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 collisions. 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
360-902-0555	360-586-1988	P.O. Box 43113 Olympia, WA 98504-3113	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) – toxicology laboratories, 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

Table of Contents

Introduction

Purpose

This document describes the data available through the TRIP 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 Data

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 – Collision Location	Washington State Department of Transportation WSDOT maintains records of Washington State crash and collision information including location.
Driver License History and Ignition Interlocks Device (IDD)	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 WSP maintains driving under the influence (DUI)-related toxicology results.
Comprehensive Hospital Abstract Reporting System	Department of Health DOH maintains record level information on inpatient and observation patient community hospital stays in Washington State.
Rapid Health Information Network	Department of Health DOH maintains real-time syndromic surveillance patient data from emergency medical services (EMS) and participating hospitals and clinics in Washington State.
Emergency Medical Services Information System (EMS) and Washington State Trauma Registry (WTR)	Department of Health 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.
Death Vital Records	Department of Health DOH maintains collection of death records that took place in Washington State.
Trauma Data	Department of Health DOH maintains data related to Washington State's Trauma Registry which captures information for seriously injured individuals.

TRIP Repository Administrative Data Limitations

While all the data sets 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

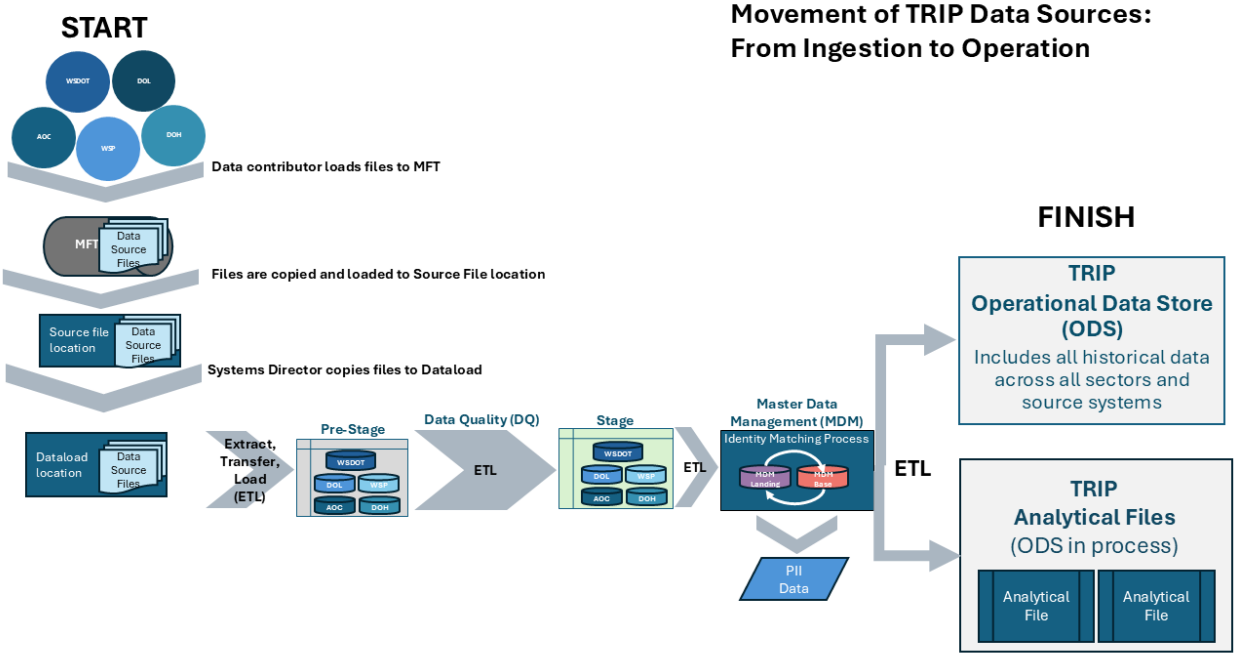
and specific state or federal reporting or monitoring requirements. Administrative data is collected as both transactional and summative datasets by local administrators and submitted to an agency authority, making variance among data collectors a potential source of bias in each dataset. Quality control 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, the revised crosswalk of P20IDs and token IDs is incorporated into P20W Education Data System Operation Data Store (ODS). Once in the ODS, data are de-identified and available for analysis. Identifiers used in the identity resolution process do not advance beyond the Master Data Management (MDM) hub.

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



Identity Resolution Process

The core 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 identity resolution is to create linkages across multiple data sources so that crash records are linked to related records 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 to create unique person identifiers. These identifiers are referred to as "P20IDs." P20IDs are assigned to all individual-level data received by TRIP from our data contributors. As additional linking activities occur, P20IDs will be updated to reflect the most recent data available.

It is important to understand the identity resolution process so the researcher can evaluate whether it may impact the analysis, especially if the research includes linking the TRIP repository data to additional data.

Creation of Token IDs and Assignment of P20IDs

Before a person-level P20ID can be created, an identity resolution "token" is created for each record in a dataset. Identity resolution tokens, known as "pkeys" in the TRIP data system, are concatenated sets of identifiers that are guaranteed to be unique to an individual. As a result, TRIP cannot rely on SSNs to uniquely identify individuals in this source of data. TRIP has found that this set of identifiers is the best way to uniquely identify individuals in WSDOT data, given the data available in the dataset to be unique to an individual in WSDOT data. The data is then loaded into the MDM system, where the system generates integer based surrogate key, called a TokenID for each pkey token. As an aside, TokenIDs are used to merge and unmerge P20IDs when over- or under-matching occurs. Over-matching occurs when P20IDs contain TokenIDs for two or more people. Under-matching occurs when two or more different P20IDs actually refer to the same person.

Table 2: Identity token components in data sources

Sector	Contributor	Birthdate	First Name	Middle Name	Last Name	SSN	License #
WSDOT	WSDOT	x	x	x	x		x
DOL	DOL						x
WSP	TOX	x	x	x	x		
AOC	AOC	x	x	x	x		
DOH	CHARS	x	x	x	x	x	
	DEATH	x	x	x	x	x	
	RHINO	x	x	x	x	x	
	TRAUMA	x	x	x	x	x	
	WEMSIS	x	x	x	x	x	

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, there is no way to directly match these together, but they can be indirectly matched by involving other sources of individual data.

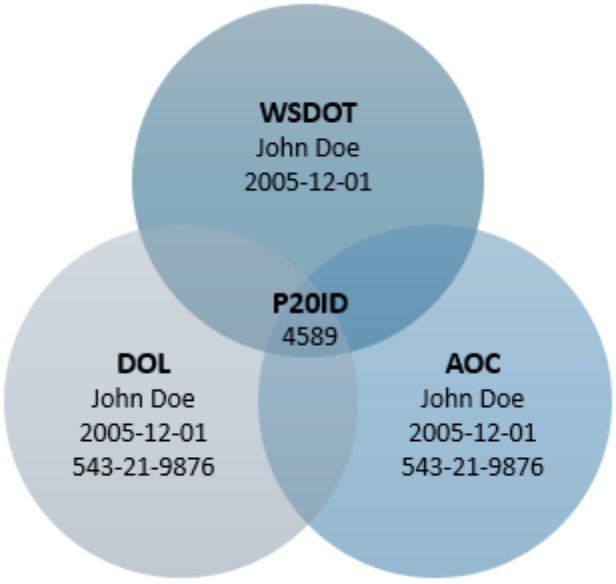
Phases of Identity Resolution

TRIP's identity resolution process has two phases.

1. High probability match phase - P20IDs are provisionally matched using MDM's "automerge" match rules. These are conservative rules for matching P20IDs. In this phase undermatching (or not correctly identifying an actual match) is not a significant concern, as the automerge match rules and SQL based rules are designed to ensure extremely low false positive rates. These match pairs are not manually reviewed, though they are subject to cardinality analysis (i.e., cardinality is a concept used in various fields such as mathematics, data modeling, and SQL to describe the relationship between sets or the uniqueness of data values), after which the match pairs are then merged.

2. Mid-probability match phase – After the high probability matches have been merged, looser match rules are used to create additional match pairs. The advantage of looser match rules is that they can potentially create additional true-positive match pairs. But this comes at the cost of also creating additional false-positive match pairs. Given that there is a mixture of true-positive and false-positive match pairs, this set of match pairs is manually reviewed. The results undergo cardinality analysis, and then the true-positive match pairs are merged.

Figure 2: Identity token components in data sources



High Probability Match Phase: Match Rules and the Creation of Match Pairs

TRIP's identity resolution process involves matching new P20IDs against the pre-existing P20IDs in MDM using a set of match rules. An example of a match rule is "Same names, same birth dates, same SSNs." When new P20IDs are matched against the pre-existing P20IDs, the result is a set of prospective P20ID match pairs.

The quality of the match pairs depends on the match rule. For match rules involving the exact match between many ID fields, such as the "same names, same birthdates, and same SSNs", the false positive rate is low and the matches can be provisionally accepted, subject to cardinality analysis described below. The quality of the match is lower the match rules involve a small number of fields, fuzzy logic is

used (such as SOUNDEX), or matching the first three characters of the last names. These lower quality matches pairs are manually reviewed.

Mid-Probability Matched Phase – Manual Review

In manual review, each match pair is assigned a probabilistic score of the likelihood of a positive match. The probabilistic score, match pairs, and associated identifiers (names, dates of birth, etc.) are brought into the same dataset. The match pairs are then manually reviewed, and 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 MDM. At this point, MDM contains the set of provisional match pairs. These prospective match pairs are then subject to cardinality analysis.

Cardinality Analysis Phase

Cardinality analysis is a key component of the identity resolution process. It allows for more aggressive matching, while at the same time, improving the quality of the existing P20IDs in MDM. In cardinality analysis, the provisional match pairs are merged on a trial basis. Then the cardinal relationships are determined between the P20IDs in the subject dataset and those in the repository being matched against. These relationships can be 1:1, 1:Many, Many:1, or Many:Many. For example, a 1:Many relationships indicates that one P20ID in the subject dataset matches multiple P20IDs in the universe of data being matched against. The 1:1 relationship, where P20ID matches exactly, are accepted. The relationships involving non 1:1 relationship need to be manually reviewed to resolve them as accurately as possible.

Once the cardinal relationships are verified, the results are fed back into MDM. The result is that some P20IDs might be unmerged, some could be merged, and some provisional match pairs potentially are shown not to be actual match pairs. At the conclusion of this iterative process, the P20IDs are integrated into the data contained in the ODS for use in analysis.

Core TRIP Contributor Data

This section provides a set of descriptions or quick references to the core data files that feed into the TRIP repository. 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 Data Handbook is designed to be used in conjunction with the TRIP Data Dictionary and the TRIP Data Status tool to understand the specific data elements and years of data available. To utilize this additional information, 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 collisions, 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 crashes, 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 crashes 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 crashes. This evidence-based approach ensures that state resources are directed toward initiatives that have a measurable impact on improving road safety. Additionally, the data is used to support grant applications and funding requests for transportation safety projects at both the state and federal levels.

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's licenses, vehicle ownership records, license suspensions, 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, and other infractions that could endanger public safety. 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.

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.

Washington State Patrol

The Washington State Patrol (WSP) Toxicology Laboratory is responsible for analyzing biological samples related to impaired driving, drug-related crimes, and forensic investigations. This laboratory plays a critical role in supporting law enforcement agencies by testing blood and other bodily fluids for the presence of alcohol, drugs, and other toxic substances. The data collected from these analyses helps determine impairment levels 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 DUI (Driving Under the Influence) 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, marijuana, prescription medications, or illegal drugs were in the driver's system at the time of the incident. This data is crucial for prosecuting 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.

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 and injury 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.

The DOH also oversees vital statistics, including birth and death records, which are crucial for public health planning and research. Birth data helps track maternal and infant health trends, while mortality data is used to identify leading causes of death, disparities in health outcomes, and the impact of diseases on different communities. This information is essential for shaping policies related to healthcare access, preventive care, and emergency response planning.

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 stakeholders 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 Target Zero 2030 goals.

The integration of diverse data sources, including emergency medical services, hospital records, crash records, 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.