Authors
Waetjen, David, PhD
Shilling, Fraser, PhD

Publication Date
2020-09-01

DOI
10.7922/G2W66J27

RESEARCH REPORT

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

Leveraging the California Highway Incident Processing System for Policy and Research

David Waetjen, Ph.D., Principal Investigator, Road Ecology Center, University of California, Davis
Fraser Shilling, Ph.D., Co-Director, Road Ecology Center, University of California, Davis

September 2020

ITS UCDAVIS INSTITUTE OF TRANSPORTATION STUDIES

Report No.: UC-ITS-2020-02 | DOI:10.7922/G2W66J27

Technical Report Documentation Page

1. Report No. UC-ITS-2020-02	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle Leveraging the California Highway Incident Processing System for Policy and Research		**5. Report Date** September 2020
		6. Performing Organization Code ITS-Davis
7. Author(s) David Waetjen, Ph.D., https://orcid.org/0000-0001-5817-3279 Fraser Shilling, Ph.D., https://orcid.org/0000-0001-8125-2186		**8. Performing Organization Report No.** UCD-ITS-RR-20-67
9. Performing Organization Name and Address Institute of Transportation Studies, Davis 1605 Tilia Street Davis, CA 95616		**10. Work Unit No.** N/A
		11. Contract or Grant No. UC-ITS-2020-02
12. Sponsoring Agency Name and Address The University of California Institute of Transportation Studies www.ucits.org		**13. Type of Report and Period Covered** Final Report (November 2019 – June 2020)
		14. Sponsoring Agency Code UC ITS

15. Supplementary Notes
DOI:10.7922/G2W66J27

16. Abstract
There are two official sources of data on traffic incidents in California: 1) the Statewide Integrated Traffic Records System (SWITRS), intended to include incidents leading to injury or death; and 2) the California Highway Patrol (CHP) data on Caltrans' Performance Measurement System, PeMS. Traffic safety researchers rely heavily on the post-processed SWITRS database, which provides only some crucial information about crashes. In 2015, the Road Ecology Center at UC Davis developed a third method to collect all incident data that appear on the CHP real-time incident-reporting web-site (https://cad.chp.ca.gov/). These data are assembled into a database called CHIPS, for California Highway Incident Processing System. Analyses indicate that the number of incidents recorded in a given period are similar in CHIPS and SWITRS but lower in PeMS. Also, many SWITRS records (e.g., 36% in 2018), but no CHIPS records, lack or have inaccurate location information on incidents. Through case studies, the research group examined three ways that CHIPS can be used to support data and policy analysis. This report proposes future pathways for creating a more integrated system for collecting and analyzing crashes.

17. Key Words Traffic incidents, databases, data fusion, data collection, case studies, crash data, animal vehicle crashes, traffic data		**18. Distribution Statement** No restrictions.	
19. Security Classification (of this report) Unclassified	**20. Security Classification (of this page)** Unclassified	**21. No. of Pages** 34	**22. Price** N/A

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

About the UC Institute of Transportation Studies

The University of California Institute of Transportation Studies (UC ITS) is a network of faculty, research and administrative staff, and students dedicated to advancing the state of the art in transportation engineering, planning, and policy for the people of California. Established by the Legislature in 1947, ITS has branches at UC Berkeley, UC Davis, UC Irvine, and UCLA.

Acknowledgments

This study was made possible through funding received by the University of California Institute of Transportation Studies from the State of California through the Public Transportation Account and the Road Repair and Accountability Act of 2017 (Senate Bill 1). The authors would like to thank the State of California for its support of university-based research, and especially for the funding received for this project. The authors would also like to thank Kathryn Harrold for originally introducing us to the CHP incident reporting website; Drs. David Shatz and Christy Adams (UC Davis Medical Center) for input about medical aspects of trauma treatments associated with traffic crashes; and Drs. Nana Tufuoh and Orion Stewart (California Department of Public Health) for their advice and input about statewide trauma reporting databases.

Disclaimer

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

Leveraging the California Highway Incident Processing System for Policy and Research

David Waetjen, Ph.D., Principal Investigator, Road Ecology Center, University of California, Davis
Fraser Shilling, Ph.D., Co-Director, Road Ecology Center, University of California, Davis

September 2020

ITS UC DAVIS INSTITUTE OF TRANSPORTATION STUDIES

Report No.: UC-ITS-2020-02 | DOI:10.7922/G2W66J27

Table of Contents

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CHIPS	California Highway Incident Processing System
CHP	California Highway Patrol
CSV	comma separated values (a type of computer file)
FARS	Fatality Analysis Reporting System
PeMS	Performance Measurement System
SWITRS	Statewide Integrated Traffic Records System
UC	University of California
WVC	wildlife-vehicle conflict

Executive Summary

Executive Summary

Understanding proximate causes and outcomes of traffic incidents relies on accurate data collection. The California Highway Patrol (CHP) helps to monitor, manage, and maintain safety on California roadways, and their personnel are often first to arrive at traffic incidents on highways, rural roads, and major arterials. The CHP publishes incident reports in real-time, along with information about road conditions, natural disasters, etc., on the public *CHP Traffic Incident Information Page*[1], in part so that other agencies can monitor activity in their respective regions. These data are invaluable because they capture the real-time communication between CHP officers (on the scene) and their dispatch center. Unfortunately, only a partial set of these data are available for public download, and therefore they do not represent the whole state.

Two official sources of data on traffic incidents occurring in California are: 1) the Statewide Integrated Traffic Records System (SWITRS), which includes incidents leading to injury or death; and 2) the CHP Incident data on Caltrans' Performance Measurement System, PeMS, which only includes incidents on state highways. Almost all traffic safety researchers in California rely upon the post-processed SWITRS database, which provides substantial information about crashes, but ~1/3 of records lack accurate location data, and all records lack detailed descriptions of events during the incident, including the incident timeline. The Road Ecology Center at UC Davis has developed a third method to collect all incident data that appear on the CHP site, and it has assembled them into a database called CHIPS, the California Highway Incident Processing System. Started in February 2015, the database documents roadway incidents across a variety of subject domains and shows temporal differences in roadway activity over time. We created CHIPS to capture incident data in real-time directly from CHP field reports, because they contain the description of the incident, as well as accurate location data.

CHIPS is a useful tool for transportation agencies and researchers because it is the most complete and accurate tool currently available to collect, manage and query incident reports for events on California state highways and other roads patrolled by the CHP. A comparative analysis indicated that the number of incidents recorded in a given period (i.e., 2018) are similar in CHIPS and SWITRS but lower in PeMS, as expected, because PeMS provides only state highway incidents (Figure ES-1). Also, many SWITRS records (e.g., 36% in 2018) lack or have inaccurate location information. However, these problems may not be present in CHIPS, as the CHIPS incidents have locations that coincide with those in PeMS and/or within the footprint of state highways and major rural roads. This agreement with road locations makes sense given that the location data are automatically uploaded from the CHP vehicle's GPS. A combination of SWITRS and CHIPS may provide a database more complete and reliable than either one individually.

1 https://cad.chp.ca.gov/

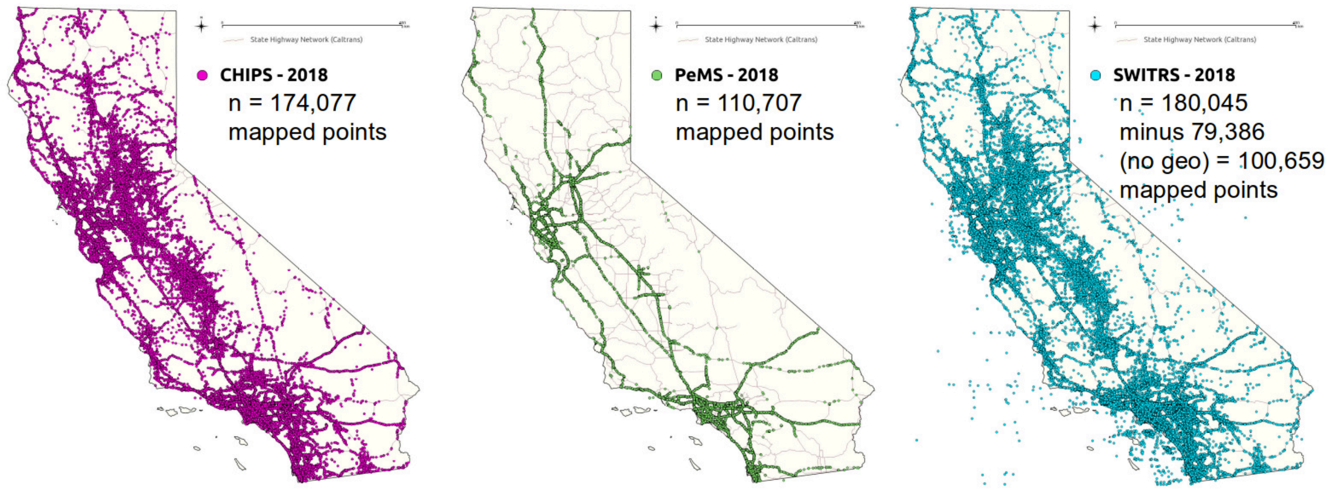


Figure ES-1. Spatial comparison of CHIPS, PeMS, and SWITRS records for 2018.

We conducted three case studies in which CHIPS played an essential role. In the first, we investigated how the California Governor's "shelter-in-place" order to mitigate the spread of COVID-19 impacted the number of traffic crashes on state highways. Because CHIPS data are automatically collected and managed in real-time, the case study results show the immediate impacts of the shelter-in-place orders on traffic crashes. The resulting report is available online (https://roadecology.ucdavis.edu/files/content/projects/COVID_CHIPs_Impacts_updated_415.pdf). In the second case study, we convened a meeting with civil engineering, public health, and healthcare professionals to consider systems that would track (non–personally identifiable) information from an accident through the victims' health outcomes. This would enable improved analyses linking crash site characteristics to health outcomes. In the third case study, we expanded CHIPS to include information on incidents involving animals. We used a subset of these data in a real-time "deer-vehicle-collision" map, which updates every 15 minutes and shows deer related incidents on highways, as well as hotspots of collisions with large mammals.

Finally, we describe possible pathways forward for creating a more integrated system for collecting and analyzing crash reporting across this arc from highway incidents to health outcomes. These potential improvements include the following:

- Identify and associate CHIPS records with their corresponding SWITRS records to improve the spatial and temporal accuracy and completeness of SWITRS (and other) datasets
- Create an interactive data portal that would allow a general user to submit ad-hoc queries that would retrieve only the type of data the user is seeking
- Create automated processes for continuing to retrieve, store, and manage data, as well as validate data quality and completeness.
- Have the narrative details included by the CHP officer processed with tools that can help identify and "discover" incidents when a search is performed.

Contents

Introduction

Analyzing proximal causes of traffic incidents requires accurate spatial location, temporal values, environmental conditions, involved parties, and infrastructure information (Yan et al., 2017). Automated reporting of traffic incident details is affected by legacy approaches to on-scene and post-hoc reporting of critical details, such as location, start and clear times, environmental conditions, and road curvature/slope (Bejleri and Brown, 2014). This has led to proposals to standardize data collection formats and to develop composites of incident information, including initial reporting of incident details (Santiago-Chaparro et al., 2016) and automatically managing emergency responses using shared and integrated computing services (Chen and Englund, 2018). With the advent of WAZE, a crowd-source method of reporting roadway incidents, new sources of information are becoming available that improve the rapidity and completeness of data collection about incidents both as they occur and in retrospect (Amin-Naseri et al., 2018; Young et al., 2019).

Jurisdictions have recently begun creating databases to capture important information associated with traffic crashes. For example, San Francisco has TransBASE.org, to "serve as the central data repository for public health–related transportation data" (Morris and Weir, 2016). This is the important first step in connecting crash incident data with health outcomes, which is critical to accurately devise plans for reducing injury and fatal crashes. A study sponsored by the National Highway Transportation Safety Administration (NHTSA) describes three primary methods to retrospectively associate hospital/health record data with crash data: 1) deterministic linkage, where identifier codes associated with incidents, emergency response, and hospital records can be associated to link the data; 2) probabilistic linkages, where crash and health characteristics are known and in common among the crash and hospital datasets and can be used to connect incidents with health outcomes; and 3) spatio-temporal linkage, which uses locational and temporal information to associate crash and hospital records (Cherry et al., 2018). All methods can have unknown sources of uncertainty and bias, and understanding and limiting these is key to developing linkages, until linkage is unnecessary due to the implementation of a single unique identification system for crash victims that follows them through treatment, such as that used in Oregon. Conderino et al. (2017) have demonstrated that it is possible to use probabilistic linkage analysis to connect specific health outcomes with the related crash incidents. However, in that study only 52% of records could be connected back to the original crash event.

The California Highway Patrol (CHP) monitors California roadways and manages most traffic crashes and hazards on the major state and federal highways. Their incident reports tell the story of what happens on California's roads and highways. They are often first to arrive at the scene of an incident, and their narrative describes important event details and observations. The descriptions cover officers' observations and actions, and provide an accurate timeline of events. The incident location is recorded automatically for an accurate geospatial reference.

The Road Ecology Center at UC Davis has been collecting all information posted to the CHP's Incident Reporting Page since February 2015 and storing them in a local database called the California Highway Incident Processing System (CHIPS). CHIPS has collected over 4 million independent incident records since inception, and while this (currently-private) database does not contain moving violations, it does include other daily CHP activities, including help following traffic collisions, traffic management (such as lane closures), natural disaster response (floods, fires), and public safety measures (during high wind or foggy conditions). All these data include CHP officer communication with their local dispatch center, who timestamp each interaction. Examples include whether an ambulance was required or not (CHP code 1141), whether there is a possible fatality (code 1144), the types of vehicles involved in the accident, and the number of lanes closed. Mining the records' textual descriptions can yield a rich set of time-series data that can be invaluable to traffic safety studies.

Recently, incident report narratives have been mined to obtain new information about incidents (Trueblood et al., 2019). While the form-based fields that an officer completes offer a more structured approach to the data record, the narrative they provide is an untapped resource which can yield critical details of the incident, and studying these details can shed new light on traffic safety. While it can be a time-consuming process for an individual to read though narratives and annotate them in a structured way for quick reference, there are text-based data mining tools which can do the work.

Caltrans publishes the CHP incident records in their Performance Measurement System (PeMS). However, as discussed later in this report, when we compare these data to CHIPS data, they appear incomplete and in some cases the data appear to be removed. Another public data source that contains California traffic accidents is the CHP published Statewide Integrated Traffic Records System (SWITRS) database. While these post-processed records provide substantial information about traffic incidents, about 1/3 lack accurate location data, and all lack narrative descriptions of the incident. Compared to both of these public systems, CHIPS is more spatially-accurate and more complete. If there was a way to integrate CHIPS and SWITRS records, then our understanding of health outcomes from traffic incidents would improve.

We have identified several areas where CHIPS data can be invaluable to understanding the characteristics of California highways and traffic safety. We present these instances as case studies showing how CHIPS data was used (or, in one case, can be used) to improve our understanding of road conditions and highway safety. We propose that CHIPS is a viable research instrument for state agencies, transportation planners, and academic researchers.

Including Stakeholders

Investigating traffic safety on California roadways includes a wide range of entities interested in proximate causes of crashes, rates and costs of crashes, connection to policies (e.g., speed limits), and ultimately engineered solutions. We decided to focus on the health and safety aspects of crashes and hosted a meeting of health-stakeholders on February 20, 2020 at the UC Davis Medical Center. Present at the meeting were representatives from UC Davis Medical Center, California Department of Public Health, UC Davis Department of Civil and Environmental Engineering, and UC Davis Road Ecology Center. While a California Highway Patrol representative was invited, they were not able to attend. We focused the meeting on the reporting arc, from the initial incident to health outcomes for those affected. We gave a presentation on CHIPS, highlighting the differences between CHIPS data and other publicly available data sources such as SWITRS, PeMS, and Fatality Analysis Reporting System (FARS, https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars). We led a round table discussion on how CHIPS could help solve the important problem of tracking crash victims from the incident scene to and out of the hospital, in order to contribute to understanding rates, costs, and ultimately solutions to crash impacts. We discussed possible solutions, including associating a unique code with a crash victim from the accident scene that follows him/her into the medical and trauma systems so the characteristics of crashes can be examined and connected to health outcomes.

Approach

The Mechanics of Incident Report Collection and Management with CHIPS

CHIPS begins with a small program that copies the information posted on CHP's public Traffic Incident Information Page (https://cad.chp.ca.gov/) to a compatible data format which is then imported into a local database. This program uses a method called "screen scraping" to extract these data and generates a delimited comma separated values (CSV) file. The data posted to the CHP Incident website are only present for a short period, presumably while the CHP officer actively works on the incident. The information gathered on each incident is continuously updated and time-stamped as dispatch centers receive additional details. Once the incident is completed, it drops off of the list on the CHP Incident site. Because of the temporary nature of these incident reports, the CHIPS program collects information every 15 minutes and either updates its active records or creates new incidents on each subsequent run.

Once the real-time incident data has been extracted from the CHP site to a CSV file, they are uploaded to a Relational Database Management System where queries can be issued against the data for more precise extraction. The records are stored in a MariaDB database, with the table structure and indexes supporting a Drupal 7 website. Post-processing techniques are used to remove serial record duplicates (e.g., sequential report updates from CHP officers) before using the data. The web framework is set up to accumulate records and provide simple ways of viewing and extracting data. The database also provides the storage system for an internal web data portal (written in Drupal 7), which allows Road Ecology Center personnel to view individual records, and, importantly, annotate the data so new datasets can be assembled for additional research. The web portal provides important data management features such as provenance (explicit recording of how the incident record changes over time), record level editing (to annotate and/or clean up the data), and filtered export of records that can be consumed by other services, such as an R model or an automatic query of incidents of a specific type, such as deer-incidents, which appear on our real-time animal crash map. Each record contains: 1) a spatial location, automatically recorded from the vehicle GPS device, as well as term-based location information; 2) the date and timeline; 3) the "Type", or category of incident; and 4) "Detailed Information", which is the CHP officer's narrative account of the incident (Figure 1). The CHIPS portal adds the map to provide context for the incident.

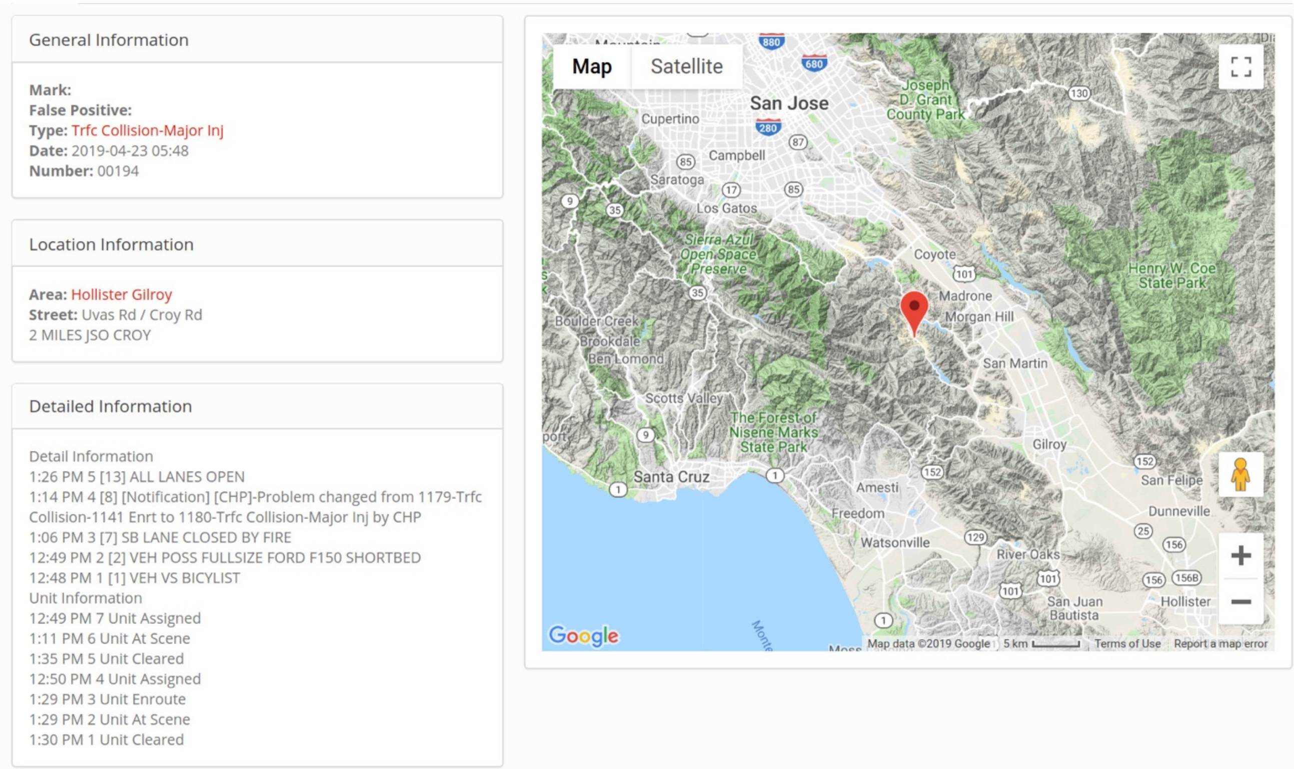


Figure 1. A processed CHIPs record with CHP officer's description of the event and a map added by the web-system.

In the example shown (Figure 1), the incident report is for a major collision between a vehicle and a bicycle. We know from the code 1141 that an ambulance was called and from the incident type that the accident was not immediately fatal. We can determine that the SB=Southbound lane closed for 20 minutes due to a fire, presumably due to the crash. The vehicle could have been a Ford F150, an attribute that can potentially help when trying to link this CHIPS record with another data source, such as SWITRS. (Note: SWITRS records will include the make and model of all vehicles involved in a collision.) We can automatically discover these keywords ("fire," "bicycle," "Ford F150") and use them to index the record within the database for easier searching and discovery.

To save valuable time at the scene, CHP officers use a set of codes to facilitate describing the incident. For example, 1141 is the code that an ambulance was called to the scene of the accident. Besides the large set of numbered codes, many officers will use a shorthand for descriptions, such as "MC" for motorcycle. Subjective observations are included in the Detailed Information field and therefore the CHIPS record, but these are not included in SWITRS. Many of the coded and consistent terms can be used as the basis for categorical and text-based queries and classification of records in CHIPS.

Native Vocabularies

The CHP Traffic Incident Information Page uses two vocabularies that have been migrated to CHIPS. “Vocabulary” is a term used in informatics describing terms that are consistently used to describe incidents. One vocabulary includes California place names (area); the second vocabulary is composed of terms that describe the types and characteristics of the incidents.

The area vocabulary provides a way to group incidents by the region in which they occurred. Many of these regions are defined based on having a local CHP office that is responsible for that vicinity. There are currently 137 recognized area names in the system. We have found that some CHP officers do not always enter this field correctly, so there might be an issue with how this area vocabulary is populated. The Road Ecology Center has had to resolve some issues with this field when it was not entered in a consistent manner.

The incident type vocabulary provides a way of classifying incidents into various types, including: Animal Hazard, Sig Alert, Traffic Collision, Fatality, and 45 others (49 total). Incident types provide a rudimentary way of understanding the nature of an incident without reading the full narrative description. Since CHP incidents can only have one incident type associated with an incident, it can sometime change while the incident is being processed. For example, a “Major Traffic Collision” incident could change to a “Fatality” if the crash was severe enough.

Please see Appendix A: CHIPS data summaries, for a listing of these two vocabularies and the number of incident records associated with each term (to date).

Comparison with Other Systems

There are several formal systems that California uses to report crashes, especially those that result in injury or death to drivers or passengers. Caltrans' Performance Measurement System (PeMS; https://pems.dot.ca.gov) shares CHP incident reports for state highways, the Statewide Integrated Traffic Records System (SWITRS, https://www.chp.ca.gov/programs-services/services-information/switrs-internet-statewide-integrated-traffic-records-system), and the Fatality Analysis Reporting System (FARS, https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars).

Since the start of this project, the CHP report data available on PeMS went from being a month or so out-of-date to being close to real-time, with a ~1.5 hour delay. In addition, the "Detail" field in CHP reports was originally not available in PeMS, but now it is, by clicking each incident id number. Some of these changes may have occurred as a result of the CHIPS system being part of frequent reports and a Transportation Research Board webinar from the Road Ecology Center on COVID-19 mitigation–related changes in traffic and crashes. Although the information content in the query view on PeMS has improved, it does not seem possible to download a complete dataset of incidents (there is a limit of 1 week), nor are coordinates provided, and incidents are exclusive to state highways, not all CHP-patrolled roadways.

Comparison with PeMS

We compared the rates of different categories of crashes as reported in CHIPS and PeMS for 2018 (Table 1). In most cases, there were fewer PeMS CHP incident records than were reported directly by CHP officers in real-time as recorded by CHIPS. In no case were there more PeMS CHP records in a particular category. For injury crashes during January 2018, PeMS CHP recorded 9,955 incidents and CHIPS recorded 15,530. This 50% higher number in CHIPS than in PeMS is in part because CHIPS includes incidents on non-state roadways, whereas PeMS only reports incidents on state highways.

Table 1. Comparison of CHIPS and PeMS records for January 2018

CHIPS Incident Type	PeMS Incident Type	CHIPS Count	PeMS Count	% (PeMS of CHIPS)
Trfc Collision-Unkn Inj	1183-Trfc Collision-Unkn Inj	8381	7056	84.2
Trfc Collision-1141 Enrt	1179-Trfc Collision-1141 Enrt	5337	2415	45.3
Trfc Collision-1141Enrt	1179-Trfc Collision-1141Enrt	398	395	99.2
Hit and Run w/Injuries	20001-Hit and Run w/Injuries	246	67	27.2
Trfc Collision-Minor Inj	1181-Trfc Collision-Minor Inj	1020	39	3.8
Fatality	1144-Fatality	99	2	2
Trfc Collision-Major Inj	1180-Trfc Collision-Major Inj	148	1	<1
Total		15,629	9975	63.8%

SWITRS

The database thought of as the official record of traffic injuries and fatalities is the Statewide Integrated Traffic Records System (SWITRS, https://www.chp.ca.gov/programs-services/services-information/switrs-internet-statewide-integrated-traffic-records-system). The data originate from CHP reports of collision scenes, are post-processed, and can take several months to a year to become available. The data are delivered as comma separated values (CSV) files and there are three main tables: collision, party, and victim, which each have 76, 33, and 12 attributes, respectively. The attributes include many aspects of an incident, from the crash severity, current road conditions, types of vehicles involved, and the age and gender of the crash victims.

We found that for records from 2018, 36% are missing spatial data, and many of the location points are located in the Pacific Ocean (Figure 2). Even though 58% of the SWITRS records that have valid spatial data appear to be located on or near a state highway, it is difficult to determine the accuracy of this apparent location given the range of inaccuracies apparent in the data as a whole. In comparison, almost 100% of CHIPS records are located within the footprint of state highways and major rural roads, which makes sense given that the incident location data are automatically uploaded from the CHP vehicle's GPS. PeMS records similarly fall on state highways (Figure 2).

The total number of CHIPS injury/fatality records and SWITRS records is similar across a complete year (2018, n-values in Figure 2) and greater than the total in PeMS, which are only for state highways. However, the SWITRS total that can be mapped is more similar to the PeMS total. This finding suggested to us that possibly the SWITRS dataset was similar to CHIPS, but that ~40% lacked spatial location. If there was a way to identify each CHIPS record corresponding to each SWITRS record, then possibly the SWITRS and CHIPS records could be combined to provide a more accurate, up-to-date (real-time), and complete record than each of its parts.

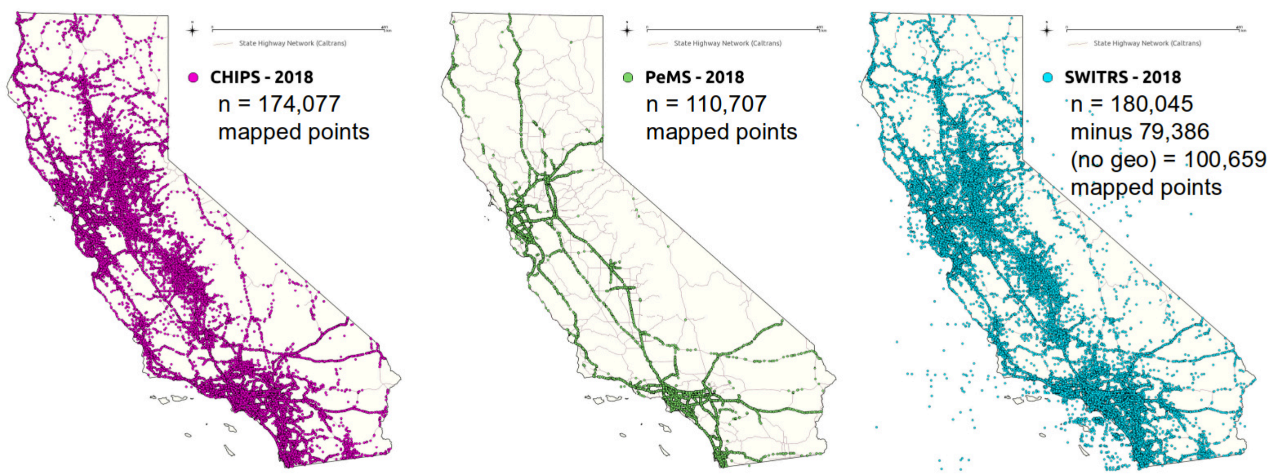


Figure 2. Spatial comparison of CHIPS, PeMS, and SWITRS records for 2018. SWITRS points lacking geographic data were not mapped.

FARS

The FARS is operated by the US Department of Transportation, National Highway Traffic Safety Administration (NHTSA). It is the repository for state reporting of fatal crashes on state highways. We found that there was a reasonable match between SWITRS and FARS fatality reports and between CHIPS and each of the other two. However, there was not a perfect match among any pair of the datasets, suggesting that there is room for improvement for this very important type of crash.

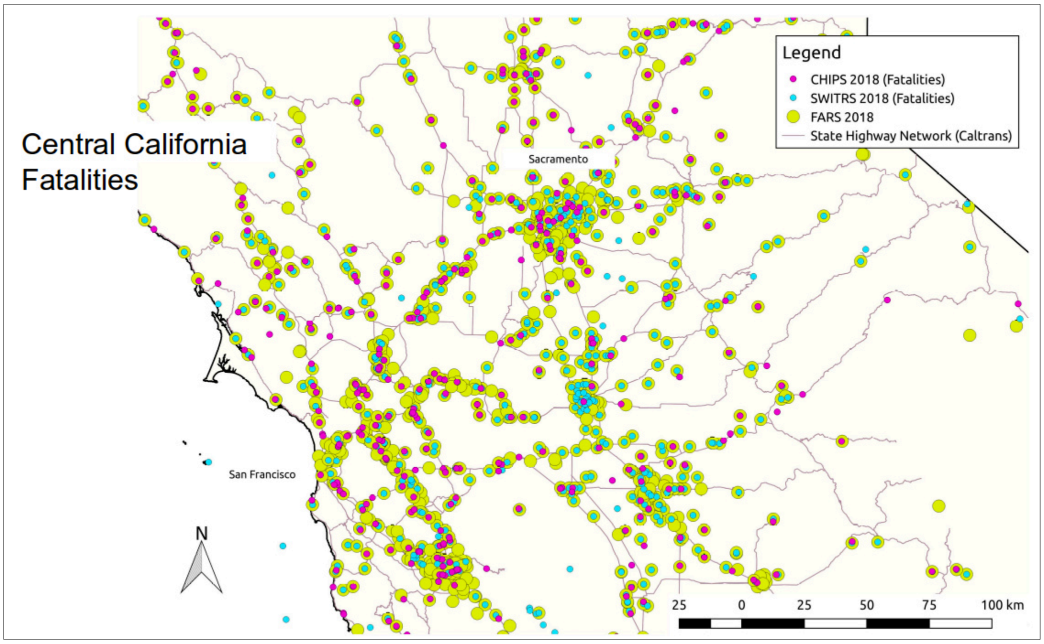


Figure 3. Map-comparison of CHIPS, FARS, and SWITRS records of fatal crashes.

Comparison Summary

The comparison among datasets for traffic crash-related injuries in California provided evidence of important gaps, as well as opportunities for developing an improved, integrated system. SWITRS is often assumed to provide complete fatality and injury data for highways and major roads; however, we found that about 1/3 of records lack accurate location data. CHIPS has accurate location data and has more collision and injury reports than SWITRS or PeMS and includes narrative and timeline data that SWTIRS lacks. The injury records do not always spatially align with SWITRS, but this seems to be because of inaccuracies in SWITRS data. PeMS has a similar data model to CHIPS, because data come from the same source, but we found large discrepancies between CHIPS and PeMS for some Incident Types. PeMS data appear to be spatially-accurate (fall on highways) but are limited to state highways, and there are unexplained discrepancies for certain incident types, such as fatalities. FARS data seem to be a close but incomplete match to SWITRS and CHIPS, leaving outstanding questions about whether this important type of incident is being captured.

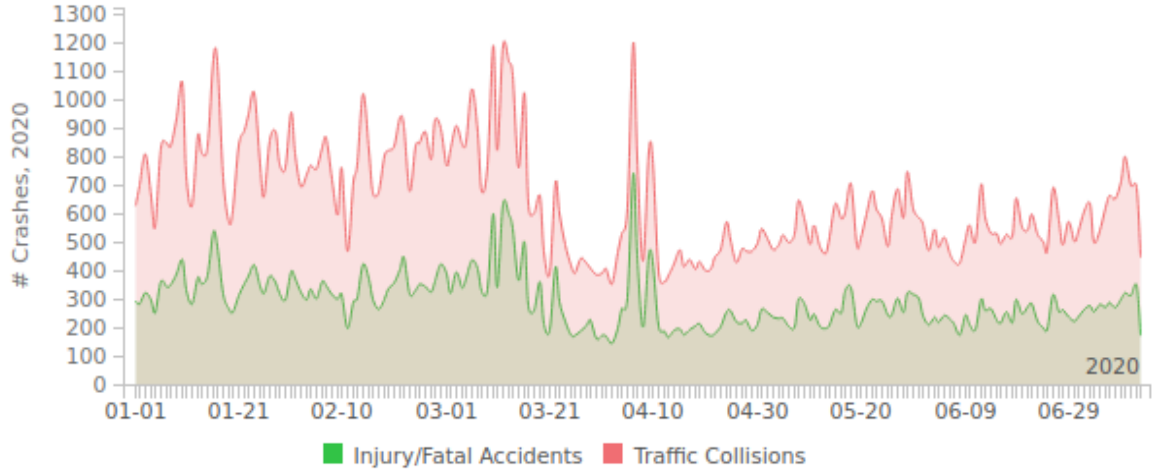
Case Studies

Through case studies, we investigated three ways that CHIPS can be used to support data and policy analysis. In the first case, we've been able to use CHIPS real-time reporting to investigate the immediate impacts of Governor Newsom's "shelter-in-place" order related to the COVID-19 pandemic on traffic crashes, especially those resulting in injury or death. In the second case, we worked with the UC Davis Medical Center and California Department of Public Health to develop improved methods to connect health outcome records with highway traffic incidents in order to improve traffic safety analyses. Finally, we use text-queries to extract deer-vehicle-collisions from the CHIPS database to publish in a real-time animal crash map (https://roadecology.ucdavis.edu/hotspots/map).

Case Study 1: Traffic Safety Related to the Governor's "Shelter-in-place" Order

The COVID-19 pandemic created unprecedented challenges to society and elected officials. One of the primary methods to mitigate the impact of the virus was to reduce contact among people and employ social distancing. In California, this was partially implemented by cities, counties, and the Governor's office through "shelter-in-place" orders and related actions (e.g., closure of non-essential businesses), beginning in early and mid-March. The orders results in a 30-60% reduction in peak traffic volumes on California highways and a 75% reduction in vehicle miles traveled on all California roads. While the reduction of traffic was expected, one of the unintended impacts of the orders was the statistically-significant reduction in the number of crashes and related injuries and fatalities. Because we have historical data (back to 2/2015), we were able to confirm that the reduction was not typical of the transition from March to April, or from winter to spring. The Road Ecology Center described these unintended impacts in two special reports on changes in rates of highway crashes (https://roadecology.ucdavis.edu/files/content/projects/COVID_CHIPs_Impacts_updated_415.pdf and https://roadecology.ucdavis.edu/files/content/projects/COVID_CHIPs_Impacts_updated_430.pdf). As part of the analyses, CHIPS data were used to look at the change before and after the orders in the rates of traffic collisions on California highways and certain major roadways patrolled by the CHP (Figure 4).

Because CHIPS data are automatically collected and managed in real-time, we were able to analyze the immediate impacts of the shelter-in-place orders on traffic crashes, including injury and fatal crashes. This rapid-response policy analysis would not have been possible without our automated process of collecting incident data.



Data source: California Highway Incidents Processing System (CHIPS)

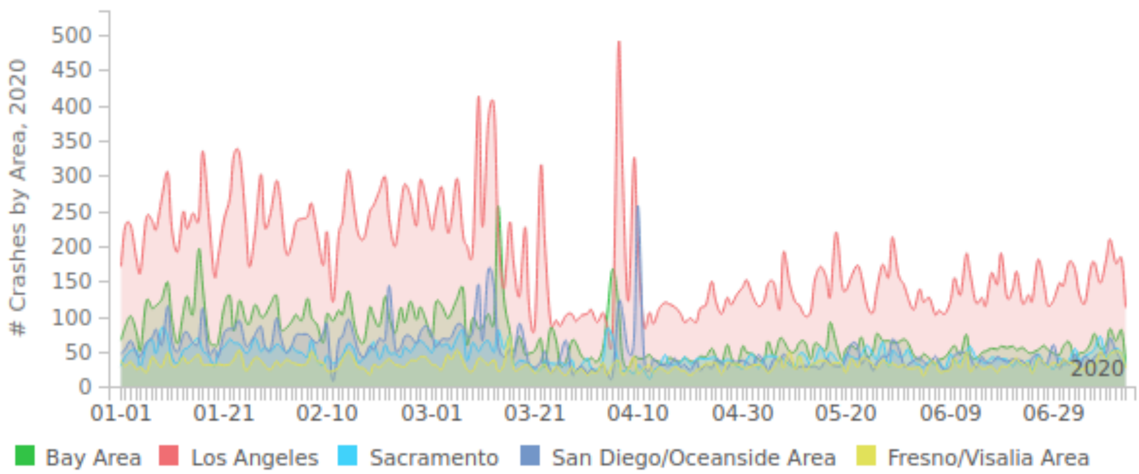


Figure 4. Rates of all crashes and injury/fatal crashes (A) across California and (B) across specific California regions.

Case Study 2: Highway Incidents to Health Outcomes (HIHO)

California does not have a way to track and investigate the arc from roadway crashes to health outcomes for impacted individuals because the crash report is not associated with the individuals after they leave the crash site. While there are data from first responders and ambulance companies which can be associated *post-hoc* with an individual as they enter medical treatment, the details about the accident *do not* accompany an individual's medical report. This prevents an important type of analysis connecting the crash site characteristics (location, climate, roadway conditions) and how those impact the health outcome of the driver and passengers.

Once the CHP officer clears the scene, there is very little information linking the incident details with the subsequent emergency transportation and later health outcomes for the accident victim. One question is how the CHIPS incident records can be integrated with SWITRS, the Fatality Analysis Reporting System (FARS), and individual medical records

Oregon has a novel system, where someone involved in a crash receives a wrist band (called Trauma Bands), which provide a unique identifier that can follow the person from the scene of the incident to and through the emergency department or trauma center. This identifier goes into the incident report as well as the person's health record, so one can look at a person's health outcomes based on the type of conditions surrounding the crash they experienced. Since 2016, Sweden has had a functionally similar system, Swedish Traffic Accident Data Acquisition (STRADA), that includes traffic crash information, including environmental and roadway conditions, connected to the health impacts of the crash. This system allows analysis of completeness of crash reporting. Researchers have used the Swedish Traffic Accident Data Acquisition to compare hospital treatments of crash victims (Held, 2016) and analyze the effectiveness of policies like "Vision Zero" (Varnild et al., 2020). In the latter case, researchers demonstrated crash and injury severity benefits for pedestrians from Vision Zero actions on specific roads. Trauma Bands themselves are relatively inexpensive. However, there are additional personnel training costs, implementation costs, and required updates to existing processes/workflows. Digital systems would need to be updated to include this potentially new field in various databases.

As part of the Stakeholder Meeting (see Including Stakeholders above), we discussed the idea of encoding the unique identifier with information about the incident. One unique identifier could include all of the basic information about the incident in a single string, while also functioning as the unique identifier for the person in the incident. For example, California counties can be represented by a number, say 01 to 58 (number of counties), and this number can be added to the identifier so that the county can easily be determined without knowing any other details. There are ways of encoding this information to be individually-unique without it being personally identifiable. In follow-up discussions, our collaborator at UC Davis Medical Center, Dr. David Shatz, started working—with others active in state policymaking regarding trauma treatment and tracking—on beginning a trauma band system for California.

Case Study 3: California Animal Crash Map

Wildlife-vehicle conflict (WVC) refers to any interaction between wildlife and vehicles/traffic that can have negative impacts for drivers and/or wildlife. This includes animals fleeing from traffic noise/light, drivers swerving around animals on the road surface, and vehicle collisions with animals. WVC is a large and growing concern among Departments of Transportation and the driving public (Bissonette et al. 2008), and it is a conservation concern regarding most animal species (Fahrig and Rytwinski 2009). Loss et al. (2014) estimated that between 89 and 340 million birds may die per year in the US from collisions with vehicles. Predicting and prioritizing places for mitigation of impacts to wildlife and drivers is an important step in reducing WVC. Many Departments of Transportation are trying different methods of reducing WVC, including fencing roadways and providing crossing structures across the right-of-way to allow safe animal passage.

To address concerns about WVC, we added animal-related fields to the CHIPS data model, broadening the scope and value of the data. CHIPS records are queried using terms like "deer," "bear," "elk," etc., and the query results are manually reviewed and updated to include information about animal outcome, driver/passenger outcome, and other details. CHIPS records are also automatically queried using deer-related terms (e.g., "deer," "buck"), while controlling for similar words that are unrelated to animals (e.g., "John Deere"). The resulting records are deer-related and are included in a real-time map of collisions with large animals (Figure 5).

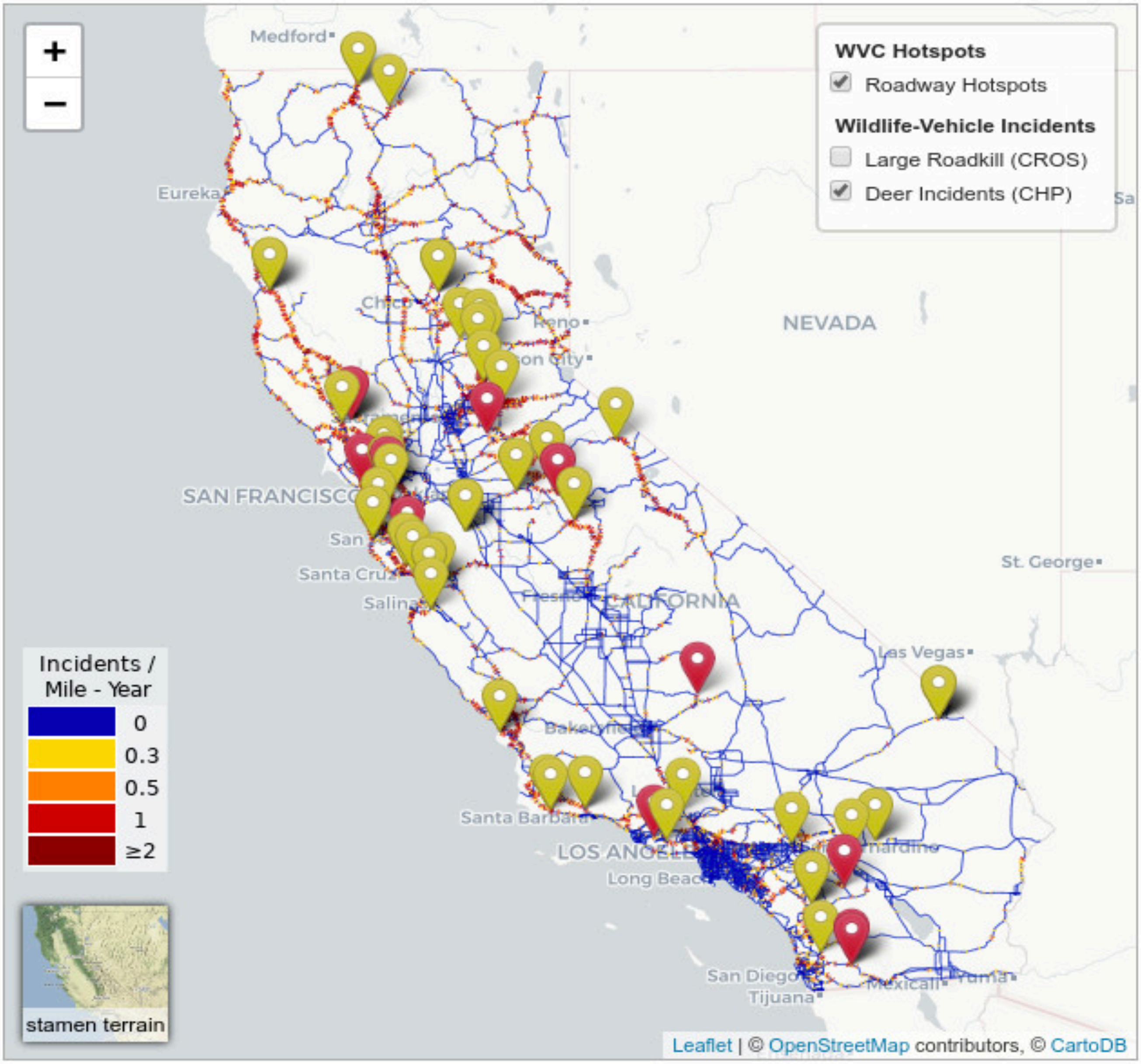


Figure 5. Real-time and recent wildlife-vehicle conflicts. Red points are from CHIPS and occurred in the previous 24 hours, olive points are also from CHIPS and occurred in the previous week, and orange-brown points represent large mammal carcasses from the California Roadkill Observation System in the previous week.

Future Directions and Needs

A critical part of this project was including collaborators from medical and public health domains in discussions of how different crash-tracking systems compare and solutions to California's fragmented system for reporting crashes, including health outcomes. The primary result of the stakeholder meeting was a general agreement that California needs a more integrated system for connecting traffic incidents involving injuries with health outcomes. The first step the group discussed was integrating existing reporting systems that each contain part of the overall arc from incident to outcome. For incidents, this includes CHIPS, PeMS and SWTIRS, all of which provide different benefits. Once emergency medical services (i.e., ambulances) respond, then a separate and more fragmented set of systems comes into play, with few obvious solutions. However, once individuals are received at trauma centers, then they enter a more systematic process of tracking treatments and health outcomes. These different systems potentially provide the raw materials for a complete integrated system of tracking individuals from traffic incidents to health outcomes in California, which is critical for comprehensive and accurate traffic safety analyses. For example, Conderino et al. (2017) demonstrated that probabilistic linkage could be used to connect ~52% of hospital records involving people affected by traffic crashes with the crash incidents themselves. The sections below describe several steps that can contribute to the immediate goal of an integrated system to supply accurate crash incident data to be used in further analysis and linkage to hospital and trauma treatment records.

Step 1: Investigate Linkage Approaches to Connect CHIPS Incidents to SWITRS Records

To improve the spatial and temporal accuracy and completeness of SWITRS (and other) datasets, we propose to identify and associate CHIPS records with their corresponding SWITRS records. We have pilot-tested a crosswalk between CHIPS and SWITRS records for 2/2015 to 8/2018 (the most recent date where SWITRS data were available, as of 2/2020). We have found varying rates of matching records depending on the spatial extent of the search window and categorical-similarities (e.g., injury severity), which can be optimized to identify CHIPS records (and corresponding data) with corresponding SWITRS records.

Step 2: Develop Crosswalk from CHIPS to SWITRS and Trauma-Treatment Datasets

We have developed a possible data model that shows how the SWITRS data could be linked to the CHIPS database (Figure 6). This model uses a "join table" to link the CHIPS ID (cid) with the SWITRS unique identifier (case_id). The join could be made by running a spatio-temporal query (Step 3) or using a probabilistic linkage approach (e.g., as in Conderino et al., 2017), and then using additional fields from both datasets to help determine the right match. The outcome of the integration could be a more spatially-accurate SWITRS record that could be used by the state and traffic safety researchers. We also plan on working with the UC Davis Medical Center to create probabilistic linkages between highway crash incident reports in CHIPS with trauma treatment data in Sacramento region hospitals, for which we have an IRB-approved protocol.

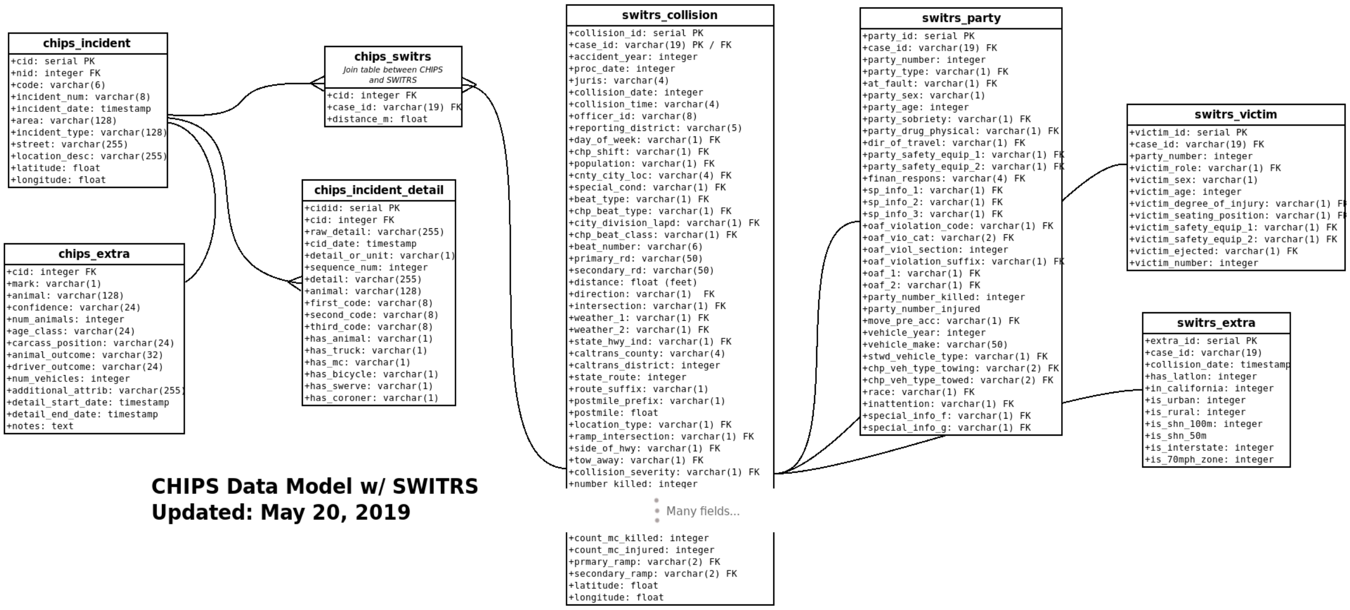


Figure 6. Data model to integrate SWITRS records with CHIPS records and database.

Step 3: Spatio-Temporal Query

Both SWITRS and CHIPS contain geospatial point data in the form of latitude and longitude fields (WGS84 coordinate system). These fields, along with the date and time of the incident, can be used to build a spatio-temporal query to associate the two datasets. The procedure begins with a CHIPS record and searches for candidate SWITRS collision records that are on the same day and one day before or one day after the CHIPS incident date (temporal variation) and within a certain radius set by the user (e.g., 1 km, Figure 7). If there is more than one candidate, then the collision severity, location description, vehicle type, and other relevant fields can be used to select the best candidate. Once the process is completed, a “crosswalk” record would be created in the join table with the best candidate.

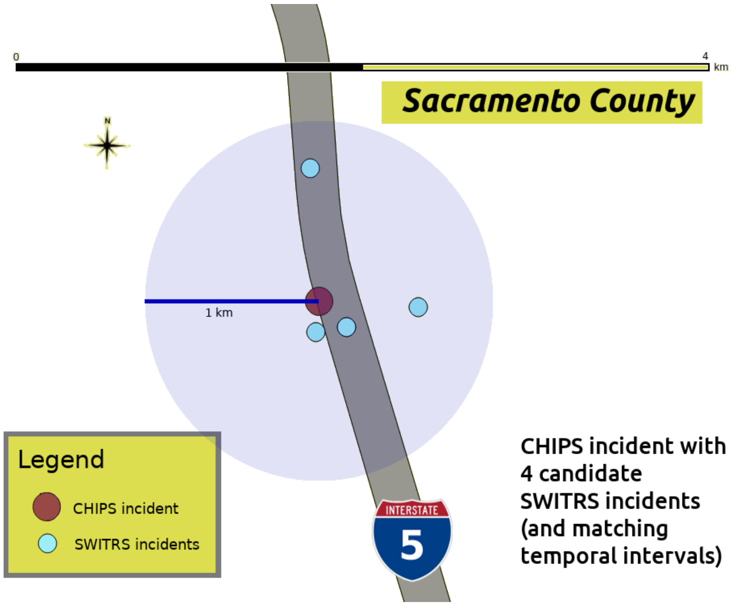


Figure 7. Model of process for associating CHIPS and SWITRS records based on location.

Interactive Data Portal

While the value of CHIPS lies in the data, with over 4 million unique incident records, one useful service for traffic safety research in California will be an interactive data portal where analysts/researchers can go to access these data. We have been providing CHIPS data on an *ad hoc* basis for several years for investigation of bicycle/scooter incidents, speed limits for truck traffic, and wildlife-vehicle conflicts. At the time of writing this report, we also provided crash/injury data to a 7-county consortium (Joint Regional Intelligence Center – JRIC) in Southern California (Los Angeles, Orange, Riverside, San Bernardino, San Luis Obispo, Santa Barbara, and Ventura) to inform their analysis of the impact of COVID-19 related stay-at-home orders on traffic volumes and crashes. This manual provision of data is not sustainable in the long-term and should be replaced by a web-portal to serve the data. The most useful approach will be a mechanism to access the data through structured queries and map-based protocols. To accommodate the general user, these data need an interactive data portal where they can submit ad-hoc queries against these data, so they get back only the type of data they need for their study. This would also require automated processes for continuing to retrieve, store, and manage data, as well as validate data quality and completeness.

One of the major improvements we propose to the existing system will be to have the narrative details included by the CHP officer processed with tools that can help identify and "discover" incidents when a search is performed. We used several tools to optimize queries and provide data suitable for several fields of research—trauma prevention, post-hoc incident analysis, traffic safety analyses. Our current web-database was built as a repository without keyword indexing, record discovery, or the ability for any user to conduct ad hoc queries. A key improvement we propose to the current methodology is the introduction of a pre-processing step before data are stored in a local repository. This will enable us to: 1) ensure we have the most complete information on an incident; 2) translate the descriptive details the officer provides into a series of attributes that can better classify the incident; 3) consolidate CHP incident types into a new vocabulary which better captures the fields of interest that CHIPS data can provide (e.g., "all injury accidents", "natural disaster response").

References

Amin-Naseri M., P. Chakraborty, A. Sharma, S.B. Gilbert, and M. Hong. 2018. Evaluating the Reliability, Coverage, and Added Value of Crowdsourced Traffic Incident Reports from Waze. Transportation Research Record, Vol. 2672(43) 34–43. DOI: 10.1177/0361198118790619

Bejleri, I. and Brown, D. 2014. Stretching Scarce Dollars Through Streamlined Processes and Partnerships: Prototype Implementation for a Unified, Intelligent, and Sustainable Geolocation Process for Roadway Incidents. Proceedings of the Transportation Research Board 93rd Annual Meeting. Location: Washington DC. Date: 2014-1-12 to 2014-1-16

Bissonette, John A., et al. "Assessment of Costs Associated with Deer–Vehicle Collisions: Human Death and Injury, Vehicle Damage, and Deer Loss." Human-Wildlife Conflicts, vol. 2, no. 1, 2008, pp. 17–27. JSTOR, www.jstor.org/stable/24875102.

Chen, L.; Englund, C. Every Second Counts: Integrating Edge Computing and Service Oriented Architecture for Automatic Emergency Management. J. Adv. Transp. 2018, 2018, 7592926.

Cherry, C., A.M. Hezaveh, M. Noltenius, A Khattak, L. Merlin, E. Dumbaugh, D. Ragland, and L Sandt. 2018. Completing the Picture of Traffic Injuries: Understanding Data Needs and Opportunities for Road Safety. A Special Report of the Collaborative Sciences Center for Road Safety, Pp. 150.

Conderino, S., L. Fung, S. Sedlar, J.M. Norton. 2017. Linkage of traffic crash and hospitalization records with limited identifiers for enhanced public health surveillance. Accident Analysis & Prevention, Vol. 101: 117-123

Fahrig, Lenore, and Trina Rytwinski. "Effects of Roads on Animal Abundance: an Empirical Review and Synthesis." Ecology and Society, vol. 14, no. 1, 2009. JSTOR, www.jstor.org/stable/26268057.

Held, F. 2016. Investigation of under-reporting and the consistency of injury severity classifications in Swedish police crash data compared to hospital injury data based on the Swedish Traffic Accident Data Acquisition (STRADA). MSc thesis in Engineering Mathematics and Computational Science

Morris, D., M. Weir. 2016. Geospatially enabled database for analyzing traffic injuries in San Francisco, California. Transportation Research Record Vol. 2595(1): 40-49, https://doi.org/10.3141/2595-05

Santiago-Chaparro, K.R., K. Brose, A. Bill, and D. Noyce. 2016. Improving the content of crash records: The case for a data fusion approach. Proceedings of the Transportation Research Board 95th Annual Meeting. Location: Washington DC, United States. Date: 2016-1-10 to 2016-1-14

Trueblood A.B. , A. Panta, J. Kim, H.-C. Kum, M. Perez, S. Das, and E.M. Shipp. 2019. A semi-automated tool for identifying agricultural roadway crashes in crash narratives. TRAFFIC INJURY PREVENTION. https://doi.org/10.1080/15389588.2019.1599873

Varnild, A., P. Tillgren, and P. Larm. 2020. What types of injuries did seriously injured pedestrians and cyclists receive in a Swedish urban region in the time period 2003-2017 when Vision Zero was implemented? Public Health 181: 59-64. https://doi.org/10.1016/j.puhe.2019.11.019

Yan, L., Y. He, L.. Qin, C. Wu, D. Zhu, and B. Ran. 2017. A Novel Feature Extraction Model for Traffic Injury Severity and Its Application to FARS Data Analysis. Proceedings of the Transportation Research Board 93rd Annual Meeting. Location: Washington DC. Date: 2017-1-8 to 2017-1-12

Young, S.D., W. Wang, B. Chakravarthy. 2019. Crowdsourced traffic data as an emerging tool to monitor car crashes. JAMA Surgery, Vol. 154(8): 777-778.

Appendix A: CHIPS data summaries

CHIPS Data Summary

Number of Incidents (2/2015 – 5/2020): 3,116,394

CHIPS Vocabulary: Area

Area	Count
Altadena	64918
Alturas	2380
Amador	6709
Antelope Valley	19764
Arrowhead	15399
Auburn	27014
Auburn FSP	1263
Bakersfield	49533
Baldwin Park	79727
Barstow	16455
Bishop	5066
Blythe	2669
Bridgeport	3679
Buellton	7088
Buttonwillow	8804
Capistrano	22556
Castro Valley	18461
Central LA	87235
Chico	9861
Clear Lake	11094
Coalinga	8030
Contra Costa	62173
Contra Costa FSP	3214
Crescent City	5280
Dublin	38841
Dublin FSP	1916
East LA	69263
East Sac	16734
El Cajon	63656
El Centro	9779
Fort Tejon	9524
Fresno	62362

Area	Count
Garberville	9033
Gold Run	7466
Golden Gate Dispatch	625
Grass Valley	11898
Hanford	8594
Hayward	29851
Hayward FSP	3370
Hollister Gilroy	23915
Hollister Gilroy FSP	10
Humboldt	16873
Indio	20911
King City	5527
LA	21106
LACC	2232
LAFSP	77824
Los Banos	11864
MY	3876
MYCC	12
MYFSP	17
Madera	15293
Marin	30525
Marin FSP	1282
Mariposa	4231
Merced	20067
Merced Dispatch	148
Modesto	40656
Mojave	9976
Monterey	34155
Monterey Dispatch	2
Moorpark	17663
Morongo Basin	3158
Napa	17485

Area	Count
Napa FSP	146
Needles	2535
Newhall	43041
North Sac	74129
North Sac FSP	2836
Oakhurst	6342
Oakland	56026
Oakland FSP	4176
Oceanside	66886
Orange County FSP	26527
Oroville	9790
Placerville	25889
Placerville FSP	255
Porterville	8606
Quincy	5409
Rancho Cucamonga	54018
Red Bluff	9053
Redding	17497
Redwood City	46031
Redwood City FSP	2916
Riverside	89690
Riverside FSP	17592
SA	7268
SACC	210
SAFSP	906
San Andreas	8871
San Bernardino	68510
San Bernardino FSP	16607
San Diego	130774
San Francisco	35305
San Francisco FSP	2343
San Gorgonio Pass	24769
San Jose	79821
San Jose FSP	6350
San Luis Obispo	17913
Santa Ana	81838
Santa Barbara	18037
Santa Cruz	32945
Santa Fe Springs	71485
Santa Maria	9308
Santa Rosa	46344
Santa Rosa FSP	606

Area	Count
Solano	37707
Solano FSP	926
Sonora	14190
South LA	83853
South Sac	61434
South Sac FSP	7409
Stockton	46635
Stockton FSP	122
Susanville	6489
Temecula	29919
Temecula FSP	573
Templeton	12934
Tracy	12307
Tracy FSP	116
Trinity River	3730
Truckee	9131
Ukiah	14542
Ventura	26618
Victorville	15004
Visalia	28996
West LA	57726
West Valley	59621
Westminster	25058
Williams	3988
Willows	4473
Winterhaven	1840
Woodland	18130
Woodland FSP	754
Yreka	9343
Yuba Sutter	10184

CHIPS Vocabulary: Incident Type

incident_type	count
AMBER Alert	114
Aircraft Emergency	236
Animal Hazard	89873
Assist CT with Maintenance	21916
Assist with Construction	36772
BLUE Alert	9
CLOSURE of a Road	16566
Car Fire	46264
Chain Control	1013
County Roads	803
Defective Traffic Signals	18543
Derailed Train	26
ESCORT for Road Conditions	361
FSP Req Traffic Break	50
Fatality	5999
Foggy Conditions	360
Hazardous Materials Inc	397
Hit and Run No Injuries	246444
Hit and Run w/Injuries	15849
JUMPER	4689
Joint Weather Ops	76
Live or Dead Animal	25258
Mud/Dirt/Rock	5611
Object Flying From Veh	6088

incident_type	count
Provide Traffic Control	10676
Report of Fire	66573
Req CHP Traffic Control	54
Request CalTrans Notify	9941
Road/Weather Conditions	3524
Roadway Flooding	7922
Rock Run	316
SIG Alert	6978
SILVER Alert	750
SNOW Information	293
SPINOUT	4106
Smuggling Fishing Boat	6
Spilled Material Inc	778
Traffic Advisory	1927
Traffic Break	14866
Traffic Hazard	1090207
Trfc Collision-1141 Enrt	291791
Trfc Collision-1141Enrt	9925
Trfc Collision-Major Inj	10913
Trfc Collision-Minor Inj	63624
Trfc Collision-No Inj	633603
Trfc Collision-Unkn Inj	330715
WIND Advisory	1551
Wrong Way Driver	12037