TRIP Core Datasets

The TRIP receives a variety of administrative datasets from agency partners that are incorporated into the TRIP repository. These administrative datasets are outlined in Table 1, based on the category of data and data source. These datasets vary in subject matter, from the roadway to the crash, to police interaction, to court interaction, to health encounters. As such, TRIP's repository is the most comprehensive longitudinal crash-record public safety system in the state.

Table 1: The TRIP Core Data Sources

Sector	Data Contributor (Source) and Agency Description
Statewide Crash Data – Crash Location	Washington State Department of Transportation - Crash Data WSDOT maintains records of Washington state crash information including contributing factors and crash location
Driver License History and Ignition Interlock Device (IID)	Department of Licensing DOL maintains driver license history and IID data (i.e., instrument to measure breath alcohol content (BAC) level)
Court Case Filings – Judicial Information System	Administrative Office of the Courts AOC maintains statewide electronic court records database for all cases seen by courts in Washington state (excluding King County Superior courts as of 2019)
DUI-related Toxicology Results	Washington State Patrol - Toxicology Laboratory WSP maintains driving under the influence (DUI)-related toxicology results.
DUI-related Breathalyzer Test Results	Washington State Patrol – Breathalyzer Test Data WSP maintains driving under the influence (DUI)-related breathalyzer test results.
Comprehensive Hospital Abstract Reporting System	Department of Health - CHARS DOH maintains information on inpatient and observation patient community hospital stays in Washington state.
Rapid Health Information Network	Department of Health - RHINO DOH maintains real-time healthcare encounter data from all non-federal emergency department (ED) and critical access hospital facilities in Washington State
Washington Emergency Medical Services Information System (WEMSIS)	Department of Health - WEMSIS DOH maintains records associated with episodes for individuals accessing (pre-hospitalization) emergency medical services (EMS).
Death Vital Records	Department of Health - Vital Records DOH maintains data associated with death that took place in Washington state.
Trauma Registry Data	Department of Health - Trauma Registry DOH maintains the registry of patient information for individuals who sustain serious injuries and are treated in trauma designated Washington state hospitals, including individuals who were dead on arrival or transferred to another acute care facility.

TRIP Repository Administrative Data Limitations

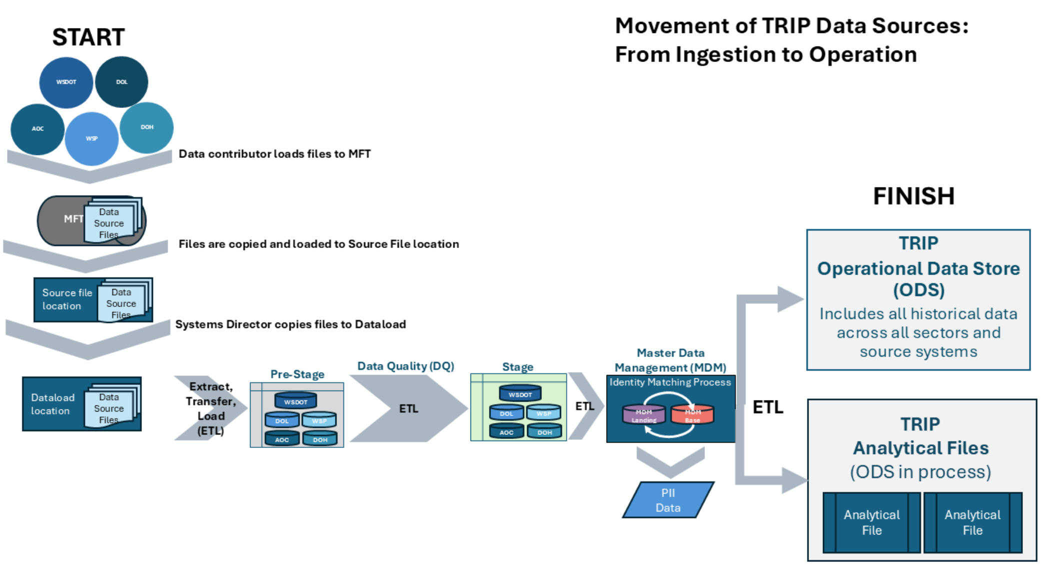
While all the datasets above are processed to the highest quality standards by the data contributing agencies, it is important to recognize that inaccuracies may exist within administrative data. Unlike other data, where both cross- and within-subject controls are possible, such measures are often unfeasible and impossible to incorporate in administrative data. Administrative data is also not typically collected for research or evaluation purposes but to meet the administrative needs of specific programs and specific state or federal reporting or monitoring requirements. Administrative data is collected as both transactional and summative datasets by local administrators and submitted to an agency authority, making variance among data collectors a potential source of bias in each dataset. Quality control processes may be imposed after data is submitted to agency authorities, which could impact data quality in ways that are difficult to detect within the final dataset.

The limitations described in this Handbook are not meant to suggest that the administrative data loaded into the TRIP repository is unreliable. TRIP advises researchers to keep these potential concerns in mind as they request data and conduct research. Administrative data must always be thought of as the combination of both the collected data and the process used to collect the data. The data summaries in this Handbook delve into these processes. Researchers who use TRIP data for analysis purposes should review all the available data documentation and adjust their models according to the research question and the administrative data collection procedures.

Flow of Contributor Data

Figure 1 illustrates how TRIP loads contributor data from contributing agencies. Once data is received through a secure file transfer process, the data is loaded to a pre-stage database, then it undergoes a series of quality checks before it is transferred to a stage database. Personally identifiable information (PII) is separated at that point from the rest of the data and used for identity resolution. Once the identity resolution process is complete, de-identified data are moved into the TRIP Operational Data Store (ODS) and become available for analysis. All PII data from the source files are excluded from the TRIP data repository and is unavailable to researchers. TRIP Analytical Files include selected de-identified data elements from the TRIP ODS that represent several public safety outcomes and measures.

Figure 1: Flowchart of data through stages of the TRIP loading process



Core TRIP Contributors and Dataset Overview

This section provides a set of descriptions or quick references to the core data files that feed into the TRIP repository. Please note, all data comes with its limitations. This report identifies a few major limitations, and there are likely more not listed that could impact work that utilize this data. This information is not an exhaustive list of data in the TRIP repository, nor does it provide the detail needed for a researcher to sufficiently complete a TRIP Data Request Form. Rather, these descriptions are designed to:

- guide researchers toward data that are relevant to their research questions
- provide metadata that will inform research design
- provide examples of how the data is used in research

This Handbook is designed to be used in conjunction with the TRIP Data Dictionary to understand the specific data elements and years of data available. See TRIP's Data Resources website.

Washington State Department of Transportation

The Washington State Department of Transportation (WSDOT) collects and maintains comprehensive crash data to improve road safety and transportation planning across the state. This data includes detailed records of traffic crashes, including factors such as location, time, weather conditions, road characteristics, and the involvement of pedestrians, cyclists, and motor vehicles. Law enforcement agencies submit crash reports, which WSDOT compiles into a centralized database. By analyzing this data, transportation officials, policymakers, and researchers can identify trends, high-risk areas, and contributing factors to accidents, ultimately informing strategies to reduce traffic fatalities and injuries.

One of the primary uses of WSDOT crash data is to enhance traffic safety by identifying patterns and implementing targeted interventions. For instance, crash data analysis helps in recognizing high-crash locations, commonly referred to as "hot spots," where accidents frequently occur. This information enables WSDOT to prioritize infrastructure improvements such as better signage, road redesigns, traffic signal adjustments, and pedestrian safety enhancements. Additionally, the data supports the development of traffic enforcement strategies, including targeted DUI checkpoints and speed control measures in high-risk areas.

The crash data is also crucial for evaluating the effectiveness of existing road safety initiatives and policies. By comparing crash trends before and after the implementation of safety measures—such as new traffic laws, roadway redesigns, or public awareness campaigns—WSDOT can assess whether these efforts are reducing accidents. This evidence-based approach ensures that state resources are directed toward initiatives that have a measurable impact on improving road safety. The data is used to support grant applications and funding requests for projects at both the state and federal levels.

Notes and Limitations of the WSDOT dataset

WSDOT fully codes, processes, and applies quality assurance checks to all "reportable" crashes which are identified by law (RCW 46.52.030) as crashes with at least $1,000 in damage and/or injury, involving a motor vehicle, and on a public roadway. Such crashes require law enforcement officer investigation and reporting. WSDOT also processes crash reports of interest to WSDOT's safety and design programs, which may fall outside the scope of the "reportable" definition. For example, WSDOT processes pedalcyclist-pedalcyclist crashes to provide safety data related to pedalcyclist and pedestrian road users. These crashes comprise the TRIP crash dataset.

However, many crash reports submitted to WSDOT are not processed by the crash processing branch. These non-reportable crashes include crashes occurring on private property and not adjacent to a public roadway, crashes occurring in parking lots, or crashes reported by private citizens to a call-in "Traffic Reporting Unit" but which was not investigated by law enforcement. Such crashes are not included in the TRIP data repository. For this reason, users may expect to see discrepancies when comparing crashes in the TRIP repository to other sources of crash information (e.g., DOH data for patients seeking care for crash-related injuries).

Tribes within the state are not required to report crashes on tribal lands but frequently do. Periodic changes to governance within the tribes (change in tribal police chief, or tribal council) or change in policy may result in changes to the reporting of the tribal crash data to WSDOT. This results in inconsistent reporting of crashes on tribal lands over time.

In many cases, identifying information for an individual involved in a crash is not captured on the Police Traffic Crash Report or the identifiers are insufficient to distinguish the individual from another with similar identifiers. In such cases, TRIP is unable to link the individual to other TRIP datasets, such as health or court interactions, and licensing history.

Washington State Department of Licensing

The Washington State Department of Licensing (DOL) collects and manages a vast amount of data related to driver licensing, vehicle registration, and professional certifications across the state. This data includes information on issued driver licenses and identification cards, vehicle ownership and registration records, driver record history (i.e. license suspensions and revocations, violations and/or convictions), driver license restrictions and/or endorsements, driver training and testing history, commercial driver certifications, and even professional and business licenses for various industries. DOL's database plays a crucial role in maintaining public safety, ensuring compliance with state laws, and facilitating efficient transportation and business operations. By managing and securing this data, the agency helps support law enforcement, insurance providers, and other government entities that rely on accurate licensing and registration records.

One of the most important uses of DOL data is ensuring road safety through driver monitoring and enforcement. The agency tracks driving records, including violations, suspensions, and revocations, helping law enforcement and the judicial system identify high-risk drivers. This data is also used to enforce penalties for DUI offenses, excessive speeding, reckless and negligent driving, and other infractions that could endanger public safety. Additionally, the DOL collaborates with the Washington Traffic Records Governance Council (TRGC) to improve data integration and analysis, ensuring that driver-related information is accurately shared across agencies for more effective safety measures. By maintaining and securing this vast collection of data, the Washington State Department of Licensing plays a critical role in promoting public safety, regulatory compliance, and efficient government operations.

Notes and Limitations of the DOL dataset

DOL generates the license history dataset from the WSDOT crash dataset by linking drivers' license numbers available on the PTCR to driver license numbers within the DOL system. This is sufficient to identify license history records for a majority of drivers in the crash dataset. However, DOL data is not provided in cases when the driver license number is missing or mistyped, or when the driver presents an out-of-state driver's license.

Washington State Administrative Office of the Courts

The Washington State Administrative Office of the Courts (AOC) plays a crucial role in managing and maintaining data that reflects the functioning and efficiency of the state's judicial system. This office is responsible for collecting, analyzing, and distributing a wide range of court-related data from across Washington's superior, district, and municipal courts. The data encompasses various types of cases, including civil, criminal, family law, juvenile justice, and appellate matters.

One of the primary uses of AOC data is to inform legislative and policy decisions. By analyzing the data on case filings, resolutions, and court processing times, policymakers can identify patterns and areas where the justice system may need reform or additional resources. For example, the AOC tracks the volume of cases in various categories, helping lawmakers gauge where interventions such as funding increases or policy adjustments are needed. Furthermore, data on case outcomes can guide decisions related to sentencing guidelines, bail reform, and other justice-related initiatives. The AOC's ability to provide detailed insights into these areas helps ensure that Washington's justice system remains efficient and responsive to the public's needs.

Notes and Limitations of the AOC dataset

The AOC dataset typically contains records from Washington State's court system, including superiors, districts, and municipal courts. It includes case filings, case types (criminal, civil, family, juvenile, etc.), dispositions, sentencing outcomes, and sometimes financial/fee data. Data is drawn from case management systems used across multiple courts. For AOC data, Washington state has a non-unified court system which means that Washington's courts operate more independently at the county and municipal level instead of being managed as one single, centrally administered system. As such, all courts are not all managed under a single statewide system – therefore, local governments (counties and cities) are primarily responsible for funding their trial courts, not the state.

Data are entered at the local court level, which can cause variation across counties in coding, timeliness, or completeness. For example, there could be variations across jurisdictions in how clerks enter data (e.g., charge codes, disposition reasons). There is also variability in local court practices – for example, municipal courts may process certain misdemeanors differently than district courts, and fines, fees, and restitution practices differ across jurisdictions and may be recorded unevenly. Juvenile, dependency, mental health, and other confidential case types are often redacted or entirely missing; and expunged/sealed records are removed, which can bias analyses of outcomes over time. Additionally, as this is administrative data with limited comments, the data does not take into account local practices (e.g., plea agreements, deferred prosecutions) which may not always map cleanly to standardized codes.

The AOC data focuses on case processing, and therefore there is no underlying circumstances (e.g., no narrative context, no details on the incident itself). What gets recorded reflects court processes, not necessarily underlying behavior or criminal activity. For example, arrests that do not result in charges are not present in this dataset; prosecutor discretion heavily shapes what enters the dataset. There is potential for charge and disposition variability which includes, for example, offense codes may change due to statutory updates, local ordinance differences, or reclassification over time. Dispositions (e.g., dismissal, plea, trial outcome) may not be consistently categorized. More so, only filed charges are included — not those screened out by prosecutors or resolved outside court.

Court data often have a lag before they are validated and uploaded to AOC systems. Furthermore, due to changes in policy over time, older records may lack detail or be inconsistent compared to more recent data – for example, charge codes and case types may be recoded or redefined over time. Additionally, there is inconsistent use of local offense codes vs. state statute codes - the exact coding practices can vary between jurisdictions or evolve over time - this data covers cases filed and processed as early as 2009, though availability of certain fields may differ by year.

There is also potential that there is missing or miscoded fields (e.g., race, ethnicity, sentence length), and certain offenses (e.g., traffic, municipal infractions) may not be uniformly reported. Furthermore, there is limited defendant characteristics (race, ethnicity, age, gender often inconsistently recorded) – this means that there is potential bias in data collection since demographic fields are often observer-coded, implicit bias from clerks or officers can affect race/ethnicity reporting. Additionally, in all criminal justice datasets, certain groups may be systematically underrepresented or misclassified – for example, race and ethnicity are often missing, inconsistently recorded, or based on officer/clerk perception rather than self-identification and gender is usually binary (male/female), which excludes nonbinary or transgender identities. The dataset provides no information on circumstances of the offense (e.g., context of a domestic violence charge, victim-offender relationship).

Washington State Patrol

The Washington State Patrol (WSP) Toxicology Laboratory is responsible for analyzing biological samples related to impaired driving, drug-related crimes, and death investigations. This laboratory plays a critical role in the criminal justice system by testing blood and other biological samples for the presence of alcohol, drugs, and toxic substances. The data collected from these analyses helps understand suspected impairment in drivers, confirm the presence of controlled substances in criminal cases, and provide crucial evidence in court proceedings. The WSP Toxicology Laboratory's work is essential for ensuring public safety and enforcing Washington State's impaired driving laws.

One of the primary uses of WSP toxicology data is in Driving Under the Influence (DUI) investigations. When a driver is suspected of impairment, law enforcement officers collect blood samples that are sent to the WSP Toxicology Laboratory for analysis. The results determine whether alcohol, cannabis, prescription medications, or illicit drugs were in the driver's system at the time of the incident. This data is crucial in DUI cases and shaping policies related to drug-impaired driving. The increasing prevalence of poly-drug use, where multiple substances are found in a driver's system, has made toxicology data even more valuable in understanding and addressing impaired driving trends.

The toxicology data also plays a key role in drug-related fatalities and forensic investigations. When a person dies under suspicious circumstances, medical examiners and coroners rely on toxicology reports to determine whether drugs, alcohol, or other toxic substances contributed to the death. These findings are used in homicide investigations, accidental overdose cases, and suicides. By analyzing trends in drug-related deaths, public health officials and policymakers can develop targeted interventions, such as improved addiction treatment programs and educational campaigns on substance abuse.

Additionally, the Washington State Patrol (WSP) collects and manages breathalyzer test data as part of its efforts to enforce impaired driving laws and promote roadway safety. Breathalyzer devices are used by law enforcement officers to measure a driver's blood alcohol concentration (BAC) through breath samples during traffic stops and DUI investigations. This data plays a vital role in identifying alcohol-impaired drivers, supporting legal proceedings, and monitoring trends in impaired driving across Washington State. The WSP's use of breathalyzer testing is essential for protecting the public and ensuring compliance with state laws regarding alcohol consumption and driving.

One of the primary uses of WSP breathalyzer data is in DUI enforcement. When a driver is suspected of alcohol impairment, officers administer a breath test to determine their BAC level. If the BAC exceeds the legal limit, the results can be used as evidence in DUI cases. This data helps law enforcement agencies track the frequency of alcohol-impaired driving incidents and evaluate the effectiveness of enforcement strategies, such as patrol emphasis and sobriety checkpoints. Breathalyzer data contributes to policy decisions aimed at reducing alcohol-related crashes and improving traffic safety.

Breathalyzer data also plays an important role in broader traffic safety analysis and public health research. By examining patterns in BAC levels among drivers, officials can identify trends such as peak times for alcohol-impaired driving, demographic risk factors, and geographic areas with higher incidence rates. This information supports the development of targeted prevention efforts, including public awareness campaigns and community-based interventions. Overall, WSP breathalyzer test data provides valuable insight into alcohol-related driving behavior and serves as a key tool in reducing impaired driving and enhancing public safety.

Notes and Limitations of the WSP data

There are several limitations regarding WSP toxicology data that will be described here. The presence and/or quantity of drugs in each blood test or sample is not necessarily an indicator of impairment. Potentially impairing drugs include all psychoactive substances. The mechanism by which drugs affect an individual, the extent to which they impair, and the time course for the impairment can differ. The Laboratory's scope of testing is routinely updated, and testing may vary based on testing services requested. In 2023, WSP implemented limited testing scopes for DUI case submissions when an ethanol result of 0.09 g/100mL or greater is detected, no subsequent drug testing is conducted. When ethanol results are below 0.09 g/100mL, then WSP tests for cannabinoids. If THC results are less than 7.0 ng/mL and more comprehensive drug testing is requested, then WSP will conduct a broader test for additional drugs. Additional information and current scope of testing can be found here.

In 2026, Washington's Senate Bill 5880 allows accredited private labs to conduct DUI toxicology testing alongside state facilities to reduce delays of up to 300 days, with implementation reporting required by October 1, 2031. This change may impact data completeness, as the Washington State Patrol may no longer receive all toxicology results when private labs are used, potentially affecting the overall scope and consistency of statewide toxicology data.

Washington State Department of Health

The Washington State Department of Health (DOH) collects and manages a vast amount of public health data to monitor the well-being of residents, track disease outbreaks, and guide policy decisions. This data encompasses a wide range of health-related areas, including vital statistics (birth and death records), disease surveillance, immunization records, healthcare facility licensing, and environmental health factors. By compiling and analyzing this information, the DOH plays a critical role in identifying public health trends, responding to emergencies, and improving healthcare outcomes across the state.

One of the most essential uses of DOH data is in tracking and managing communicable diseases. The agency collects real-time information on infectious diseases such as influenza, COVID-19, tuberculosis, and sexually transmitted infections. Healthcare providers and laboratories are required to report cases of certain diseases to the DOH, allowing for quick responses to outbreaks. This data enables the department to coordinate with local health agencies, issue public health alerts, and implement disease control measures such as vaccination campaigns and quarantine guidelines.

The DOH also oversees vital statistics, including birth and death records, which are crucial for public health planning and research. This information is essential for shaping policies related to healthcare access, preventive care, and emergency response planning.

Table 2: The DOH Core Data Sources

Sector	Dataset	Description
DOH	CHARS	Comprehensive Hospital Abstract Reporting System; collection of record level information on inpatient and observation patient community hospital stays.
	DEATH	Death rates and counts for leading causes of death in Washington state.
	RHINO	Rapid Health Information Network; automated, electronic reporting from all Washington state emergency departments with a focus on identifying, investigating and designing rapid responses to public health threats.
	TRAUMA	Trauma Registry; demographics, injuries, care, and outcomes of trauma patients.
	WEMSIS	Washington Emergency Medical Services Information System; emergency medical services/prehospital data repository for electronic patient care records.

Notes and Limitations of the DOH data

DOH TRIP data comes from 5 different resources (i.e., CHARS, WEMSIS, RHINO, Death, and TRAUMA). TRIP receives PII data from these data sets but does not house additional information such as injury data. Researchers can work with TRIP and DOH to link this data to gather additional insights.

For Death, CHARS, and Trauma data, the earliest data is from 2009. For RHINO, DOH filters records that are associated with traffic injury related incidents/conditions. They provide records that meet the "All Traffic related V2" ESSENCE syndrome definition. Automated, electronic reporting of syndromic surveillance data from all Washington emergency departments is required by RCW 43.70.057. For WEMSIS, RCW 70.168.090 requires all licensed ambulance and aid services in Washington to report response data to WEMSIS This requirement was implemented beginning in 2024. Recent years contain around 98% of EMS responses and earlier years having considerably less. The earliest WEMSIS data is from January 1, 2016 but coverage was limited across the state in 2016 and 2017.

Levels of Linkage in the TRIP Repository

The TRIP repository relies on two complementary forms of linkage to support cross-sector transportation safety research: person-level linkage and episode-level linkage. Together, these processes allow researchers to examine both the longitudinal histories of individuals across systems and the specific outcomes associated with individual crash events.

The first level, person-level linkage, is accomplished through the identity resolution process described in the next section. This process identifies records belonging to the same individual across contributing datasets and assigns a unique PersonID. Person-level linkage enables researchers to follow individuals across multiple sectors and over time, connecting records from crashes, licensing, court, emergency medical services, hospital, trauma, toxicology, and mortality datasets into a unified longitudinal record.

While person-level linkage establishes that records belong to the same individual, it does not determine whether specific records are related to the same crash event. For example, an individual may experience multiple crashes, emergency department visits, hospital admissions, or court interactions within a short time period. As a result, an additional layer of linkage is required to associate records with the appropriate crash incident.

The second level, episode-level linkage, addresses this need by identifying which records across datasets are likely associated with a particular crash event. Episode linkage uses combinations of PersonID, event dates and times, location information, injury indicators, case numbers, and other contextual variables to associate crash records with related outcomes and interactions. Because available identifiers and contextual information differ across datasets, the linkage logic is tailored to each data source.

The following sections describe (1) the person-level linkage methodology, and its identity resolution processes and (2) the episode-level linkage methodologies, along with important limitations and considerations relevant to interpretation and analysis.

TRIP Person-Level Linkage

The core feature of the TRIP repository is the linking of cross-sector data. Through an identity resolution process, TRIP links individuals across data files from contributing agencies to facilitate longitudinal and cross-sector analysis. Identity resolution is the process of identifying records that belong to the same entity (e.g. person or household). The purpose of TRIP's identity resolution process is to identify and create linkages across multiple data sources so that crash records associated with a given individual are linked to related records associated with that individual and event in the roadway, police interaction, court interaction, and health encounters datasets. For the TRIP repository, this involves linking individual-level data, such as names and birth dates, across multiple sources and identifying these individuals with unique person identifiers. These identifiers are referred to as "PersonIDs." PersonIDs are assigned to all individual-level data received by TRIP from our data contributors. As additional linking activities occur, PersonIDs are updated to reflect the most recent and complete data available.

Creation of PKeys and Assignment of PersonIDs

Before individual-level PersonIDs can be created or assigned, identity resolution tokens referred to as a "PKeys" are created for each record in a dataset. A "PKey" is an identifier or combination of identifiers from a dataset that are unique to an individual within the dataset. TRIP reviews each dataset to establish a PKey definition specific to the dataset. Identifier and individual characteristic data from the datasets and associated with the PKeys are then loaded into the identity resolution system for linkage. PKeys that already exist in the identity resolution system are attributed with the PersonID assigned to that PKey. PKeys that do not already exist in the identity resolution system are assigned a unique preliminary PersonID which may be overwritten in the identity resolution steps described below.

Because no common set of identifiers exists across all data sources, identity resolution processes and match rules are tailored to each dataset and limited to the identifiers available for linkage (see Table 3).

Table 3: Identity Components in Data Sources

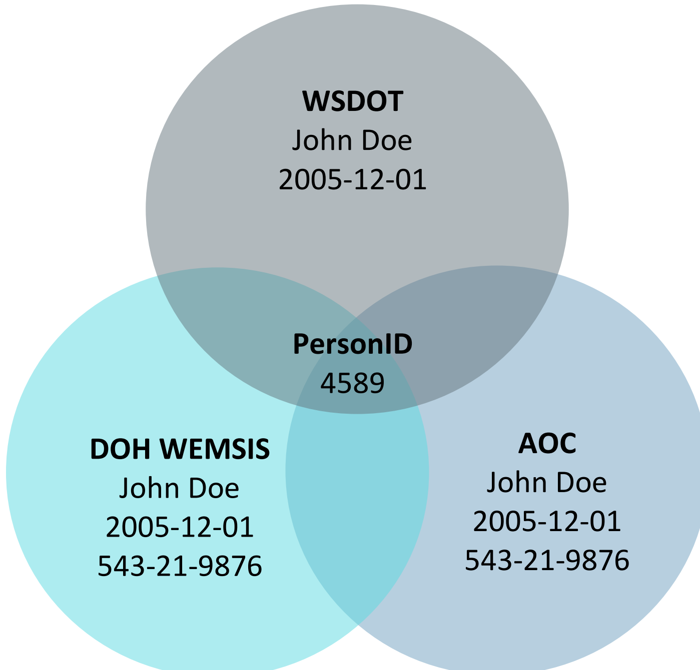
Sector	Dataset	Birthdate	First Name	Middle Name	Last Name	SSN	License #
WSDOT	Crash	✓	✓	✓	✓		✓
DOL	Licensing						✓
WSP	Toxicology	✓	✓	✓	✓		
	Breathalyzer	✓	✓	✓	✓		✓
AOC	Court Filings	✓	✓	✓	✓		
DOH	CHARS	✓	✓	✓	✓	✓	
	DEATH	✓	✓	✓	✓	✓	
	RHINO	✓	✓	✓	✓		
	TRAUMA	✓	✓	✓	✓	✓	
	WEMSIS	✓	✓	✓	✓	✓	

No single set of identifiers is common to all data sources, so the identity resolution process and match rules are tailored to the source of data being matched. For example, DOH data has names, birth date, and Social Security Number (SSN) whereas AOC data has names and birth date. As a result, rules linking each are limited to the available identifiers.

Two datasets with different sets of available identifiers can be indirectly matched by involving other sources of individual data. For example, the Licensing dataset above can only be directly linked to the

Crash data, but the Crash data can be linked to all other datasets. Consequently, the licensing data may be linked indirectly to all datasets in the TRIP repository (see Figure 2).

Figure 2: Identity Token Components in Data Sources



Phases of Identity Resolution

TRIP's identity resolution process has four phases which are performed sequentially each time a new dataset or dataset update is linked.

Blocking

Blocking is the process of generating potential match pairs between new PersonIDs and existing PersonIDs within the identity resolution system using predefined match rules. These rules compare combinations of identifiers— such as names, birth dates, and SSNs—to identify records that may belong to the same individual. Match quality depends on the strength of the rule applied: exact matches across multiple identifiers generally produce low false positive rates, while rules based on fewer fields, partial strings, or fuzzy matching techniques increase the likelihood of uncertainty. After potential match pairs are generated, the identity resolution system applies the Expectation Maximization (EM) algorithm to calculate a probabilistic match score for each pair.

Evaluation

The set of potential match pairs is split into three categories:

- **High-probability matches.** These PersonID match pairs are the result of application of conservative match rules (e.g., "same name, same birth date, same SSN"). For this set, undermatching (or not correctly identifying an actual match) is not a significant concern, as the conservative match rules are designed to ensure extremely low false positive rates. Probabilistic match scores might also be used to further delineate the set of high probability matches.
- **Mid-probability matches.** These potential match pairs result from the application of looser match rules than the ones used to create the high probability matches (e.g., "same name, same

birth date"). The match pairs and probabilistic scores, and associated identifiers (names, dates of birth, etc.) are brought into one dataset. The match pairs are then manually reviewed. The match pairs that are deemed to be actual positive match pairs are flagged, and the results are integrated into the prospective match pairs within the identity resolution system. At this point, the identity resolution system contains the set of provisional match pairs.

- **Low-probability matches.** These potential match pairs result from very loose match rules (e.g., "same county of residence, same gender, same first name"). Based on low probabilistic match scores, none or very few of these potential match pairs are provisional match pairs, hence these potential match pairs are ignored.

Cardinality Analysis

Cardinality analysis is a key step in identity resolution that enables more aggressive matching while improving the quality of existing PersonID linkages. In this process, provisional match pairs are temporarily merged and evaluated to determine the relationships between PersonIDs in the subject dataset and those in the broader repository. These relationships may be 1:1, 1:Many, Many:1, or Many: Many – for example, a 1:Many relationship indicates one PersonID in the subject dataset matches multiple PersonIDs in the broader repository. 1:1 relationships are accepted, while non-1:1 relationships are manually reviewed to resolve discrepancies as accurately as possible. This step also allows analysts to reassess prior matches using new information, potentially merging or unmerging records from earlier iterations. As a result, identity resolution linkages continuously improve over time as new information is obtained. Once verified, results are fed back into the identity resolution system, where PersonIDs may be merged, unmerged, or provisional matches rejected.

Merging

After the cardinality phase is concluded, the match table now contains a list of positive match pairs of PersonIDs. These match pairs are then incorporated into the identity resolution system using an automated process. The result is that people who had been previously represented by multiple preliminary PersonIDs are now represented by a single PersonID.

The result of the person-level linkage process is that all records associated with a distinct individual are assigned the same, unique PersonID across all tables in the repository, allowing researchers to identify and compile records associated with an individual.

TRIP Episode-Level Linkage

The TRIP repository is an event-based dataset. Through the identity resolution process described in the previous section, individuals are identified and linked both within and across datasets. However, to support meaningful research, the crash and outcome records associated with a specific individual and crash event must also be connected. This process, referred to as “episode-level linkage,” is described in the following sections.

Linkage of all contextual data to the associated crash records relies upon the availability of person-level identity resolution – that is, unique individuals have been identified and linked within and across datasets. In general, episode-level linkage is performed by linking records associated with a given person and date. However, this approach is complicated by the fact that multiple crashes may occur on a given date, and that multiple outcome records (e.g., EMS incidents, toxicology cases, hospital admits, charges) may also correspond to that same date.

Some datasets include additional data elements or dataset definitions that might be used to inform or refine the linkage results. For example, the toxicology dataset includes the requesting agency case number which can also be listed in the crash record. Furthermore, the DOH’s RHINO dataset is limited by the source program to include only incidents associated with motor vehicle crashes.

Linkage Logic by Dataset

AOC Data

Initially, motor vehicle drivers, pedalcyclists, and pedestrians are associated with charge records related to motor vehicle operation or roadway usage.[1] This linkage is based on the PersonID and a charge violation date occurring within one day of the coded crash date, allowing for potential typographical errors in date entry.

Charge records that are preliminarily linked with multiple crashes are subsequently refined using date, court county (under the assumption that charges are adjudicated in the county where the corresponding violation(s) occur), and indication of citation in the PTCR (under the assumption that charges related to a crash are generally preceded by a citation from the investigating officer). These refining rules are only applied when a single charge is associated with multiple crashes.

Figure 3 illustrates an example of a single crash linked to multiple court cases. In this example, a single crash has been preliminarily linked to two cases. The case violation date for both matches the crash date. A single incident yielding multiple court cases is not unusual and the crash record is linked to both court case records.

Figure 3. Single Crash, Linking to Multiple Court Cases

Crash Data			
Crash ID	**Date**	**County**	**Cited?**
10145	5/16/2019	Clallam	Y

Court Data		
Case ID	**Court County**	**Violation Date**
CC-0918	Clallam	5/16/2019
CM-8710	Clallam	5/16/2019

Figure 4 shows the refinement of linkages using similarities across additional data field. In this example, two court cases are each preliminarily linked to two crashes involving the same individual on adjacent

[1] It is assumed that charges associated with passengers are generally not relevant to the crash cause or result.

dates. Each court case linkage is refined independently. The first court case (SCC-09980) shares a county and date with crash 44154. Conversely, it is on a different date and in a different county than crash 46632 and the individual did not receive a citation in the second crash. Therefore, SCC-09980 is linked to the first crash, but not the second.

Likewise, the second court case (SCC-10041) occurs in a different county than either crash, but matches the date of the first, which also shows a citation. Case SCC-10041 is also linked to the first crash but not the second.

Figure 4. Refining Linkages Based on Similarities in Additional Fields

Crash Data			
Crash ID	**Date**	**County**	**Cited?**
44154	1/27/2023	Spokane	Y
46632	1/28/2023	Klickitat	N

Court Data		
Case ID	**Court County**	**Violation Date**
SCC-09980	Spokane	1/27/2023
SCC-10041	Stevens	1/27/2023

Figure 5 illustrates situations in which linkage refinement is not possible. Here, a single court case is preliminarily linked to two crashes for an individual occurring on the same day. Since the refining variables (crash date, county, citation indicator) are the same between the two crashes, there is no way to distinguish which crash(s) might be the cause of the charges, and which might not. The case remains linked to both crashes.

Figure 5. Linkage Where Refining Associations is Not Possible

Crash Data			
Crash ID	**Date**	**County**	**Cited?**
00341	2/1/2023	Whatcom	Y
01590	2/1/2023	Whatcom	Y

Court Data		
Case ID	**Court County**	**Violation Date**
C63301	Whatcom	2/1/2023

AOC Dataset-Specific Limitations & Notes

The AOC dataset contains records for charges that proceed into the court system. However, not all citations result in filed charges, and not all charges are ultimately recorded with AOC. As a result, this linkage does not provide a complete representation of unlawful activity associated with crashes.

Court cases generally include all charges associated with a single investigation. In the context of traffic violations related to crashes, a case may contain multiple distinct charges and, in some instances, charges associated with multiple crashes arising from a broader incident. For example, a single case may include charges related to several crashes occurring during a hit-and-run sequence. When those crashes occur on the same date, it may not be possible to reliably determine which specific charges correspond to which crashes.

The linkage refinement addresses cases where a single charge is preliminarily associated with multiple crashes for an individual. However, it is not uncommon to see multiple court cases related to a single crash incident. For example, civil infractions might be charged at the municipal court and felonies at the superior court. Furthermore, charges might be brought at the superior court, but if all felonies are dismissed the other misdemeanors might be remanded to a lower court for disposition.

Since the AOC dataset contains all charges, it must first be filtered to traffic-related charges. To the degree that the initial traffic charge filter captures an incomplete list of traffic-related charges, the final linkage may include inappropriate linkages to non-collision-related charges (e.g., charges related to a traffic stop earlier in the day) or exclude appropriate linkages to traffic-related charges.

DOH CHARS Data

The CHARS dataset is linked to crash records based on PersonID and an admission year that matches either the crash year or the year immediately following the crash. This broad linkage is necessitated by the fact that the CHARS dataset TRIP currently obtains from DOH has no explicit data elements that can be used to link admissions to specific incidents or to each other, outside of the admission year which is derived from the record identifier.

Figure 6 shows the crash linkage to CHARS admission/visit records. In this example, two crashes are each linked with three admission records in the CHARS datasets. Without additional information, it is not possible to tell which hospital records are associated with which crash(es).

Figure 6. Crash Linkage to CHARS Admission/Visit Records

Crash Data	
Crash ID	**Crash Date**
63121	8/8/2022
94438	12/18/2022

CHARS Data	
Record ID	**Derived Admission Year**
2022_0171390	2022
2022_0298816	2022
2023_0045112	2023

DOH CHARS Dataset-Specific Limitations & Notes

Identifying whether a particular hospital visit can be associated with a specific crash is difficult to determine programmatically. Medical conditions may result from one or more incidents, persist long after the originating event, or require treatment well after the initial injury occurred. Individuals may also be involved in multiple incidents involving similar or different injuries that require inpatient admission or outpatient observation, and a single admission may address multiple unrelated conditions. Together, these factors complicate the determination of which CHARS records should be considered outcomes of a given crash.

The preliminary linkage approach described above almost certainly results in some crashes being associated with admissions related to different crashes or to conditions unrelated to any crash. Conversely, it likely fails to identify admissions associated with crash-related injuries requiring long-term or delayed treatment. Because the available CHARS dataset does not include an admission date, the linkage process may also incorrectly associate crashes with admissions that occurred prior to the crash event. For these reasons, requestors may need to obtain and use additional CHARS data elements—such as admission date, reason for admission, symptom onset date, and related clinical information—to more accurately determine whether an admission is associated with a particular crash.

It is also reasonable to expect that a single crash may result in multiple admissions and that injuries from multiple crashes may be treated during a single admission. Accordingly, the relationship between crashes and admissions may legitimately be many-to-many, with one crash linked to multiple admissions and one admission linked to multiple crashes. However, the extent to which these relationships appear in the linkage output is likely driven as much by the data limitations described above as by true underlying clinical relationships.

In addition, the CHARS dataset captures only inpatient admissions and certain observation stays. As a result, collision-related injuries treated exclusively in emergency departments without admission, outpatient settings, or out-of-state facilities are not represented, leading to incomplete outcome capture. Variability in clinical coding practices across hospitals and across time may also contribute to inconsistent identification of collision-related conditions.

DOH Death Vital Records

While many crash participants will have associated death records, the episode linkage logic seeks to link crash incidents with *related* outcomes for the individual. For this reason, crash participants are linked to death records based on PersonID alone but only when the crash record indicates a traffic-related fatality. All other crash participants may be linked to death record data as needed based on PersonID.

Figure 7 shows the crash to death record linkage. In this example, an individual is involved in three crashes over multiple years. This individual also has a death record. The death record is linked only to the final crash, in which the injury type indicates the crash resulted in the individual's death.

Figure 7. Crash to Death Record Linkage

Crash Data			
Crash ID	**PersonID**	**Crash Date**	**Injury Type**
073725	561572	12/16/2012	Suspected Minor Injury
267441	561572	5/3/2018	No Apparent Injury
428112	561572	8/19/2022	Died in Hospital

Death Data	
PersonID	**Death Date**
561572	8/21/2022

DOH Death Vital Records Dataset-Specific Limitations & Notes

Federal reporting requirements for traffic fatalities, together with WTSC's fatal case review process, result in generally consistent and reliable fatality indicators within PTCR records when an individual dies as a result of a crash. In a small number of cases, however, the linked death record reflects a death date that precedes the crash date. This discrepancy may arise either from limitations in the identity resolution process that result in records from different individuals being incorrectly linked, or from inaccuracies in the recorded collision or death dates within the source systems. Review of these cases suggests that date-entry errors are substantially more likely than mismatched identities.

Not all crash participants coded as fatalities are successfully linked to a death record. One legitimate explanation is that an injured individual may have died while being transported to, or treated at, an out-of-state medical facility and therefore would not appear in the available death records. In most cases, however, the absence of a linked death record is more likely attributable to limitations in the identity resolution process that prevent successful linkage. In limited circumstances, separate collision reports may be submitted for multiple distinct collisions that together comprise a single event, with fatality coding applied to more than one crash record. For example, in one collision a driver may rear-end another vehicle without injury, then exit the vehicle and subsequently be struck by a third vehicle in a second collision, later dying at the hospital. In such cases, the investigating officer may code the individual's injury type as "Died at Hospital" in both crash reports. This results in multiple crashes being linked to a single death record. Unfortunately, crash data cannot always reliably establish the sequence of related collisions because the recorded crash times may be identical across reports.

Although it may often be reasonable to assume that the final collision involving an individual was the primary cause of death, this assumption does not always hold. For example, an individual may sustain fatal injuries in a serious collision and later be involved in a minor secondary incident while being transported for medical care, such as an ambulance being rear-ended without causing additional injuries. In this circumstance, the death should properly be attributed to the earlier collision rather than the later event. In both examples described above, the linkage logic associates all crashes in which the individual is coded as deceased with the corresponding death record.

Importantly, the existence of a linkage between a crash participant and a death record does not, by itself, establish that the death was caused by the crash. The Injury Type value recorded on the crash report remains the authoritative indicator of a crash-related fatality.

DOH RHINO Records

Initially, crashes are linked to emergency department (ED) visits occurring within seven days of the crash. In addition, subsequent visits will be linked to the same crash as long as they occur within three days of a prior visit already associated with that crash.

Associating visits with specific crashes is complicated when multiple crashes occur in a short period of time. Given the limitations of the available data in the two datasets, it is not possible to definitively determine which collision resulted in which specific emergency department visits. In general, of the preliminarily linked visits, all visits on or after a crash date will be associated with the crash up to (but not including) the date of a subsequent crash. This will result in all crashes for a day being associated with the same set of ED visits. Two special cases are described below.

Special Case 1: If multiple crashes occur on one day, all RHINO records on that day and after will be associated only with the crash(s) in which an injury was incurred, as indicated by the crash report, and none with the non-injury crash(s), for the individual. If all crashes are non-injury crashes for the individual, all RHINO records will be associated with all crashes (see Figure 9).

Special Case 2: If at least one crash occurs on adjacent days, ED visits that occur on or after the second day will be handled using the rules laid out in Special Case 1, adjusted to consider all crashes between the two days. Note that if at least one crash occurs on *more* than two consecutive days, each pair of consecutive days will be treated independently (see Figure 10).

Figure 8 describes an example of a driver involved in collisions on two non-consecutive days. Any visit on or after the first crash, but prior to the date of a subsequent crash are associated with the first crash. Any visits on or after the date of the second crash are associated with that crash, regardless of the presence of an injury.

Figure 8. Multiple Crashes Occur on Two Non-Consecutive Days

Day	Event - Injury	Visits on this day associated with
1	Crash 1 – Injury	Crash 1
2		Crash 1
3	Crash 2 – No injury	Crash 2
4+		Crash 2

Figure 9 describes a situation where Special Case 1 is employed. Any visit occurring on or after Day 3 is attributed to Crash 2 regardless of the injury severity indicated in the crash data. For Days 1 and 2, visits are associated only with the injury crash from Day 1 (special case 1). For visits on or after Day 3, all crashes are associated with Crash 3 regardless of the injury severity indicated in the crash data.

Figure 9. Multiple Crashes Occur on a Single Day, Followed by Another Crash Several Days Later

Day	Event - Injury	Visits on this day associated with
1	Crash 1 – Injury Crash 2 – No injury	Crash 1
2		Crash 1
3	Crash 3 – No injury	Crash 3
4+		Crash 3

Figure 10 describes a situation where Special Case 2 is employed. Visits on Day 1 are associated with Crash 1. Visits on Day 2 would be associated with both Crash 1 and Crash 3 (Special Case 2). No visits would be associated with Crash 2 because it is a non-injury crash among two consecutive days containing injury crashes.

The crashes occurring on Day 2 and 3 also comprise Special Case 2. Consequently, and because both injury and non-injury crashes exist on these days, visits on and after Day 3 are associated with all injury crashes between the two days. No visits on or after Day 3 are associated with Crash 1 because it is more than one day after the date of Crash 1.

Figure 10. Multiple Crashes Occur on Three Consecutive Days

Day	Event - Injury	Visits on this day associated with
1	Crash 1 – Injury	Crash 1
2	Crash 2 – No injury Crash 3 – Injury	Crashes 1 & 3
3	Crash 4 – No injury	Crash 3
4+		Crash 3

DOH RHINO Records Dataset-Specific Limitations & Notes

RHINO data only include visits from participating ED facilities, meaning that ED visits to non-participating hospitals,[2] clinics, or out-of-state facilities are not captured and thus not linked. Additionally, visits may be included in the dataset based on mentions of a crash in chart notes even when the visit is unrelated to a crash, which can inflate apparent delays or follow-up counts. Delayed emergency department visits or follow-up visits for long-evolving injuries may fall outside the standard linkage window.

The RHINO dataset provided for the linkage includes visit dates but not times and lacks detailed diagnoses or outcome codes, limiting the ability to accurately sequence multiple visits or confirm collision-related care. Individuals involved in multiple collisions within a short period may have visits that cannot be definitively attributed to a single collision, resulting in potential conflation of injuries. Finally, variation in patient behavior – such as seeking care in outpatient settings, not seeking care at all, or seeking medical assessments even when uninjured, results in relationships between collisions and ED visits that cannot be fully resolved given the available data.

DOH Trauma Registry

Crash records are initially linked to trauma admissions using PersonID, including all admissions occurring within three days following a crash. This window allows for the inclusion of delayed admissions, as well as initial and transferred trauma admissions related to the crash event. If no admissions are identified on or within the three days following a crash, but at least one admission exists within up to two days prior to the crash date, the most recent of these prior admissions is linked to the crash. This approach accounts for potential errors in the recording of either crash or admission dates.

Preliminarily linkages that result in multiple crashes being associated with a single admission are refined by unlinking all crashes from a given admission that occur more than 30 minutes prior to the most recent crash associated with that admission. If multiple crashes remain linked to a single admission after

[2] Currently, the only non-participants are federally-owned facilities such as Madigan Army Hospital and the Veteran's Association (VA) Hospital.

this step, and at least one of those crashes is coded as having no injury based on officer assessment, all non-injury crashes are unlinked from the admission record.

In Figure 11, three trauma admissions are initially linked to a single crash record for an individual. Because admissions exist within three days following the crash, the admission prior to the crash date is unlinked.

Figure 11. Crash Potentially Linked to Trauma Admissions Before and After Crash Date

Crash data		
Crash ID	Crash Time	Injury Severity
451	7/7/2019 22:14	Suspected Serious Injury

Trauma data	
Admit ID	Admit Date
90367	7/5/2019
74111	7/7/2019
65012	7/8/2019

In Figure 12, three crashes are initially linked to a single trauma admission. The first crash is unlinked because it falls more than 30 minutes prior to the latest crash on or prior to the admission date. The second crash is removed because, while it is less than 30 minutes prior to the latest crash on or before the admission date, it is a non-injury crash whereas the last crash is an injury crash (see Figure 12).

Figure 12. Multiple Crashes Linked to a Single Admission

Crash data		
Crash ID	Crash Date	Injury Severity
4590	11/24/2023 09:14	Suspected Minor Injury
4607	11/25/2023 20:37	No Apparent Injury
5513	11/25/2023 20:43	Suspected Serious Injury

Trauma data	
Admit ID	Admit Date
55431	11/25/2023

In Figure 13, two crashes are each initially linked to the same two admits. Since at least one admission occurs on or after the date of the second crash (098312), the first admission (110581), which occurs prior to the crash date, is unlinked. The other admission (264751) is linked to two crashes. Because the first crash (094692) falls more than 30 minutes prior to the latest crash still associated, it is unlinked.

Figure 13. Multiple Crashes Linked to Multiple Admissions

Crash data		
Crash ID	Crash Date	Injury Severity
094692	5/1/2024 14:43	Suspected Serious Injury
098312	5/3/2024 08:31	*NULL*

Trauma data	
Admit ID	Admit Date
110581	5/1/2024
264751	5/3/2024

DOH Trauma Registry Dataset-Specific Limitations & Notes

This methodology produces cases where a single admission is associated with multiple crashes for an individual. This occurs when the crashes are coded as occurring at the same time – this typically reflects a single incident that involves multiple distinct crashes events. Without more information from the PTCR, it is impossible to reliably determine which of these crashes results in the trauma admission.

Likewise, the methodology produces cases where a single crash participant is associated with multiple trauma admissions. A valid explanation for this would be when a crash results in an individual being admitted to a trauma hospital but is subsequently transported and admitted to a second trauma hospital, perhaps one with a higher trauma level designation. Admissions related to a crash that occur more than three days after the crash will not be linked, and trauma admissions that fall within three days of a crash, but which are not, in fact, related to the crash may be erroneously linked to the crash.

DOH WEMSIS

Crashes are initially linked to WEMSIS records based on PersonID and the EMS incident falling within 24 hours after the crash. EMS incidents coded up to an hour prior to the crash time are also included to account for the rounding of crash times commonly found on crash reports.

Preliminary linkages that associate a single EMS record with multiple crashes are then refined. If an EMS record is preliminarily linked to multiple crashes, it will remain linked to all crashes within one hour of the closest crash (temporally), all other preliminary linkages associated with the EMS are removed. Of the crash records that remain linked, if at least one indicates an injury, any non-injury crashes are unlinked.

Figure 14 shows a single crash linked to multiple EMS records. Even though the second EMS incident is on the subsequent calendar day, both fall within 24 hours of the crash time and remain linked.

Figure 14. A Single Crash Linked to Multiple EMS Records

Crash Data		
Crash ID	**Crash Time**	**Injury Severity**
009172	5/11/2024 23:42	Suspected Minor Injuries

WEMSIS Data	
Incident ID	**Incident Time**
71172	5/11/2024 23:45
75871	5/12/2024 01:14

Figure 15 shows multiple crashes linked to a single EMS record. The first two occurred less than 24 hours before the EMS incident date, and the third less than one hour after. However, because the first crash is more than one hour prior to the latest linked crash (331726), it is unlinked. Furthermore, though the second crash is within one hour of the latest crash time, it is coded as a non-injury crash while the last is an injury crash. The second crash is therefore, also unlinked from the WEMSIS record.

Figure 15. Multiple Crashes Linked to a Single EMS Record

Crash Data		
Crash ID	**Crash Time**	**Injury Severity**
236190	2/23/2019 20:17	*NULL*
300682	2/24/2019 08:22	No Apparent Injuries
331726	2/24/2019 08:24	Suspected Major Injuries

WEMSIS Data	
Incident ID	**Incident Time**
31988	2/24/2019 08:23

DOH WEMSIS Dataset-Specific Limitations & Notes

Limiting linkages of crashes with WEMSIS records within 24 hours following a crash will likely exclude crash-related EMS utilization that occurs in the days following a crash such as transportation between care facilities due to downgraded care requirements or subsequent home-to-facility transport due to latent related conditions. However, including WEMSIS records from a broader time frame introduces potential for erroneously linking unrelated WEMSIS records with a particular crash.

Although WEMSIS contains personally identifiable information, other patient-level variables such as detailed injury descriptions, treatments, or outcomes are not available for this linkage process. Location information is sometimes missing or inconsistent, requiring derivation for linkage to collision datasets. EMS personnel may also classify incidents differently across agencies, leading to potential misclassification of motor vehicle collisions. Timing differences between EMS response and the actual collision event can further complicate temporal linkage. Finally, because linkage relies on PII, date, time, and location, missing or inconsistent identifiers can result in false matches or missed connections.

DOL License History

The DOL dataset is constructed from a subset of the WSDOT crash file and includes crash record identifiers. Consequently, DOL data is linked to crash data using those identifiers, with no need for more complex linking logic.

DOL License History Dataset-Specific Limitations & Notes

The DOL dataset is constructed based on DOL accounts associated with driver license numbers provided in the crash report. As a result, it does not include driver history for drivers without licenses, those providing invalid driver license numbers, or out of state drivers.

WSP Toxicology

Initially, potential linkages between WSP toxicology data and crash-involved motor vehicle drivers, pedalcyclists, and pedestrians are identified by matching records using PersonID. These linkages are limited to instances where the toxicology offense date matches the crash date, or where case numbers match between the crash and toxicology datasets, regardless of date.

While a single toxicology result may, in some circumstances, be associated with multiple crashes, it is not reasonable to assume general applicability across crashes that are separated by several hours. In such cases, the methodology attempts to identify the most appropriate crash or crashes for association. This is accomplished by first identifying the latest crash occurring on the relevant day. If at least one crash has a case number matching the toxicology record, the latest crash on that day with a matching case number is selected. Any additional crashes occurring within 120 minutes prior to this reference crash—including those extending into the previous day—are also included in the linkage set, while all other crashes are excluded. This approach is based on the assumption that a DUI-related blood draw or warrant execution typically results in arrest and booking, after which the individual is generally not operating a vehicle again that day.

Figure 16 illustrates a case in which multiple crashes are linked to a single WSP toxicology record. In this example, two crashes involving the same individual are initially linked to one toxicology case. The second crash has a matching case number and occurs on the toxicology offense date. It is therefore linked to the toxicology result. The first crash, while occurring on the preceding day, falls within 120 minutes of the second crash and is therefore also included in the linked set.

Figure 16. Multiple Crashes Linked to a Single WSP Toxicology Case

Crash data	
Date/Time	**Case Number**
12/31/2020 23:57	*NULL*
1/1/2021 00:02	21-00011

Toxicology data		
Case ID	**Case Number**	**Offense Date**
ST21-00001	21-00011	1/2/2021

Figure 17 shows the use of case number to determine appropriate linkages. In this example, two crashes involving the same individual are preliminarily linked to a single toxicology case. The first crash has a matching case number and is therefore linked to the toxicology result. The second crash does not share a matching case number and is therefore not linked to the toxicology result.

Figure 17. Using Case Number to Determine Appropriate Linkages

Crash data	
Date/Time	**Case Number**
3/15/2022 02:01	22-43031
3/15/2022 20:16	22-43488

Toxicology data		
Case ID	**Case Number**	**Offense Date**
ST21-00001	22-43031	3/15/2022

Figure 18 shows the use of case number and crash time to determine linkages. In this case, four crashes for an individual are preliminarily linked to a single tox case. One of the four is from the day prior to the toxicology offense date due to the matching case number. The reference date is the offense date because a crash with matching case number exists on that day. The third crash is the latest crash on the reference date with a matching case number, but the fourth is within 120 minutes of this and therefore becomes the reference crash. The last two crashes are then linked and the previous ones are not.

Figure 18. Using Case Number and Crash Time to Determine Linkages

Crash data	
Date/Time	**Case Number**
1/1/2021 23:04	21-00011
1/2/2021 09:10	21-00005
1/2/2021 15:19	21-00011
1/2/2021 15:25	*NULL*

Toxicology data		
Case ID	**Case Number**	**Offense Date**
ST21-00001	21-00011	1/2/2021

WSP Toxicology Dataset-Specific Limitations & Notes

The 120-minute window is an approximate threshold intended to reflect the expected duration of drug-related impairment affecting a driver's behavior. However, impairment durations vary substantially across substances, with some effects lasting shorter and others longer. As a result, when multiple crashes occur in close temporal proximity but span more than 120 minutes, earlier crashes may be excluded from linkage even when plausibly related. Conversely, in cases where a driver is involved in a minor crash, later ingests an impairing medication, and then drives again resulting in a second crash associated with a DUI assessment, the logic may link both crashes to the toxicology result even though only the second crash is causally related.

Crash data also exhibit patterns in which multiple crashes for the same individual, sometimes across multiple days, are assigned the same case number. This complicates the assignment of toxicology results to specific crash events. When multiple crashes sharing the same case number as a toxicology record occur more than 120 minutes apart on different days, the crash closest to noon on the toxicology offense date is selected for linkage, while the other is excluded. When multiple such crashes occur within 120 minutes of each other on the same offense date, the later crash is linked and the earlier crash is not. The extent to which these rules accurately reflect true causal relationships is uncertain.

Limitations & Notes

Data Quality

Various factors affect the accuracy of the episode linkage methodology described above. Ultimately, the ability of this process to reliably and accurately identify records associated with a single crash incident depends on the availability and quality of source data elements, as well as the accuracy of Forecasting and Research's identity resolution results. A review of the data indicates periodic discrepancies in reporting across datasets. In particular, inconsistencies exist among fields that record dates of crash-related incidents, resulting in documented dates that may differ by one or more days across WSDOT crash dates, WSP toxicology offense dates, WEMSIS incident dates, and related measures such as RHINO visit dates and Trauma Registry admission dates.

In general, the algorithms described above attempt to accommodate these discrepancies by initially linking records where incident dates are closely aligned, typically within one or two days. In subsequent steps, linkages are refined to prioritize records occurring on the same day when available. As noted above, additional variables used in the episode linkage process include WSDOT citation indicators, injury severity, crash county, AOC court county and charge codes, and WSP agency case numbers. The reliability of the episode linkage results is therefore largely dependent on the completeness, accuracy, and consistency of these data elements across systems.

Multiple Crashes

Identifying episodes and the corresponding crash-related outcome records in different datasets is relatively simple when a single crash exists for an individual. However, when an individual is involved in multiple crashes within a short period, the accurate association of the crash records with the corresponding outcome records becomes increasingly difficult. In many cases, a single outcome record can be associated with multiple crash records, or a single crash can be associated with multiple outcomes for an individual, and insufficient information exists to accurately distinguish which crash(es) are truly associated with which outcome(s). When this occurs, the linkage logic results in a linkage between all relevant crash and outcome records, instead of employing arbitrary or low-confidence assumptions to limit the linkage to a single crash, or outcome. This is done to avoid underrepresenting the presence of outcome data for the individual.

Cross-sector Research

By fostering collaboration among transportation agencies, public health organizations, law enforcement, and research institutions, the TRIP will ensure that crash data is not only comprehensive but also actionable. A well-integrated repository allows interested parties to analyze crash trends, identify high-risk areas, and develop data-driven strategies to enhance roadway safety. This level of coordination will enable policymakers to implement targeted interventions, such as improved infrastructure design, enhanced enforcement measures, and public education campaigns aimed at reducing risky driving behaviors. Additionally, sharing standardized data across agencies will facilitate more accurate and timely assessments of traffic safety initiatives, ensuring continuous improvements toward the Vision Zero 2030 goals.

The integration of diverse data sources, including emergency medical services, hospital records, insurance claims, and roadway characteristics, will provide a deeper understanding of crash risk factors and their broader societal impacts. This comprehensive dataset will enable researchers to examine how factors such as road conditions, vehicle technology, driver behavior, and environmental influences contribute to crash outcomes. Additionally, the ability to track injuries and fatalities over time will help identify disparities in traffic safety, ensuring that vulnerable populations, such as pedestrians, cyclists, and motorcyclists, receive the attention and resources necessary to enhance their protection. The TRIP repository will serve as a critical tool for guiding evidence-based policy decisions that prioritize equity and effectiveness in traffic safety measures.

By leveraging cross-sector collaboration, the TRIP will also support the development of proactive safety measures that prevent crashes before they occur. With access to high-quality, integrated data, researchers and policymakers can identify patterns and emerging risks, allowing for timely interventions such as road design modifications, enhanced traffic law enforcement, and public awareness campaigns. These preventative strategies will not only save lives but also contribute to a safer and healthier transportation system for all Washington residents. Ultimately, the TRIP's comprehensive approach to crash data analysis will play a vital role in shaping policies that reduce fatalities and serious injuries while promoting long-term public health and safety.

Cross-sector research and analyses are essential for improving traffic crash safety and public health by fostering collaboration between transportation, healthcare, law enforcement, and policy sectors. Traffic crashes are a leading cause of injury and death worldwide, and addressing this issue requires an integrated approach that considers road design, driver behavior, emergency response, and medical outcomes. By combining crash data from transportation agencies with injury reports from hospitals and public health departments, experts can identify high-risk areas, common injury patterns, and the most effective interventions. This data-driven approach helps design safer roadways, improve vehicle safety standards, and implement policies that reduce crash-related fatalities and long-term health consequences.

Technological advancements and innovative policies in traffic safety also benefit significantly from cross-sector collaboration. For example, partnerships between the automotive industry, urban planners, and public health experts have led to advancements such as crash-avoidance systems, better pedestrian infrastructure, and improved post-crash care. Research integrating data from law enforcement and medical institutions can also highlight the impact of impaired driving, speeding, and seatbelt use on crash severity. By analyzing this information, policymakers can implement targeted safety campaigns, stricter enforcement measures, and infrastructure improvements such as roundabouts, speed bumps, and dedicated bike lanes. These strategies not only reduce crash frequency but also enhance public health by preventing severe injuries and fatalities.

Cross-sector analyses also inform long-term strategies that promote sustainable transportation and healthier communities. Public health professionals working alongside transportation planners can advocate for policies that encourage walking, cycling, and public transit while reducing reliance on private vehicles. This shift not only decreases traffic congestion and crash risks but also improves air quality and reduces chronic health conditions such as respiratory diseases and obesity. Economic research further strengthens these efforts by demonstrating the financial benefits of investing in safer road infrastructure and preventative public health initiatives. Through cross-sector collaboration, societies can develop comprehensive, evidence-based solutions that improve both traffic crash safety and overall public health.

Critical questions related to data quality, linkage, and outcomes using TRIP may include:

- What is the data quality of the source data needed for linking? What can be improved?
- What percentage of crash records are successfully linked to corresponding records? What is the frequency of unlinked records?
- How often is there a positive toxicology result for a likely impairing drug and the officer did not code a contributing circumstance for under the influence of drugs? What are those drugs involved in the cases with drug toxicology positive for a likely impairing drug but no crash contributing circumstances for that driver?
- What is the agreement between BAC values from a toxicology record and police-reported BAC on crash reports? Does impairment information on the PTCR match the toxicological outcomes reported? i.e. How often is there a positive BAC toxicology that was not reported to the crash record?
- Can hospital records be used to derive a clinical assessment of injury severity to compare to/supplement the officer's assessment of injury severity? How do these assessments differ?
- What additional insights and risk factors can be gathered from linked data sources in non-fatal injury crashes? To what extent can non-fatal injury crashes inform the likelihood of fatal crashes in the same location?
- Can driver citation and adjudication history be a predictor of future crashes?
- What are the total objective hospital costs associated with non-fatal serious crashes resulting in medium- and long-term health care?
- What is the post 30-day mortality rate related to a serious crash (fatalities are currently defined as death within 30 days)?
- What are the characteristics and potential interventions for drivers involved in multiple crashes over time?
- What proportion of crashes involve RCW violating contributing circumstances, such as speeding, but no citations/adjudication were issued due to behaviors leading to the crash?