Using multiple data sets for public health tracking of work-related injuries and illnesses in California

Authors:
Lauren Joe[1] MPH, Rachel Roisman[1] MD, MPH, Stella Beckman[1,2] MPH, Martha Jones[3,4] PhD, MPA, John Beckman[5], Matt Frederick[5], Robert Harrison[1] MD, MPH

Corresponding author:
Robert Harrison, California Department of Public Health, Division of Environmental and Occupational Disease Control, 850 Marina Bay Parkway, Building P3, Richmond, CA 94607; email: Robert.Harrison@ucsf.edu

ABSTRACT

Background: Research suggests the U.S. Bureau of Labor Statistics Survey of Occupational Injuries and Illnesses underestimates the magnitude of workplace injuries and illnesses. Enumerating workplace injuries and illnesses may be improved utilizing multiple state-based data sources.
Methods: Using California-based datasets (workers' compensation claims, health care facility data, and physician reports), we enumerated unique cases of amputations and carpal tunnel syndrome (2007-2008), and evaluated the datasets for usefulness in occupational health tracking by performing record linkage across all datasets and calculating match rates between them.
Results: 6,892 amputation and 39,589 CTS cases were identified. Match rates between the datasets ranged from 34.0-45.6% (amputations) and 3.0-43.5% (CTS). Enumerated amputation and CTS cases from state-based sources were about five and ten times greater than the BLS SOII estimates (1,390 and 3,720).
Conclusions: Successful demonstration of this state level approach has broad implications for improving federal and state efforts to track and prevent work-related injuries and illnesses.

Conflict of interest statement:
The authors do not have any conflicts of interest.

KEYWORDS
Surveillance; Carpal Tunnel Syndrome; Amputation; Occupational; Record Linkage

[1]California Department of Public Health, Richmond, CA
[2]School of Public Health, University of California, Berkeley, CA
[3]California Department of Industrial Relations, Oakland, CA
[4]Center for Medicine, Health and Society, Vanderbilt University, Nashville, TN
[5]Impact Assessment Incorporated, La Jolla, CA

INTRODUCTION

Occupational health surveillance, also referred to as tracking, relies on worker injury and illness reporting, and surveillance efforts are a necessary foundation for addressing workplace hazards through targeted prevention strategies. The identification of occupational injury and illness trends is useful for prioritization of occupational risks and follow-up prevention and intervention strategies. As previously demonstrated in the literature, workplace injuries and illnesses are not accurately identified for a variety of reasons, including harassment, fear of employer retaliation, ignorance, lack of training in occupational health among health care providers, and administrative barriers (Azaroff et al. 2002; Probst and Estrada 2010). Despite these limitations, valuable information can be garnered from existing tracking systems, and understanding how these systems differ in their ability to capture the totality of worker injuries and illnesses enables us to better utilize them to protect workers and prevent workplace injuries and illnesses.

At the national level, the U.S. Bureau of Labor Statistics (BLS) conducts the Census of Fatal Occupational Injuries (CFOI) and the Survey of Occupational Injury and Illness (SOII). The CFOI has been counting every work-related fatal injury occurring in the U.S. since 1992 and is a comprehensive and valuable tool for collecting detailed data on workplace deaths. Although death is the most severe outcome of occupational injuries, occupational morbidity is just as important to track, as injuries and illnesses can be devastating and result in lifelong disability. Reducing causes of morbidity may also prevent death, as morbidity risks may have the potential to escalate into life-threatening risks when left unaddressed. The SOII has been the basis of epidemiologic surveillance of workplace injuries and illnesses since 1972. It differs from the CFOI in that the SOII is not a census, but based on a survey method that provides estimates of injuries and illnesses (BLS 2008). Many attributes of the SOII system make it ideal for tracking workplace injuries and illnesses. For example, the comprehensive nature of reporting and sampling characteristics can generate state-based data and annual rates that can be used to evaluate the impact of interventions over time. However, recent studies have suggested that the SOII may undercount injuries and illnesses and may benefit from the additional ascertainment of cases using state-based data sources (Leigh et al. 2004; Rosenman at al. 2006; Boden and Ozonoff 2008; Boden and Ozonoff 2008a).

State-based data sources that are not available at the national level can be used to target specific cases and/or work sites for investigations, thereby coordinating efforts at the individual worker and work site levels to reduce the burden of workplace injuries, diseases, and deaths. One example is the California Occupational Pesticide Illness Prevention Program, which utilizes reported cases of acute pesticide illness from various sources to identify the specific occupations and types of pesticides that are related to health problems in workers (CDPH OHB 2009). With funding from the National Institute for Occupational Safety and Health (NIOSH) and the California Department of Public Health (CDPH), this program collects and examines various reports to learn more about occupational pesticide poisoning and how to prevent toxic exposures. This system allows for timely identification of pesticide poisoning outbreaks, which are investigated to provide assistance to employees and their employers, and to develop educational materials and recommendations to prevent future outbreaks of similar nature from occurring in the future (CDC 2011).

In addition to identifying instances of acute injury and illness, occupational health surveillance data can be used to strengthen ongoing public health action that has already begun through the efforts of workers themselves. Using surveillance data to characterize the burden of specific occupational injuries and illnesses can provide the scientific evidence needed to propose policy change and long-term solutions. With the adoption of electronic data systems for hospital discharge, emergency department, ambulatory surgery, and workers' compensation claims, we have an opportunity to improve the ability of state and federal agencies to perform coordinated and timely surveillance that can more closely estimate the true nature and extent of workplace morbidity and mortality. Whether the data are utilized to provide routine injury-specific surveillance, to track emerging and acute occupational injuries, or to provide scientific evidence to strengthen worker-initiated public health efforts, it is important to understand the various data systems available so that they can be harnessed most effectively.

In order to demonstrate the utility and feasibility of using multiple state-based data sets for injury-specific occupational health tracking, the CDPH Occupational Health Branch (OHB) enumerated cases of work-related amputations and carpal tunnel syndrome (CTS) that occurred in California in 2007 and 2008. We focused on three datasets that are available for tracking work-related illness and injury in California: the Workers' Compensation Information System (WCIS), Doctors' First Reports of Occupational Injury and Illness (DFRs), and health care facility data collected by the California Office of Statewide Health Planning and Development (OSHPD), which includes ambulatory surgery, inpatient discharge, and emergency department visits. We also utilized a fourth dataset that includes cases from the BLS SOII (2007 and 2008) that were provided to CDPH under a cooperative agreement with the BLS for this analysis and are not typically available for routine surveillance at the state level. The overall purpose of this study was to examine these various sources of work-related injury and illness reporting to improve our ability to perform occupational public health prevention and intervention activities.

MATERIALS AND METHODS

Data Sources

The four sources of data utilized for this analysis have varying administrative purposes and data elements. A summary of case inclusion criteria from the four data sources is presented in Table 1.

Bureau of Labor Statistics (BLS) Survey of Occupational Illness and Injuries (SOII)

The SOII is an annual survey of a sample of workplace establishments utilizing data collected on Occupational Safety and Health Administration (OSHA) injury and illness logs (known as OSHA 300 Logs) maintained by employers. The SOII uses a survey design to estimate the number and frequency of work-related injury and illness in the U.S. and participating states, and includes detailed data on industry and the nature and circumstances of illness or injury. In addition to these data, the SOII collects descriptive case information, including the demographic characteristics of the injured and ill workers who require at least one day of recuperation away from work. Under a cooperative agreement with the BLS to conduct a pilot study to enumerate cases of amputations and CTS for the years 2007 and 2008, BLS provided the requested SOII microdata to our research team. Each case in the SOII dataset was assigned a code indicating the specific type of injury or illness based on the Occupational Injury and Illness Classification System (OIICS, BLS 2008). Amputation cases were extracted from the SOII dataset using the following OIICS nature of injury codes: "0310," "0311," or "0319." CTS cases were extracted using the OIICS nature of injury code: "1241." The SOII estimated a total of 1,390 amputation cases of amputation and 3,720 cases of CTS in California for the years 2007 and 2008.

California Division of Workers' Compensation (DWC) Workers' Compensation Information System (WCIS)

WCIS has been collecting workers' compensation data in electronic format since March 2000. Claims administrators must submit electronic First Reports of Occupational Injury (FROI) to the California Department of Industrial Relations, Division of Workers' Compensation (DWC), within five working days after knowledge of the injury or illness. Though WCIS is an administrative database, the data elements it contains make it a valuable tool for occupational health surveillance purposes, including narrative text describing the injury (Sorock et al. 1997). Claims must be reported to WCIS if a claims administrator receives any of the following: Employer's First Report or Doctors' First Report of Occupational Injury or Illness, an application for adjudication, or any indication that an injury requiring medical treatment by a physician occurred. Self-employed individuals are not required to report to WCIS.

CDPH accesses the WCIS database by requesting data from the DWC based on various predefined criteria. In this study, WCIS extraction criteria were based on case definitions for amputations and CTS developed in conjunction with other collaborators (BLS, Boston University, Washington State Department of Labor and Industries, Massachusetts Department of Public Health). For claims identified as potential amputations or CTS with an injury date in 2007-2008, the dataset used in this study was comprised of extracts from WCIS containing claims data on injury type, employee name, employer name, and benefit payments, and medical billing data on clinical procedure and diagnosis codes. At the time of analysis, 2008 was the most recent year available with complete WCIS claims data. Our original amputation extract from the WCIS database included all claims with "amputation" in the nature of injury field, with amputation-related keywords in the injury description field, or with appropriate diagnosis or procedure codes in the medical billing data. Our original CTS extract from the WCIS database included all claims with "carpal tunnel syndrome" in the nature of injury field, CTS-indicating or -related keywords in the injury description field, or appropriate diagnosis or procedure codes in the medical billing data (**Supplemental Material A**).

Manual review of a sample of amputation and CTS claims revealed that the original extracts contained some claims that were not amputations or CTS. As a result, detailed case classification schemes were developed using a combination of values in the following fields: diagnosis, procedure, nature of injury, part of body, cause of injury, and injury description. The final case classification scheme for amputations (classified as probable or uncertain cases) and CTS (classified as probable, possible, or uncertain cases) are detailed in **Supplemental Material B**. Medical records for a sample of amputation and CTS claims were used to validate the case classification schemes. Fifty-three amputation and 59 CTS medical records were reviewed independently by two physicians, and positive and negative predictive values (PPV, NPV) were calculated. NPVs for amputations and CTS were greater than 0.5, thus cases classified as "uncertain" based on the case classification schemes were removed (1,508, or 22.8%, of the amputation cases, and 12,106, or 29.3%, of the CTS cases) from the record linkage analysis to reduce chances of misclassification. (Joe et al. 2012; Roisman et al. 2013)

Health Care Facility Data

Hospital Discharge (HD), Emergency Department (ED), and Ambulatory Surgery (AS) data are collected by the California OSHPD from all licensed health care facilities in California, approximately 5,000. Data related to financial performance, utilization, patient characteristics, and services provided are publically available, and are often utilized by health care facilities to monitor patient outcomes and hospital performance. Under a Data Use Agreement, the CDPH Center for Health Statistics provided our research team with more detailed data, including ICD-

9 codes, social security numbers (SSN), dates of birth, and dates of service. Each dataset (HD, ED, and AS) has a primary diagnosis and primary procedure field, as well as up to ten additional diagnosis and procedure fields. Due to concerns about misclassification, we limited our analysis to cases for which amputation or CTS was the primary diagnosis or procedure code, using the same ICD-9 codes utilized for the extraction of WCIS claims (**Supplemental Material A**). Only work-related cases were included in the analysis and were determined by the designation of workers' compensation as the payer or by a "place of occurrence" code consistent with a workplace.

The OSHPD dataset contains social security numbers (SSN) but no names, making it challenging to match to other datasets (e.g., SOII has names but no SSN). Therefore, we utilized a health care data service (SearchAmerica) to obtain first and last names for the SSNs in our OSHPD dataset. The service identified names for 96.2% of the OSHPD cases, which we subsequently used for matching with the other data sets (see below).

Doctors' First Reports of Occupational Injury or Illness (DFR)

DFRs have been a reporting source for California work-related injury and illness data since 1949, and have been used routinely by CDPH under numerous CDC/NIOSH surveillance cooperative agreements since 1987. DFRs must be completed within five days by all physicians in California who suspect work-related injury or illness. They contain detailed case and employer information, and are submitted to the workers' compensation insurance carrier (or administrator if self-insured), who then must forward the DFRs to the California Department of Industrial Relations (DIR). CDPH obtains the DFRs under a Memorandum of Understanding with DIR. Approximately 600,000 DFRs are received annually. DFRs are reviewed manually, sorted into selected categories for data analysis and follow-up, and then the remaining DFRs are archived. The CTS DFR cases were available for the present analysis as they had been collected for 2007 and 2008 as part of CDPH's ongoing surveillance activities. Amputation cases had not been collected and were not available for analysis. All CTS DFR cases were classified into 4 categories (Definite, Probable, Possible, and Uncertain) based on criteria previously developed by CDPH OHB as part of an earlier CTS surveillance project (**Supplemental Material C**).

Data Linkage

A summary of the variables used for record linkage across datasets is provided in Table 1. We linked cases across datasets in order to identify cases that appeared in more than one dataset using an iterative process with record linkage software (Jurczyk et al. 2008; Jurczyk et al. 2008a). We conducted deterministic record linkage based on exact SSN when possible, and probabilistic record linkage based on criteria when a deterministic match using SSN was not successful (Mason and Tu 2008; Meray et al. 2007). The linkage process for each dataset was iterative such that remaining unmatched records were matched again in order to maximize the number of matches identified (**Supplemental Material D**). After linkage was performed to obtain raw matches, matches were de-duplicated and refined (described below) to restrict the linkages to one-to-one matched pairs for the purpose of enumerating unique cases of amputation and CTS across the datasets.

For each series of record linkage, many different cases of amputation or CTS in a dataset matched to a single case in the other dataset. For example, in the WCIS-OSHPD linkage series, multiple OSHPD cases matched to a single WCIS case (resulting in duplicate OSHPD cases), and multiple WCIS cases matched to a single OSHPD case (resulting in duplicate WCIS cases). In order to enumerate the unique number of cases across the different datasets, these duplicates were removed to obtain a final list of one-to-one matches. This de-duplication process differed with each linkage series based on the variables available in each dataset (Supplemental Material D details this stepwise process for each linkage series). The most common variables utilized for de-duplication were dates of injury (SOII, WCIS, DFRs), dates of admittance or service (OSHPD), and case classification (WCIS). Linkages were also refined so that the matches retained were limited to amputation or CTS cases, resulting in a final list of one-to-one, unique matches for each linkage series that could be used for enumeration.

Calculating Match Rates and Case Enumeration

Cases in WCIS and OSHPD that remained unmatched after performing record linkage to all datasets were de-duplicated based on exact SSN when possible, or first name, last name, and birthdate. Match rates between datasets were calculated by dividing the number of unique matches by the total number of de-duplicated cases in each dataset. Cases were enumerated by counting the number of unique amputations and CTS that appeared in the WCIS, DFR, and OSHPD datasets. We included both lost-time and non-lost-time cases from the WCIS, DFR, and OSHPD datasets. For purposes of enumeration, we also included cases involving mining, railroad and water transportation, temporary employment, membership organizations and small agricultural establishments, which are excluded from the SOII. As SOII represents a sample of cases only, these were not included in the final enumeration.

This study received Common Rule approval (Code of Federal Regulations 45 46.111) by the State of California, Health and Human Services Agency, Committee for the Protection of Human Subjects.

RESULTS

Data Linkage

BLS Survey of Occupational Illness and Injuries (SOII)

In total, 65.9%of SOII amputation and 60.4% of CTS cases were linked to the WCIS data set (Table 2). Thus, approximately one-third of amputation and CTS cases from the OSHA 300 Logs could not be found in our workers' compensation database. As the SOII cases in our analysis represent work-related lost-time injuries, we expect these workers to have filed workers' compensation claims for medical treatment and/or lost work time. We would particularly expect that a workers compensation claim would be filed for acute traumatic amputations where the connection with work is usually obvious.

Only 29.9% of SOII amputation and 27.0% of SOII CTS cases were linked to the OSHPD data set (Table 2). This low match rate is to be expected, as most SOII cases of amputation or CTS may not be treated in an emergency department, or require inpatient hospitalization or surgery. In contrast, only a small fraction (3.7%) of SOII CTS cases could be linked to the DFR data set. We expect that most SOII-eligible CTS cases would seek physician care and a DFR should be filed under existing California regulations. Many physicians may fail to recognize work-related injuries and illnesses and/or submit a DFR, and some insurers may not send the DFRs to DIR as required.

Overall, 70.5% of all SOII amputation cases and 66.9% of all SOII CTS cases were linked to at least one other dataset. Thus, about one in three cases of amputation and CTS reported by employers on their OSHA 300 Logs were not recorded elsewhere – including physician reports or workers' compensation claims.

Workers' Compensation Information System (WCIS)

Only a small percentage of WCIS claims were linked to SOII cases (2.9% of amputation claims and 0.9% of CTS claims, Table 2). This is expected, as the SOII is a relatively small sample of all cases, and the WCIS is a statewide system that is designed to capture all claims. We found 34.0% of WCIS amputation claims and 10.5% of WCIS CTS claims were linked to an OSHPD case. This finding is to be expected, as most work-related injuries are not treated in an emergency room, admitted to a hospital, or require surgery. A large majority of work-related cases of amputations (65.9%) and CTS (86.5%) were found only in WCIS (Figures 1 and 2).

Health Care Facility Data (OSHPD)

Only 45.6% of OSHPD amputation cases and 24.5% of OSHPD CTS cases were linked to a WCIS claim (Table 2). All work-related cases requiring ambulatory surgery should have a workers' compensation claim, as health care providers typically must obtain authorization from the workers' compensation insurance carrier prior to surgery. Overall, 54.4% of amputation cases and 73.8% of CTS cases were found only in OSHPD (Figures 1 and 2).

Doctors' First Reports of Occupational Injury or Illness (DFR)

Only 43.5% of DFR CTS cases were linked to WCIS claims (Table 2). After recognizing an injury or illness as work-related, California regulation requires that physicians submit a DFR to the workers' compensation insurance carrier. The DFR then becomes a basis for the insurance carrier to submit an electronic FROI to DWC's WCIS database. Therefore, we expect that all DFR cases, absent an administrative problem in claims management or an immediate determination that the injury was not work-related, would be matched to a claim in the WCIS database. Overall, 46.0% of CTS cases were uniquely identified from DFRs (Figure 2).

Enumeration

For the years 2007 and 2008, a total of 6,892 amputation cases were identified from WCIS and OSPHD (Figure 1), and 39,589 CTS cases were identified from WCIS, OSPHD and DFRs (Figure 2). Of the 6,892 amputation cases, almost half (3,216 or 46.9%) were found only in the WCIS. Of the 39,589 CTS cases, almost two-thirds (25,193 or 63.6%) were found only in the WCIS. Amputation and CTS cases identified from California's state-based data systems (6,895 and 39,589) were about five and ten times greater, respectively, than the number estimated from the BLS SOII (1,390 and 3,720) for 2007 and 2008.

DISCUSSION

There are several unexpected findings from the linkage of work-related amputations and CTS in California that deserve discussion. First, more than one-third of SOII cases of CTS and amputations could not be found in WCIS as a workers' compensation claim. SOII cases represent injured employees who have notified their employer of their injury. If these employees had a work-related injury requiring medical care beyond first aid, a claim should have been filed with the employer's workers' compensation insurance carrier to pay for medical care and indemnity payments of temporary or permanent disability where appropriate. CTS is often a cumulative injury that occurs over months to years, and it is conceivable that some SOII cases filed workers' compensation claims in other years. However, amputations are usually immediate and fairly obvious injuries, and we expect that a workers' compensation claim should be filed for these cases within weeks of the incident.

There are several reasons why SOII cases may not be found as workers' compensation claims. Workers' compensation claims may not be filed if the treating physician does not recognize the injury as work-related (e.g., does not submit a DFR), if the employer does not notify the workers' compensation insurance carrier of the work-related injury, or if the employee procures medical treatment outside of the workers' compensation system. In some cases, employers may pay medical providers directly for the injured workers' care. A recent survey of Log 300 reporting in Washington State suggests that some employers simply list all workers' compensation claims on their OSHA 300 Log, others carefully follow the OSHA recordkeeping guidelines, and others have difficulties interpreting the applicable regulations and how they apply (particularly among small employers) (Wuellner and Bonauto 2013). For example, some employers may list a case on the OSHA Log 300 to meet the recordkeeping requirements, but then make a separate administrative decision about whether to notify their workers' compensation insurer about an injury. The extent to which these issues arise in OSHA recordkeeping and workers' compensation claims reporting in California is not known, and would require follow-up surveys of both employers and employees to uncover the actual chain of events that finally records a work-related injury or illness or not.

Based on the OSHPD linkage results, one of two cases of work-related amputation and three of four cases of work-related CTS that required an emergency room visit or surgery could not be found in the workers' compensation claims database. Injured workers with amputations or with CTS that need surgery are likely suffering from more severe injuries, and therefore we expect that the majority of these cases would have filed a workers' compensation claim. In addition, health care providers usually require insurance authorization prior to proceeding with a surgical procedure. Informal telephone interviews with several ambulatory surgery center billing departments confirmed that prior insurance authorization is indeed obtained. There are several possible reasons why we were not able to find OSHPD cases in WCIS. For CTS cases requiring surgery, workers' compensation claims may have been filed in prior years when the case was initially identified as work-related. Indeed, when 2007-2008 ambulatory surgery CTS cases were matched to prior years, approximately 66% of cases were matched in WCIS (results not shown). The ambulatory surgery center may initially expect payment for the surgical procedure from the workers' compensation insurance company, but after additional review these cases may be determined not to be work-related and final payment for the procedure is made from another source. In-depth surveys of workers' compensation provider billing practices, administrative procedures and workers' compensation claims practices are needed to determine the extent to which these issues explain the large discrepancy in case identification.

Third, three of five CTS cases reported by physicians on the DFR could not be found in the workers' compensation claims database. Under California law, all physicians are required to submit a DFR to the workers' compensation insurance carrier for a suspected work-related injury and illness. The DFR is one basis for creating and submitting a FROI to the WCIS. Therefore, we expect that every CTS DFR should be matched to a workers' compensation claim. The cumulative nature of CTS suggests that some workers' compensation claims may have been filed in years prior or subsequent to the DFR. It is possible that some physicians file DFRs for work-related injuries (including CTS) but a workers' compensation claim is never established, or there is inconsistent interpretation by insurance carriers of reporting requirements. Since DFRs are required to be filed for any suspected work-related injury or illness, it is possible for there to be no associated workers' compensation claim file if the injury or illness was later deemed to be not work-related. Additional studies are needed to track the "life of an injury" from physician reporting to detection in the workers' compensation claims database.

The enumeration of all work-related amputation and CTS cases suggests that the number of these injuries that occur annually in California is much greater than those estimated by the BLS SOII. The greatest proportion of all cases was found only in the workers' compensation database, which includes cases with and without lost work time (days away from work). Likewise, many cases were found only in the OSHPD database, which includes hospital discharge, emergency department, and ambulatory surgery data, representing cases that are more severe in nature. A subset of cases is reported only by physicians on the DFR, perhaps representing suspected injuries or those requiring only medical treatment. Although not included in the overall enumeration, there were SOII cases that were not detected elsewhere as well.

These findings suggest that there are numerous pathways by which work-related injuries may be reported, each of which adds to the overall estimated magnitude of work-related injury and illness in California. After an employee is injured at work, workers' compensation regulations trigger numerous administrative requirements involving many individuals including the injured worker, the worker's supervisor, the employer's personnel or human resources manager, the insurance claims administrator, and the health care provider.

With approximately 1.3 million workplaces in 2008 (California Employment Development Department), over 200 insurers that wrote workers compensation premiums in 2012 (California Department of Insurance 2013), and 136,000 health care providers in California (Medical Board of California, Osteopathic Medical Board of California, California Board of Registered Nurses, and California Physician Assistant Committee), it is not surprising that multiple data sources are needed to ascertain the burden of work-related injury and illness.

Other studies of work-related injuries and illnesses in California have suggested a significant BLS SOII undercount ranging from 25 to 40 percent (Boden and Ozonoff 2008; Boden and Ozonoff 2008a). The undercount estimates in our analyses may differ from these previous studies due to the addition of other data sources in addition to workers' compensation claims, specific endpoints (amputations and CTS), and use of a different WCIS extraction criteria. While Boden and Ozonoff utilized capture-recapture methods to estimate the undercount of WCIS and the BLS SOII, our analysis focused on enumerating endpoint-specific cases and evaluating the utility of state-based data sources for tracking occupational injuries and illnesses. From the public health perspective, the BLS SOII undercount reflects an employer-based system of reporting that provides a partial description of the actual burden of work-related injury and disability in the U.S. Many authors have described barriers to reporting of work-related injuries and illnesses by workers themselves, and these cases will never be reported anywhere (Azaroff et al. 2002) or detected by our public health system. Indeed, in California there is no systematic collection of worker-reported injury or illnesses directly to the California Department of Public Health or other state regulatory authorities.

There are a number of limitations in our analyses. First, because of the large number of cases reported in California, we adopted strict matching criteria and could not review individual matches by hand. Therefore, there may have been matches that we missed. Due to the large number of cases and frequency of "close" matches on SSN or birth date, it was not feasible to include matching criteria that accounted for single digit differences or transpositions in these data. The numbers of matched pairs with one and two digit differences in SSN when matching the full WCIS and OSHPD data files would be significantly larger and prohibit manual review. Pairs of dates of birth with digit differences are similarly numerous.

Second, our analyses included both accepted and denied workers' compensation claims. The overall claims denial rate in WCIS was approximately 8% in 2007-2008 (California Department of Industrial Relations 2013). It is not known to what extent claim denial may influence OSHA 300 Log reporting and subsequent inclusion in the SOII survey for a sampled employer.

Third, the case definitions for CTS and amputation may be subject to misclassification due to inaccurate physician diagnosis, administrative claims processing, or other unknown factors. Although we reviewed records and confirmed the medical diagnoses in a sample of CTS and amputation cases, we did not ascertain the extent to which other diagnoses (such as tendinitis or avulsion) might be "true" cases of CTS or amputation. Matching was performed between datasets using our *a priori* case definition, and the extent to which matching results may vary using related diagnoses is unknown.

Finally, a number of challenges are intrinsic to each data set that we utilized. The WCIS industry information is incomplete and/or has inconsistent coding; for example, some employers provide the same corporate mailing address on every claim, while other employers provide the physical address of the establishment where the injured employer works. On different claims, a single employer might use different company names, different Federal Employer Identification Numbers, or different industry or class codes, even sometimes for the same employee at the same location. It is thus sometimes difficult to determine which claims come from the same employers. Occupation coding is not feasible as a unique "class code" is assigned for administrative purposes. The sheer size of the WCIS dataset means that individual claim review is not feasible for all records.

OSHPD data reporting requirements do not include employer information or a work-related variable other than expected payer (where workers' compensation is one of ten different possible categories). This limits our ability to identify work-related injuries among workers who are uninsured and to characterize disparities that exist in these injuries based on workers' compensation coverage status (Berdahl and Zodet 2010; Nicholson et al. 2008). Use of E-coding is not feasible for identifying work-relatedness due to the limited use of this field by health care providers, and the date of injury is not a variable included in the OSHPD dataset so was not available for matching to SOII and WCIS. Furthermore, due to changes in California licensing requirements, beginning in 2008, physician-owned ASCs are not required to report to OSHPD.

The DFR is a paper-based form completed by physicians, and some information may be unavailable due to illegible handwriting or incomplete fields. There are no automatic quality checks on the DFR, thus resulting in inaccurate data completion or misdiagnosis.

CONCLUSION

An ongoing system using multiple data sources can add to federal and state efforts to prevent work-related injuries and illnesses. The total number of amputations and CTS in California is significantly greater than the BLS SOII estimates, suggesting that a multisource surveillance system is a valuable adjunct to employer-based reporting. From the public health perspective, BLS SOII data are a valuable tool that can be used to analyze trends and compare relative risk of injuries and illness across industries and occupations. The BLS SOII system was not designed for use by public health departments for case identification, disease outbreak detection, or work site investigations. In contrast, the workers' compensation claims system may be used by public health agencies for case identification, leading to workplace interventions that can prevent additional cases. Physician reports are a valuable and timely source of clinical information about both individual and multiple cases from a work site, leading to investigations of injuries and disease outbreaks. Hospital discharge, emergency department, and ambulatory surgery records may identify cases of severe work-related injury or disease that require public health action. Developing and maintaining occupational epidemiology as a core component of public health capacity at the state level is critical to the use of these data sources in the ongoing prevention of work-related injuries and illnesses.

ACKNOWLEDGEMENTS

This project was supported by a U.S. Bureau of Labor Statistics grant, "Enumerating Workplace Injuries and Illnesses with Multiple Data Sources" (RFO BLS-0910-002). We wish to acknowledge other state and federal collaborators: Darrin Adams, Les Boden, Dave Bonauto, Letitia Davis, Kathleen Grattan, John Ruser, John Russell, Eric Sygnatur, SangWoo Tak, and Sara Wuellner.

This work was supported in part by an appointment to the Applied Epidemiology Fellowship Program administered by the Council of State and Territorial Epidemiologists (CSTE) and funded by the Centers for Disease Control and Prevention (CDC) Cooperative Agreement Number 5U38HM000414-5.

REFERENCES

Azaroff LS, Levenstein C, Wegman DH. September 2002. Occupational injury and illness surveillance: Conceptual filters explain underreporting. American Journal of Public Health: 1421–1429.

Berdahl TA and Zodet M. 2010. Medical care utilization for work-related injuries in the United States 2002-2006. Medical Care. 48: 645-651.

Boden LI and Ozonoff A. June 2008. Capture-Recapture Estimates of Nonfatal Workplace Injuries and Illnesses. Annals of Epidemiology: 500-506.

Boden LI and Ozonoff A. August 2008a. Reporting Workers' Compensation Injuries in California: How Many are Missed? The California Commission on Health and Safety and Workers' Compensation.

Bureau of Labor Statistics. September 2008. Occupational Safety and Health Statistics. In: BLS Handbook of Methods: United States Department of Labor, Bureau of Labor Statistics webpage. Retrieved from: https://www.bls.gov/opub/hom/homch9.htm. Date accessed: 12/30/2010.

California Board of Registered Nurses. September 2013. Monthly Statistics webpage. Retrieved from: http://www.rn.ca.gov/about_us/stats.shtml. Date accessed: 2/6/2014

California Department of Industrial Relations. June 2013. In: Table 9. California Workers' Compensation Claims: FROI and SROI Data Summary, by Year of Injury 2000-2012, California Department of Industrial Relations Worker's Compensation Information System Database Tables webpage. Retrieved from: http://www.dir.ca.gov/dwc/wcis/WCIS_tables/Table9.xls. Date accessed: 2/4/2014.

California Department of Insurance, Rate Specialist Bureau. June 2013. In: California Life and Annuity Insurance Industry 2012 Market Share Report, California Department of Insurance Market Share Reports webpage. Retrieved from: http://www.insurance.ca.gov/0400-news/0200-studies-reports/0100-market-share/2012/upload/LifeMktShr2012.pdf. Date accessed: 2/7/2014.

California Department of Public Health, Occupational Health Branch (CDPH OHB). August 2009. In: Occupational Pesticide Illness in California, 1998-2007, CDPH OHB Tracking Pesticide-Related Illness and Injury – Publications webpage. Retrieved from: http://www.cdph.ca.gov/programs/ohsep/Documents/pestillness.pdf. Date accessed: 5/29/2013.

California Employment Development Department, Quarterly Census of Employment and Wages. In: Size of Business Data – 2003-present, California Employment Development

Department webpage. Retrieved from: http://www.labormarketinfo.edd.ca.gov/Content.asp?pageid=138. Date accessed: 2/5/2014.

California Physician Assistant Committee. In: PAC Annual Report, 2011-2012, California Physician Assistant Committee Forms and Publications webpage. Retrieved from: http://www.pac.ca.gov/forms_pubs/annual_report_1112.pdf. Date accessed: 2/6/2014.

Centers for Disease Control and Prevention (CDC). October 2011. Acute Illness and Injury from Swimming Pool Disinfectants and Other Chemicals --- United States, 2002—2008. MMWR. Morbidity and Mortality Weekly Reports. 60:1343-1347.

Friedman L, Krupczak C, Brandt-Rauf S, Forst L. 2012. Occupational amputations in Illinois 2000–2007: BLS vs. data linkage of trauma registry, hospital discharge, workers' compensation databases and OSHA citations. Injury. 44:667-73.

Joe L, Beckman S, Roisman R, Frederick M, Beckman J, Rempel D, Jones M, Harrison R. June 2012. Using an Administrative Workers' Compensation Claims Database for Occupational Health Surveillance in California: Validation of a Case Classification Scheme for Carpal Tunnel Syndrome (Abstract 1326952). Council of State and Territorial Epidemiologists - Annual Conference.

Jurczyk P, Lu JJ, Xiong L, Cragan JD, Correa A. 2008. FRIL: A tool for comparative record linkage. AMIA Annual Symposium Proceedings: 440-444.

Jurczyk P, Lu JJ, Xiong L, Cragan JD, Correa A. 2008a. Fine-grained record integration and linkage tool. Birth Defects Research Part A: Clinical and Molecular Teratology. 82: 822-829.

Leigh JP, Marcin JP, Miller TR. 2004. An estimate of the U.S. government's undercount of non-fatal occupational injuries. Journal of Occupational and Environmental Medicine. 46: 10–18.

Mason CA and Tu S. 2008. Data linkage using probabilistic decision rules: a primer. Birth Defects Research Part A: Clinical and Molecular Teratology. 82: 812-821.

Medical Board of California. Physician and Surgeon License by County, 2011-2012, webpage. Retrieved from: http://mbc.ca.gov/About_Us/Statistics/Licenses_by_County.aspx. Date accessed: 2/6/2014.

Meray N, Reitsma JB, Ravelli AC, Bonsel GJ. 2007. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. Journal of Clinical Epidemiology. 60: 883-891.

Mustard CA, Chambers A, McLeod C, Bielecky A, Smith PM. May 2012. Comparison of data sources for the surveillance of work injury. Occup Environ Med. 69: 317-324.

Nicholson V., Bunn TL, Costich JF. Jun 2008. Disparities in work-related injuries associated with worker compensation coverage status. Am J Ind Med. 51: 393-398.

Osteopathic Medical Board of California. In: OMBC 2012 Oversight Report, Osteopathic Medical Board of California Forms and Publications webpage. Retrieved from: http://www.ombc.ca.gov/forms_pubs/sunset_2012.pdf. Date accessed: 2/6/2014.

Probst TM and Estrada AX. 2010. Accident under-reporting among employees: Testing the moderating influence of psychological safety climate and supervisor enforcement of safety practices. Accident Analysis and Prevention. 42: 1438-1444.

Roisman R, Joe L, Frederick M, Beckman S, Beckman J, Jones M, Harrison R. May 2013. Using an Administrative Workers' Compensation Claims Database for Occupational Health Surveillance in California: Validation of a Case Classification Scheme for Amputations. Use of Workers' Compensation Data for Occupational Safety and Health: Proceedings from June 2012 Workshop: 121-125.

Rosenman KD, Kalush A, Reilly MJ, Gardiner JC, Reeves M, Luo Z. April 2006. How Much Work-Related Injury and Illness is Missed by the Current National Surveillance System?" Journal of Occupational and Environmental Medicine: 357–365.

Sorock GS, Smith GS, Reeve GR, Dement J, Stout N, Layne L, Pastula ST. 1997. Three perspectives on work-related injury surveillance systems. American Journal of Industrial Medicine. 32: 116-128.

Wuellner S and Bonauto D. November 5, 2013. Occupational injury recordkeeping among BLS sampled establishments; Implications for national surveillance. Conference presentation at: 2013 American Public Health Association Annual Meeting, Boston, MA.

FIGURES

Figure 1. Enumeration Results for Amputation Cases in 2007-2008 from Workers' Compensation Information System (WCIS) and Health Care Facility Data (OSHPD)

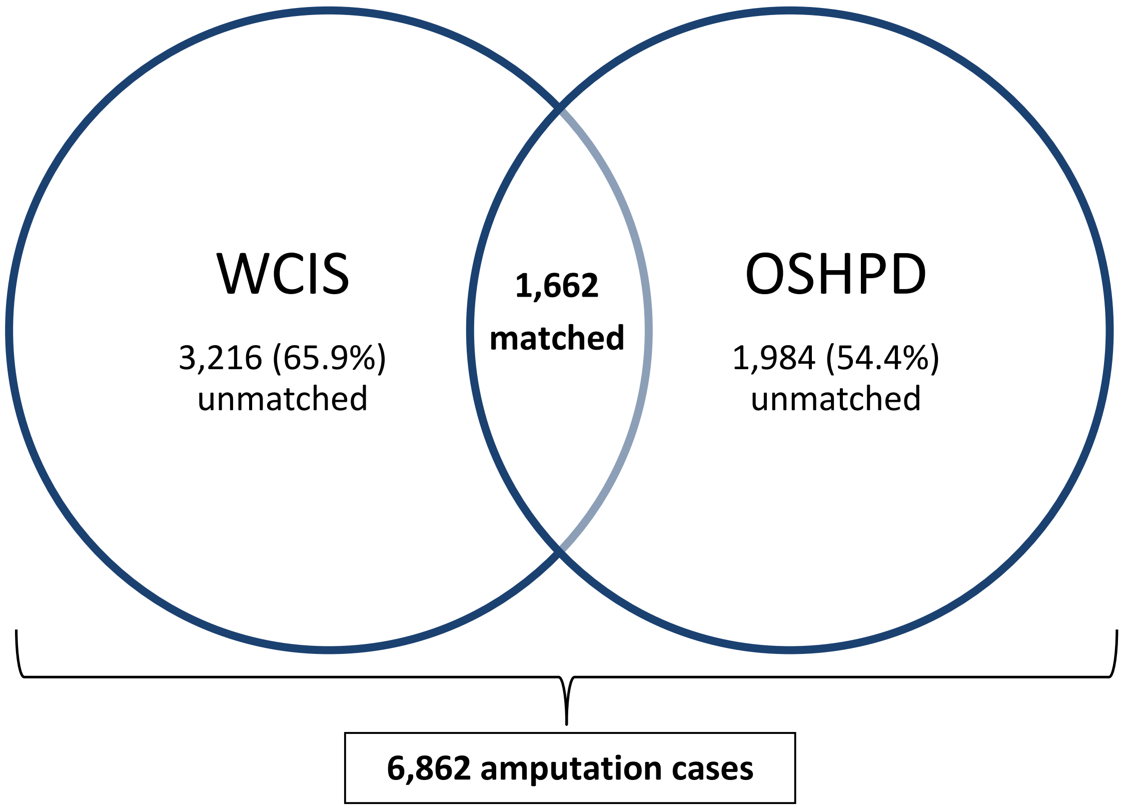
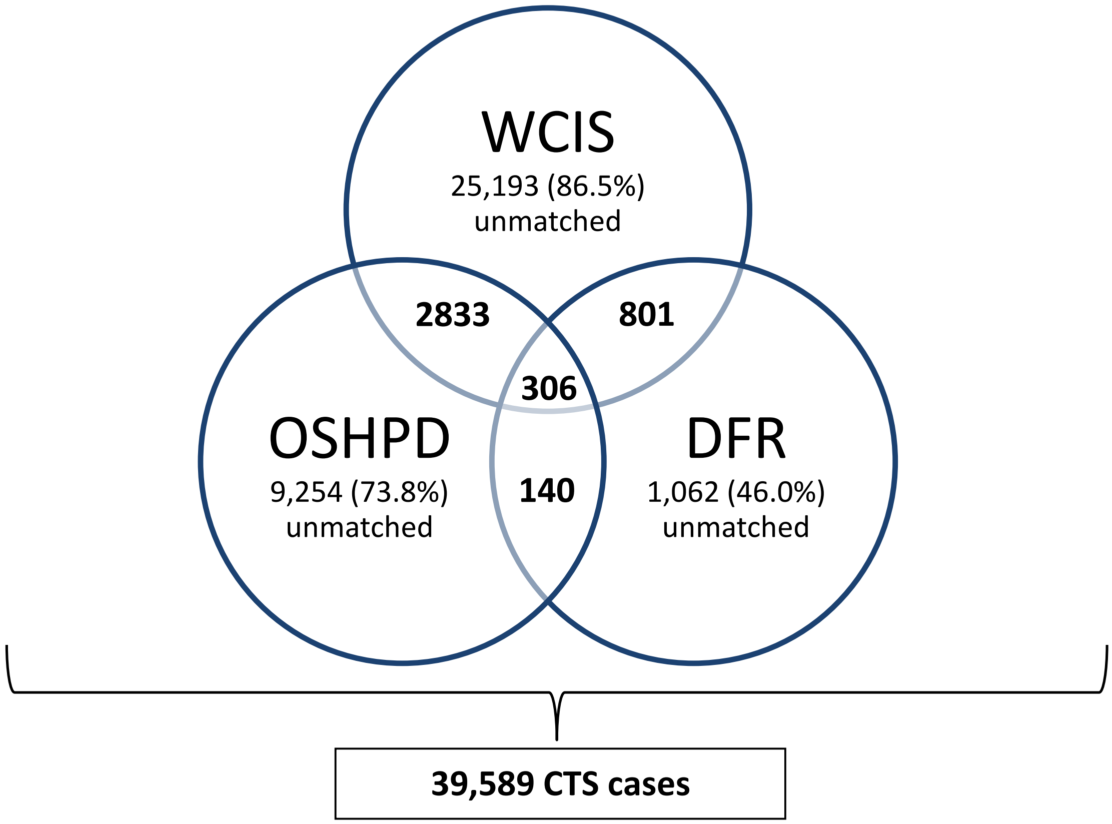


Figure 2. Enumeration Results for CTS Cases in 2007-2008 from Workers' Compensation Information System (WCIS), Health Care Facility Data (OSHPD), and Doctors' First Reports of Occupational Injury or Illness (DFR)



TABLES

Table I. Dataset Inclusion Criteria and Data Linkage Elements

Dataset (2007-2008)	**Inclusion Criteria**		**Data Linkage Elements**
	CTS	**Amputation**	
SOII	- **Nature of injury:** OIICS code for CTS	- **Nature of injury:** 3 OIICS codes for amputation	Name Employer name Date of birth Date of injury
WCIS	- **Nature of injury:** 4 WCIO codes - **Cause of injury:** 4 WCIO codes - **Part of body:** 8 WCIO Part of body codes - **Injury description:** "carpal", "CTS", "numbness", or "tingling" - **Procedure:** CPT code for CT release - **Diagnosis:** ICD-9 code for CTS	- **Nature of injury:** WCIO code - **Injury description:** "bony loss", "cut off", "amputation" or some variation in text - **Diagnosis:** 5 Diagnosis related group codes; 68 ICD9-CM codes - **Procedure:** 83 CPT; 46 ICD-9-CM codes; 124 Healthcare procedure coding system codes	Name Employer name Date of birth Date of injury
OSHPD	- **Payer:** Workers' compensation or work-related ICD-9 e-code - **Primary procedure:** CPT code for CT release - **Primary diagnosis:** ICD-9 code for CTS	- **Payer:** Workers' compensation or work-related ICD-9 e-code - **Primary diagnosis:** 5 Diagnosis related group codes; 68 ICD9-CM codes - **Primary procedure:** 83 CPT; 46 ICD-9-CM codes; 124 Healthcare procedure coding system codes	SSN Name Date of birth
DFR	- **Symptoms**: paresthesia, hypoesthesia, pain, burning/numbness affecting median nerve of hand(s) - **Physical exam findings**: Tinel's sign, Phalen's test, diminished/absent sensation to pin prick in median nerve distribution of hand; positive median nerve compression - **Electrodiagnostic findings** (NCS/EMG): Median nerve dysfunction across the CT	N/A	SSN Name Employer name Date of birth

Table II. Overall Record Linkage Results by Dataset for Cases of Amputation and CTS that Occurred in 2007-2008

Dataset	**SOII**			**WCIS**			**OSHPD**			**DFR**		
Amputation cases	217			4,881			3,646			(n/a)		
Matched to	**WCIS**	**OSHPD**		**SOII**	**OSHPD**		**SOII**	**WCIS**				
Records matched	143	65		143	1,662		65	1,662				
Match rate (%)	65.9	29.9		2.9	34.0		1.8	45.6				
CTS cases	459			29,133			12,533			2309		
Matched to	**WCIS**	**OSHPD**	**DFR**	**SOII**	**OSHPD**	**DFR**	**SOII**	**WCIS**	**DFR**	**SOII**	**WCIS**	**OSHPD**
Records matched	277	124	17	277	3,069	1,005	124	3,069	377	17	1,005	377
Match rate (%)	60.4	27.0	3.7	0.9	10.5	3.4	1.0	24.5	3.0	0.7	43.5	16.3

Supplemental Material

Section A. WCIS Extraction Criteria

Table A1. Amputation Extraction Criteria for WCIS

Criteria Variable	Values
Nature of Injury	02 - Amputation
Injury Description	Has “bony loss”, “cut off”, “amputation”, or some variation in the text
Diagnosis Related Group Codes	5 DRG codes
Diagnosis Codes	68 ICD9-CM Disease Codes
Procedure Codes	83 CPT Codes; 46 ICD-9 CM Procedure Codes; 124 Possible Healthcare Procedure Coding System Codes

The original amputation extract from the WCIS database included all claims with Nature of Injury = “Amputation”, or with appropriate keywords in the Injury Description, or with appropriate diagnosis or procedure codes in the medical billing data.

Table A2. CTS Extraction Criteria for WCIS

Criteria Variable	Code	Major Category	WCIO Description*
Nature of Injury	49	Sprain or Strain	Sprain or Tear
	52	Sprain or Strain	Strain or Tear
	78	Computer Cumulative Trauma	Carpal Tunnel Syndrome
	80	Other Cumulative Trauma	All Other Cumulative Injuries, NOC
Cause of Injury	60	Strain	Strain or Injury by, NOC
	94	Rubbed or Abraded	Repetitive Motion
	97	Strain	Repetitive Motion
	98	Miscellaneous Causes	Cumulative, NOC
Part of Body	30	Upper Extremities	Multiple Upper Extremities
	33	Upper Extremities	Lower Arm
	34	Upper Extremities	Wrist
	35	Upper Extremities	Hand
	36	Upper Extremities	Finger(s)
	37	Upper Extremities	Thumb
	39	Upper Extremities	Wrist(s) & Hand(s)
	90	Multiple Body Parts	Multiple Body Parts
Injury Description	Contains a variation of the term "carpal", "CTS", etc. or "numbness" or "tingling"		

*Source: https://www.wcio.org/Document%20Library/InjuryDescriptionTablePage.aspx

The original CTS extract from the WCIS database included all claims with Nature of Injury = “Carpal Tunnel Syndrome”, or with appropriate keywords in the Injury Description, or with appropriate diagnosis or procedure codes in the medical billing data.

Section B: WCIS Case Classification

Amputation Case Classification

Cases obtained from WCIS (using the amputation extraction criteria agreed upon in conference calls) were classified as probable or uncertain cases of amputation. Classification data were obtained from medical billing data (diagnosis and procedure codes), nature of injury, part of body, and injury description. Diagnosis for amputation and amputation-related procedure codes were considered probable amputation codes. The lists of probable diagnosis and procedure codes are in Tables B1 and B2.

A nature of injury (NOI) signifying amputation was also considered indicative of an amputation case. Both part of body and injury description were used as "criteria variables" to support probable codes and the nature of injury code. Acceptable values for these variables, which are consistent with a diagnosis of amputation, are in Table B3.

A classification scheme was established by reviewing cases and examining characteristics that suggested the level of probability that an extracted case was a true amputation case (Table B4). The scheme was confirmed on a preliminary basis by further review of classified cases to ensure that the majority of cases in a given class appeared to be properly classified. The expanded table of the case classification scheme shows all the possible combinations of the variables and the resulting case classification (Table B5). Further testing of the classification scheme by reviewing a sample of medical records is described in the body of the main article.

Table B1: List of probable amputation diagnosis codes

Description	ICD-9-CM
Amp of Thumb	885, 885.0, 885.1
Amp of Finger	886, 886.0, 886.1
Amp of Arm/hand	887
Amp of Arm/hand unilateral	887.0, 887.1, 887.2, 887.3, 887.4, 887.5
Amp of Arm/hand bilateral	887.6, 887.7
Amp of Toe	895, 895.0, 895.1
Amp of Foot	896
Amp of Foot, unilateral	896.0, 896.1
Amp of Foot, bilateral	896.2, 896.3
Amp of Leg	897
Amp of Leg, unilateral	897.0, 897.1, 897.2, 897.3, 897.4, 897.5
Amp of Leg, bilateral	897.6, 897.7
Lower limp amp status	V49.70
Upper limb amp status	V49.60

Table B2: List of probable amputation procedure codes

Description	CPT	ICD-9-CM
Amp of upper limb	23900, 23920, 23921, 24900, 24920, 24925, 24930, 24931, 24940, 25900, 25905, 25907, 25909, 25915, 25920, 25922, 25924, 25927, 25929, 25931	84
Amp finger	26910, 26951, 26952	84.01
Amp thumb		84.02
Amp upper limb		84.03 – 84.09
Amp lower limb	27290, 27295, 27590, 27591, 27592, 27594, 27596, 27598, 28800, 28805	84.1; 84.12 – 84.19
Amp lower limb NOS		84.1
Amp toe	28810, 28820, 28825	84.11
Amputation, NOS		84.91
Amp tuft of distal phalanx	11752	

Table B3. Acceptable values for criteria variables of amputation case classification scheme

Criteria Variable	Acceptable Values*	
Part of Body	30 - Multiple Upper Extremities 31 - Upper arm 32 - Elbow 33 - Lower Arm 34 - Wrist 35 - Hand 36 - Finger(s) 37 - Thumb 38 - Shoulder(s) 39 - Wrist(s) & Hand(s)	50 - Multiple Lower Extremities 52 - Upper Leg 53 - Knee 54 - Lower Leg 55 - Ankle 56 - Foot 57 - Toe(s) 58 - Great Toe 90 - Multiple Body Parts
Injury Description	Contains “amputate” Contains “severed” (without “tendon”)	

*Numeric codes are WCIO codes

Table B4. Amputation case classification scheme

	Strong Injury Description and/or Acceptable Part of Body	Neither Strong Injury Description nor Acceptable Part of Body
Nature of Injury = Amputation and/or Probable Code (Procedure or Diagnosis)	Probable	Uncertain

Table B5. Expanded amputation case classification scheme, all possible combinations

Probable Code	Nature of Injury = Amputation	Strong Injury Description	Acceptable Part of Body	Case Classification
Yes	Yes	Yes	Yes	Probable
Yes	Yes	Yes	No	Probable
Yes	Yes	No	Yes	Probable
Yes	No	Yes	Yes	Probable
Yes	No	Yes	No	Probable
Yes	No	No	Yes	Probable
No	Yes	Yes	Yes	Probable
No	Yes	Yes	No	Probable
No	Yes	No	Yes	Probable
Yes	Yes	No	No	Uncertain
Yes	No	No	No	Uncertain
No	Yes	No	No	Uncertain

Probable case:
[Probable code and/or NOI = Amputation] and a [strong ID and/or acceptable POB]

Uncertain case:
Does not have a strong injury description or acceptable POB.

Carpal Tunnel Syndrome Case Classification

Cases obtained from WCIS using the CTS extraction criteria (agreed upon in conference calls) were classified as probable, possible, or uncertain cases of CTS. Classification data were obtained from medical billing data (procedure or diagnosis codes), nature of injury, cause of injury, part of body, and injury description. Procedure codes for carpal tunnel release[1] were considered indicative of CTS, as well as diagnosis codes for CTS or mononeuritis of upper limb[2]. The remaining variables were used as "criteria variables" to support procedure or diagnosis codes, and values for these variables consistent with a diagnosis of CTS were considered acceptable as shown in Table B6.

A case classification scheme (Table B7) was established by reviewing cases and examining characteristics that suggested the level of probability that an extracted case was a true case of carpal tunnel syndrome as described below. The scheme was confirmed on a preliminary basis by further review of classified cases to ensure that the majority of cases in a given class appeared to be properly classified. Further testing of the classification scheme by reviewing a sample of medical records is described in the body of the main article.

Table B6. CTS extraction criteria from WCIS

Criteria Variable	Code	Major Category	Description
Nature of Injury	49	Sprain or Strain	Sprain
	52	Sprain or Strain	Strain
	78	Computer Cumulative Trauma	Carpal Tunnel Syndrome
	80	Other Cumulative Trauma	All Other Cumulative Injuries
Cause of Injury	60	Strain	Strain or Injury by, NOC
	94	Rubbed or Abraded	Repetitive Motion
	97	Strain	Repetitive Motion
	98	Miscellaneous Causes	Cumulative, NOC
Part of Body	30	Upper Extremities	Multiple Upper Extremities
	33	Upper Extremities	Lower Arm
	34	Upper Extremities	Wrist
	35	Upper Extremities	Hand
	36	Upper Extremities	Finger(s)
	37	Upper Extremities	Thumb
	39	Upper Extremities	Wrist(s) & Hand(s)
	90	Multiple Body Parts	Multiple Body Parts
Injury Description	Contains a variation of the term "carpal","CTS", etc. or "numbness" or "tingling"		

[1] CPT codes 64721 or 29848, or ICD-9-CM code 04.43
[2] ICD-9-CM 354.0 or 354

Table B7. CTS case classification scheme

Procedure code	ICD-9 Dx Code	Number of Acceptable Criteria Variables				
		4	3	2	1	0
64721 or 29848	**Any**	Probable	Probable	Possible	Uncertain	Uncertain
Any	**354 or 354.0**	Probable	Probable	Possible	Uncertain	Uncertain
Other or N/A	**Other or N/A**	Possible	Uncertain	Uncertain	Uncertain	Uncertain

Probable case:

1. Procedure code is 64721, 29848, or 04.43 and 3 or 4 out of 4 criteria variables acceptable
2. Diagnosis code is 354 or 354.0 and 3 or 4 out of 4 criteria variables acceptable

Possible case:

1. Procedure code is 64721, 29848, or 04.43 and 2 criteria variables acceptable
2. Diagnosis code is 354 or 354.0 and 2 criteria variables acceptable
3. 3 or 4 out of 4 criteria variables acceptable (no ICD-9)

Uncertain case:

1. Procedure code is 64721, 29848, or 04.43 and 0 or 1 criteria variables acceptable
2. Diagnosis code is 354 or 354.0 and 1 or 0 criteria variables acceptable
3. 2, 1, or 0 out of 4 criteria variables acceptable (no procedure or diagnosis codes available)

Section C: DFR CTS Case Classification Criteria

A- Symptoms of CTS: paresthesia, hypoesthesia, pain, burning or numbness affecting at least part of the median nerve distribution of the hand(s);[3] symptoms should have lasted at least one week or, if intermittent, have occurred on multiple occasions.

B- Objective evidence in the affected hand(s) or wrist(s), either:

- physical exam findings - positive Tinel's[4], positive Phalen's[5], diminished or absent sensation to pin prick in median nerve distribution of the hand, positive median nerve compression (aka Durkin's);

  OR

- electrodiagnostic findings (NCS/EMG) indicative of median nerve dysfunction across the carpal tunnel (old or new NCS/EMG with positive CTS reported)

<u>1 - Definite CTS</u>

A - symptom specific to the median nerve distribution reported
B - positive nerve conduction results reported by physician
OR
A - positive Tinel's, Phalen's, pin prick, atrophy in median nerve distribution, or median nerve compression (aka Durkin's) reported
B - positive nerve conduction results reported by physician

<u>2 - Probable CTS</u>

A - symptom specific to the median nerve distribution reported
B - positive Tinel's, Phalen's, pin prick, atrophy in median nerve distribution or median nerve compression (aka Durkin's) reported
OR
A - no symptom specific to the median nerve distribution reported
B - positive nerve conduction results reported by physician

<u>3 - Possible CTS</u>

A - symptom specific to the median nerve distribution reported
B - no positive test results reported
OR
A - no symptom specific to the median nerve distribution reported
B - positive Tinel's, Phalen's, pin prick, atrophy in median nerve distribution or median nerve compression (aka Durkin's) reported

<u>4 - Uncertain CTS</u>

A - no symptoms reported
B - no positive test results reported
(physician diagnosis of CTS or rule out CTS with no other information)

---

[3] Median nerve distribution: palmar side of thumb, index finger, middle finger, and radial half of ring finger; dorsal (back) side of the same digits above the proximal interphalangeal joint; and radial half of palm. Pain and paresthesia may radiate proximally into the arm.

[4] Tinel's sign – paresthesia elicited or accentuated by gentle percussion over the carpal tunnel.

[5] Phalen's test – paresthesia elicited or accentuated by maximal passive flexion of the wrist for one minute.

Section D: Data Linkage Methods

The matching software used is an open source tool available online (http://fril.sourceforge.net/) called Fine-grained Records Integration and Linkage Tool (FRIL). The variables available for use in record linkage are presented by dataset in Table D1. As summarized in Tables D2a-f, the record linkage process for each dataset pair was iterative such that remaining unmatched records were matched again to maximize the number of matches identified. Deterministic record linkage based on exact SSN was performed when possible, and probabilistic record linkage was used when SSN was not available (Table D2a). Tables D3a-e summarize the record linkage sensitivity analyses conducted to determine the most appropriate match score for each record linkage iteration. The match score is a number calculated by FRIL for each matched pair based on the record linkage configuration or criteria set by the user (Tables 2a-f). After linkage was performed to obtain raw matches, matches were deduplicated and refined based on the criteria summarized by Tables D4a-b.

Table D1. Variables Used for Matching across Datasets

Variable	SOII	WCIS	DFR	OSHPD
Case Name (First and Last)	✓	✓	✓	✓
Employer Name	✓	✓	✓	
Case Social Security Number (SSN)			✓	✓
Date of Birth (DOB)	✓	✓	✓	✓
Date of injury (DOI)*	✓	✓		

* Date of injury was used to deduplicate matches where multiple cases in a dataset linked to one case in another dataset, in order to obtain matches at a ratio of 1:1 (see Table D4).

Tables D2a-f: Criteria used for iterative record linkage

Table D2a: Summary of criteria used for iterative record linkage				
Criteria #	**WCIS Variable**	**SOII variable**	**Comparison type**	**Parameters**
A	SSN	SSN	Exact	N/A
B	First name	First name	Edit distance*	Match level start=0.2, end=0.4
C	Last name	Last name	Edit distance	Match level start=0.2, end=0.4
D	Date of birth	Date of birth	Exact	N/A
E	Employer name	Employer name	Edit distance	Match level start=0.2, end=0.4
F	First name	Last name	Edit distance	Match level start=0.2, end=0.4
G	Last name	First name	Edit distance	Match level start=0.2, end=0.4

* Metric of difference between two character strings based on the number of operations needed to transform one string into the other.

Table D2b: WCIS-OSHPD Linkage

Iteration #	Criteria Used (Weight)	Match Score Accept Level	Linkage Method	Raw CTS Matches	Raw Amp Matches
1	A (100)	100	Nested loop join*	8988	3040
2	B (34) C (33) D (33)	100 (CTS) 90 (Amp)	Nested loop join	172	89

* Compares each record to all other records (computationally intensive, not suitable for larger data sets.

Table D2c: SOII-OSHPD Linkage					
Iteration #	Criteria Used (Weight)	Match Score Accept Level	Linkage Method	Raw CTS Matches	Raw Amp Matches
1	B (34) C (33) D (33)	75	Nested loop join	524	205
2	B (25) C (25) D (50)	100	Nested loop join	168	15

Table D2d: WCIS-SOII Linkage					
Iteration #	Criteria Used (Weight)	Match Score Accept Level	Linkage Method	Raw CTS Matches	Raw Amp Matches
1	B (25) C (25) D (25) E (25)	75	Nested loop join	452	195
2	D (25) E (25) F (25) G (25)	75	Blocked* by DOB	4	2

* Stratifies by DOB and compares records with the same DOB with one another.

Table D2e: WCIS-DFR Linkage				
Iteration #	Criteria Used (Weight)	Match Score Accept Level	Linkage Method	Raw CTS Matches
1	A (100)	100	Nested loop join	2107
2	B (25) C (25) D (25) E (25)	65	Nested loop join	251
3	D (25) E (25) F (25) G (25)	65	Nested loop join	4

Table D2f: OSHPD-DFR Linkage

Iteration #	Criteria Used (weight)	Match Score Accept Level	Linkage Method	Raw CTS Matches
1	A (100)	100	Nested loop join	867
2	B (34) C (33) D (33)	75	Nested loop join	70

Tables D3a-e: Matching Methods Sensitivity Analyses

Table D3a: WCIS-OSHPD Amputation Iterative Record Linkage Sensitivity Analysis

	Total match pairs, all scores		Acceptable match pairs based on manual review* (from random sample, if applicable)	
WCIS-OSHPD Amputation Iteration 1 (n=3040 total match pairs)				
Match Score	**n**	**%**	**n**	**%**
100	3040	100	50 (50 pairs sampled)	100 (of sample)
WCIS-OSHPD Amputation Iteration 2 (n=92 total match pairs)				
Match Score	**n**	**%**	**n**	**%**
100	88	95.65	50 (50 pairs sampled)	100 (of sample)
91**	1	1.09	1	100
77	2	2.17	0	0
70	1	1.09	0	0

* Manual review based on first name, last name, date of birth, and employer name (if applicable)

** Acceptable score

<u>WCIS-OSHPD Amputations Linkage (Table D 3a)</u>

The first iteration was performed using exact SSN, yielding 3,040 raw matches. Review of a sample (n=50) of these showed that they were acceptable matches based on first name, last name, and date of birth.

For the second iteration, there were 89 raw matches. Of these, 88 had a confidence level of 100 and one had a confidence level of 91. A manual review of a sample of those with a confidence level of 100 (n=50 sample) and the record with a level of 91 (n=1) confirmed that they were good matches based on first name, last name, and date of birth. The three matched pairs below a score 91 did not match on first or last name or date of birth, and were therefore not considered acceptable raw matches.

Table D3b: WCIS-OSHPD CTS Iterative Record Linkage Sensitivity Analysis

	Total match pairs, all scores		Acceptable match pairs based on manual review* (from random sample, if applicable)	
WCIS-OSHPD CTS Iteration 1 (n=8988 total match pairs)				
Match Score	**n**	**%**	**n**	**%**
100	8988	100	50 (50 sampled)	100 (of sample)
WCIS-OSHPD CTS Iteration 2 (n=197 total match pairs)				
Match Score	**n**	**%**	**n**	**%**
100**	176	89.34	50 (50 sampled)	100 (of sample)
91	6	3.05	0	0
85	8	4.06	0	0
78	2	1.02	0	0
77	5	2.54	0	0

* Manual review based on first name, last name, date of birth, and employer name (if applicable)

** Acceptable score

<u>WCIS-OSHPD CTS Linkage (Table D3b)</u>

The first iteration was performed using exact SSN, yielding 8,988 raw matches. Review of a random sample (n=50) of these showed that they were acceptable matches based on first name, last name, and date of birth.

For the second iteration, the acceptance level was set at 100 which yielded 172 raw matches, after removing linkages between first names “Maria” in one data set and “Mario” in the other (n=6). Manual review of a sample of these confirmed that they were acceptable matches based on first name, last name, and date of birth. All matches below 100 did not match first name, last name or date of birth, and were not considered acceptable matches.

Table D3c: SOII-OSHPD Amputation Iterative Record Linkage Sensitivity Analysis

	Total match pairs, all scores		Acceptable match pairs based on manual review* (from random sample, if applicable)	
WCIS-OSHPD Amputation Iteration 2 (n=2408 total match pairs)				
Match Score	**n**	**%**	**n**	**%**
100	200	8.31	50 (50 sampled)	100 (of sample)
85	1	0.04	1	100
78	1	0.04	1	100
77**	3	0.12	3	100
67	2184	90.7	0 (70 sampled)	0
66	19	0.79	0	0
WCIS-OSHPD Amputation Iteration 2 (n=77 total match pairs)				
Match Score	**n**	**%**	**n**	**%**
100**	15	19.48	15	100
75	62	80.52	0	0

* Manual review based on first name, last name, date of birth, and employer name (if applicable)

** Acceptable score

SOII-OSHPD Amputations Linkage (Table D3c)

Because this match is limited to the use of only three variables (first name, last name, and date of birth), our acceptance level was conservative and subject to manual review of the linkages. We were also limited in the number of iterations that could be successfully performed with only three variables available.

For the first iteration, the acceptance level was set at 75 because manual review of the 205 raw matches with a confidence level above 75 revealed that they were good matches based on first name, last name, and date of birth. Manual review of a sample of matches below 75 revealed that these were not acceptable matches. Most of these unacceptable matches did not match on date of birth but matched only on common first and last names, so we are certain that these were not true matches (confidence scores of 66 and 67).

The weight of exact date of birth was increased for the second iteration, which linked the unmatched BLS SOII records from the first iteration to the full OSHPD amputation dataset. The acceptance level was set at 100 because manual review of the 15 raw matches with a confidence level of 100 revealed that they were good matches based on first name, last name, and date of birth. Manual review of a sample of matches below 100 revealed that these were not acceptable matches. However, these 15 matches were already identified in the first iteration and therefore did not contribute to the total number of linkages found.

Table D3d: SOII-OSHPD CTS Iterative Record Linkage Sensitivity Analysis

	Total match pairs, all scores		Acceptable match pairs based on manual review* (from random sample, if applicable)	
WCIS-OSHPD Amputation Iteration 2 (n=524 total match pairs, acceptance level=75)				
Match Score	**n**	**%**	**n**	**%**
100	513	97.9	50 (50 pairs sampled)	100
96	4	0.76	4	100
91	2	0.38	2	100
88	1	0.19	1	100
85	4	0.76	4	100
Below 75**	-	-	0 (70 pairs sampled)	0
WCIS-OSHPD CTS Iteration 2 (n=168 total match pairs)				
Match Score	**n**	**%**	**n**	**%**
100**	168	100.00	50 (50 pairs sampled)	100
Below 100	-	-	0 (70 pairs sampled)	0

* Manual review based on first name, last name, date of birth, and employer name (if applicable)

** Acceptable score

SOII-OSHPD CTS Linkage (Table D3d)

For the first iteration, acceptance level was set at 75 which yielded 524 raw matches. Of these matches, 96% had a confidence level of 100, and a manual review of a sample of these confirmed that they were acceptable matches based on first name, last name, and date of birth. The remaining 4% had confidence levels that ranged from 85 to 96. Manual review of these matches confirmed that these were acceptable matches as well. A sample (n=70) of matches below 75 revealed that these were not acceptable. Most of these unacceptable matches did not match on date of birth but matched only on common first and last names, so we are certain that these are not true matches.

The weight of exact date of birth was increased for the second iteration, which linked the unmatched BLS SOII records from the first iteration to the full OSHPD amputation dataset. Acceptance level was set at 100 because manual review of a sample of the 168 raw matches with a confidence level of 100 revealed that they were acceptable matches, while a sample (n=70) of matches below 100 were not. However, these matches were already identified in the first iteration and therefore did not contribute to the total number of linkages found.

Table D3e: WCIS-SOII Iterative Record Linkage Sensitivity Analysis

	Total match pairs, all scores		Acceptable match pairs based on manual review* (from random sample, if applicable)	
WCIS-SOII all Iteration 1 (n=39052 total match pairs)				
Match Score	**n**	**%**	**n**	**%**
100	12274	31.43	50 (50 sampled)	100.00
77 to 99	4901	0.14	50 (50 sampled)	100.00
76	250	0.64	96 (100 sampled)	96.00
75**	20465	52.4	86 (100 sampled)	86.00
74	25	0.06	14	56.00
71 to 73	346	0.45	-	-
70	112	0.29	63	56.25
69	19	0.05	8	42.11
68	424	1.09	54	12.74
67	117	0.3	79	67.52
66	51	0.13	26	50.98
65	68	0.17	11	16.18
WCIS-SOII all Iteration 2 (n=405 total match pairs)				
Match Score	**n**	**%**	**n**	**%**
100	64	15.8	64	100.00
76 to 99	30	0.25	30	100.00
75*	290	71.6	95 (100 sampled)	95.00
72	2	0.49	1	50.00
70	3	0.74	0	0.00
65 to 69	16	0.49	7	43.75

* Manual review based on first name, last name, date of birth, and employer name (if applicable)

** Acceptable score

<u>WCIS-SOII all Linkage (Table D3e)</u>

For the first iteration, acceptance level was set at 75 based on manual review of different levels of acceptance scores starting at a score of 65. Only 16% of matches with a score of 65 were acceptable matches after review. Cases from each score were reviewed until the proportion of acceptable matches appeared to be maximized and proportion of bad matches was minimized. Of matches with a score of 74, only 56% of matches with a score were acceptable matches, while 85% of a sample of 100 pairs with a score of 75 were acceptable matches. Of a sample of 100 pairs with a score of 76, 96% were acceptable matches. Therefore, 75 was chosen as an appropriate acceptance level for this iteration. There were 37,890 raw matches identified: 452 (1.19%) had a CTS endpoint in SOII, and 195 (0.51%) had an amputation endpoint in SOII.

The second iteration of matching was blocked, or stratified, by exact date of birth in order to link the records in a more computationally efficient manner. First names were compared to last names in order to pick up any matched cases whose name may have been entered incorrectly into the databases. An acceptance level was determined by manual review in the same manner as the first iteration. Of a sample of 100 cases with a score of 75, 95% were acceptable matches, while only 50% or less of the cases from lower scores were good matches. There were 384 raw matches identified: 4 (1.04%) had a CTS endpoint in SOII, and 2 (0.52%) had an amputation endpoint in SOII.

Table D4: Deduplication for Enumeration

Table D4a: WCIS – SOII Linkage De-duplication and Refinement for Enumeration		
	Amputation	**CTS**
Raw matches	**197**	**456**
SOII duplicates removed because • Not amputation/CTS in WCIS • Not highest case class in WCIS • Not closest injury date	28	99
WCIS duplicates removed because • Not closest injury date	0	10
One-to-one matches for enumeration	**169**	**347**

TableD4b: WCIS – OSHPD Linkage De-duplication and Refinement for Enumeration		
	Amputation	**CTS**
Raw matches	**3129**	9160
WCIS duplicates removed because • service/admit date (in OSHPD) was not closest to injury date (in WCIS)	482	508
OSHPD duplicates removed because • not highest case class in WCIS • service/admit date was not closest to injury date	415	1740
Matches removed because • service date occurred before injury date	30	**n/a**
Matches removed because • not amputation or CTS in WCIS	293	1322
One-to-one matches for enumeration	**1909**	**4167**

Table 4c: SOII – OSHPD Linkage De-duplication and Refinement for Enumeration		
	Amputation	**CTS**
Raw matches	**205**	**524**
SOII duplicates removed because • service/admit date (in OSHPD) was not closest to injury date (in WCIS)	37	134
OSHPD duplicates removed because • service/admit date was not closest to injury date	0	9
Matches removed because • service date occurred before injury date or	8	n/a
One-to-one matches for enumeration	**159**	**381**
Amputation/CTS in SOII (used for enumeration)	**72 (45.28%)**	**132 (34.65%)**
Non-amputation/CTS in SOII	**87 (54.72%)**	**249 (65.35%)**

Table 4d: DFR – OSHPD De-duplication	
	CTS
Raw matches	**937**
DFR duplicates removed because • not primary procedure/diagnosis in OSHPD • not first OSHPD case	13
DFR date of injury not in 2007-2008	72
One-to-one matches for enumeration	**625**