The California Commission on Health and Safety and Workers' Compensation

Reporting Workers' Compensation Injuries in California: How Many are Missed?

Prepared for CHSWC by:

Leslie I. Boden
Al Ozonoff
Boston University School of Public Health

CHSWC Members
Angie Wei (Chair 2008)
Catherine Aguilar
Allen Davenport
Sean McNally
Kristen Schwenkmeyer
Robert B. Steinberg
Darrel "Shorty" Thacker

Executive Officer
Christine Baker

State of California
Labor and Workforce Development Agency
Department of Industrial Relations

August 2008

Table of Contents

Acknowledgments

Many organizations and individuals contributed to this study. From the U.S. Department of Labor, Bureau of Labor Statistics, John Ruser, Tony Barkume, Jim Barnhardt, Brooks Pierce, Terry Burdette, Jim Spletzer, and John Bishow provided access to data, assistance in interpretation, and much other help. Data access and assistance were also provided by Christine Baker, Bill Kahley, Martha Jones, and Liza Dizon (California); Teri Van Hoomissen and Brian Zaidman (Minnesota); Mark Llewellyn and Tom Dauphinee (New Mexico); Dan Adelman, Ed Bissell, and Mike Maier (Oregon); Barbara Silverstein, Heather Grob, and Bill Blanford (Washington); Frances Huntley-Cooper, Lee Shorey, and Duane Frisch (Wisconsin); and Syed Islam and Ed Doyle (West Virginia).

Frank Neuhauser, Brian Zaidman, Mike Maier, Lori Carleton, and Chris Bailey provided helpful comments on an earlier draft.

Stephen Evans has been an able and good-natured programmer and colleague for the duration of this research, and Greg Howard provided excellent research assistance. Wayne Gray provided advice on how to link the BLS and workers' compensation data.

Our thanks also go to the National Institute for Occupational Safety and Health, Grant 5 R01 OH 007596, and the State of California, Commission on Health and Safety and Workers' Compensation for the funding to carry out this research.

Executive Summary

In recent years, California newspapers have reported about employers who have systematically underreported workplace injuries. In 2004, the Los Angeles Times reported that Southern California Edison Co. had received safety bonuses of $35 million from the State of California based on Edison's reporting of worker injuries that was later found to be incomplete (Douglass 2004). In 2005 and 2006, the Oakland Tribune published a series of articles questioning injury reporting by KFM, the consortium hired to rebuild the San Francisco Bay Bridge (Tucker 2005).

These newspaper articles and research studies of workplace injury and illness reporting have raised concerns that underreporting may be much more widespread. For this reason, the Commission on Health and Safety and Workers' Compensation (CHSWC) asked Boston University School of Public health to analyze existing injury reporting data to determine whether underreporting is a substantial issue in the California workers' compensation system. In this study, we also compare estimates of underreporting in California with estimates for Minnesota, New Mexico, Oregon, Washington, West Virginia, and Wisconsin.

Findings

Underreporting in California

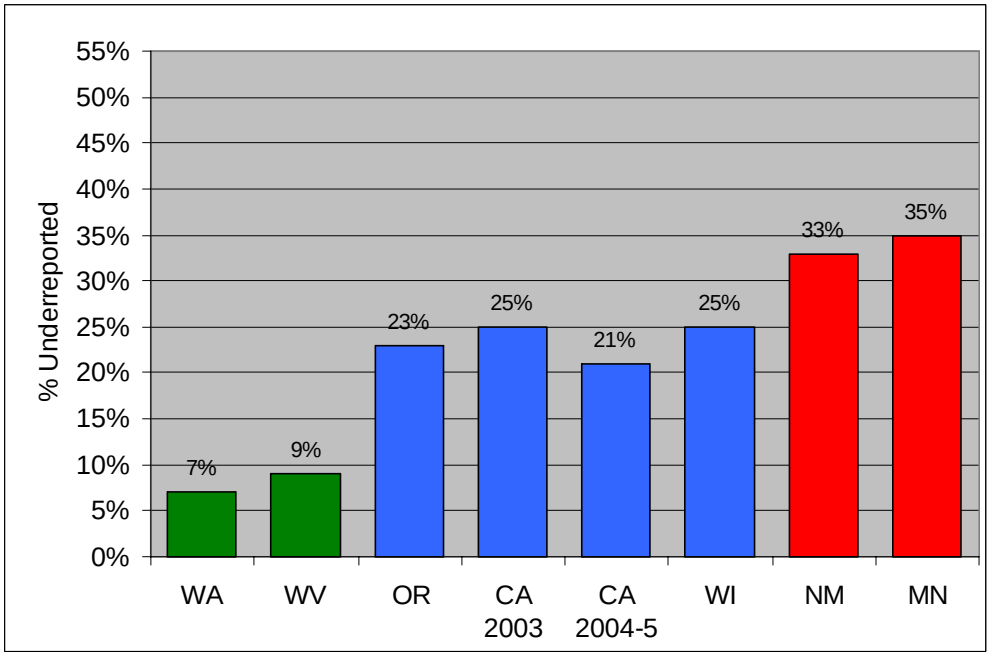
- In 2003, at least 25% of lost-time injuries in California were not reported to the Workers' Compensation Information System (WCIS). A less conservative estimate of underreporting implies that 40% of lost-time injuries went unreported.

- Underreporting did not appear to increase after the 2004 reforms were implemented. In fact, it looks like reporting improved. Our conservative estimate of underreporting for injuries occurring between July 1, 2004, and June 30, 2005, is 21%. Our less conservative estimate is 29%.

Underreporting compared with 6 other states

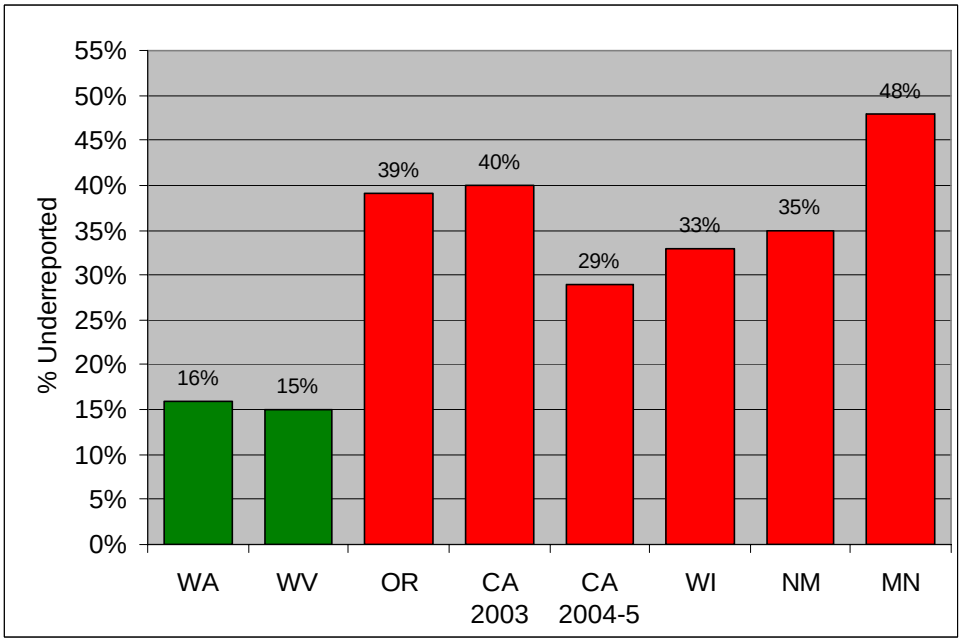
- The most conservative estimate of underreporting of lost-time cases in the other 6 states ranges from 7% in Washington to 35% in Minnesota. Estimates for California are in the middle of this range.

Figure S1. Estimates of Underreporting of Workplace Injury and Illness in 7 States: Lost-time Cases, Independent Reporting to BLS and WCIS



- Less conservative estimates for the other states suggest a range of underreporting from 15% to 48%. Again, California is not the highest nor the lowest state of those studied.

Figure S2. Estimates of Underreporting for 7 States: Strong Relationship between BLS and WCIS Reporting



Other findings

- Almost 400,000 WCIS cases from 2003 lacked information on benefit payments. It seems unlikely that no benefits were paid for all these cases, so we infer that there is considerable underreporting of benefit payments to WCIS.

- Many cases were reported to WCIS long after the end of the injury year. We initially collected data on 2003 cases more than 27-1/2 months after the end of the injury year. More than 43,000 new first reports of injury were filed in the next 20 months.

Recommendations

We have several specific recommendations that could help to identify reporting problems and reduce underreporting. We understand that it may be infeasible to adopt all of our recommendations, but we think that they are a useful starting place for planning. However, our first recommendation is more general.

- CHSWC, the California Division of Workers' Compensation (DWC), the California Department of Public Health (CDPH), the Division of Labor Statistics and Research (DLSR), and Cal/OSHA should convene an interagency underreporting task force to develop a plan for improving WCIS reporting. This would include identifying employers, insurers, and third party administrators (TPAs) that do not report compensated injuries. It would also focus on identifying underreported benefit payments and late reports. They should convene an interagency underreporting task force to develop a plan for improving WCIS reporting. This task force could include not only knowledgeable people from these agencies, but also people involved in other relevant activities, like California's reporting to the BLS survey and planning for the California Trauma Registry.

Specific recommendations include:

- DWC could strengthen its efforts to identify problem areas in reporting of workplace injuries and illnesses. In doing so, DWC may identify problems in the way reporting systems work, in addition to identifying noncompliance with reporting requirements.

- DWC and Cal/OSHA could consider collaborating to identify employers who underreport injuries. Employers who engage in substantial underreporting to either system could be given substantial penalties, and the program and penalties could be publicized. DWC could also consider penalties for late reporting to WCIS. If current laws and regulations are inadequate to support such a program, this could be addressed.

- The DWC could begin an inquiry into the 40-50% of reported claims that lack information about benefit payments. DWC could draw a random sample of such cases with dates of injury at least 3 years in the past from a subset of claims administrators for insurers, TPAs, and self-insured employers. Initially the claims administrators might be chosen because they have a relatively high proportion of cases lacking benefit reports. The DWC could submit the sample to the trading partners and request up-to-date information on benefit payments and claim status. From this information and discussions with trading partners, the DWC may be able to diagnose systematic problems and fashion solutions.

- It may be possible to compare Workers' Compensation Insurance Rating Bureau (WCIRB) data by reporting entity with WCIS data as a way of identifying reporting entities that substantially underreport compensated cases.

- California collects data on hospital and emergency room discharges and from ambulatory surgery clinics through the Medical Information Reporting for California (MIRCAL) program. The Department of Industrial Relations (DIR) might explore whether these data could be used to look for unreported workplace injuries and illnesses. The data contain diagnosis and Social Security Number of the patient and identify the expected source of payment. They do not identify the employer. If WCIS data included state Employment Development Department (EDD) account numbers (EANs), cross-matching with EDD wage files to determine the employer would be easier and more accurate than otherwise. We do not know if there are any legal issues precluding this use of MIRCAL data.

- CHSWC could explore linking other state occupational safety and health information systems with WCIS data to determine whether injuries and illnesses have been reported and compensated where appropriate.

- DIR could explore automating the doctor's first report of occupational injury or illness and requiring all doctor's first reports to be electronically transmitted. For example, reports could be filled out on the Internet and automatically transmitted to DIR. These reports could be compared with WCIS files to determine where underreporting occurs.

- CHSWC has recently funded a study of employer underreporting and misreporting of payroll to workers' compensation insurers (Neuhauser and Donovan 2007). This includes, among others, underreporting payroll, misclassifying the occupation of employees, and misclassifying employees as independent contractors. If an injured worker is misclassified as an independent contractor, the injury is not reported. Also, employers who underreport or misreport payroll may be more likely than others to suppress or simply not report workplace injuries and illnesses. If DWC or Cal/OSHA develops programs to detect WCIS underreporting and enforce reporting requirements, reporting at these employers might be examined first.

- DWC may want to consider rejecting reports of injury with invalid or incorrect Federal Employer Identification Numbers (EINs). These numbers can be valuable for potential uses of WCIS, including but not limited to the underreporting issue.

- DWC should consider adding the state EAN as a required field in the First Report of Injury. This would allow easier and more accurate linkage with EDD wage files and other state data collected from employers.

- California has recently added workers' compensation questions to the states' Behavioral Risk Factor Surveillance System (BRFSS) survey. This could be used as another way of getting a handle on the extent of workers' compensation underreporting. Over time, it could be used to determine whether reporting is improving.

Introduction

Electronic reporting of injuries and illnesses to WCIS became mandatory in 2000. Since then, claim administrators have been required to submit electronic data about all workers' compensation claims, including information about the injured worker, the injury, and benefit payments. In principle, WCIS should have information on every compensable injury occurring at a covered employer. Coverage is mandatory for all employers with at least one employee, except for those covered by the Federal Employees' Compensation Act, the Longshore and Harbor Workers' Compensation Act, or the Federal Employers' Liability Act.

Still, eligible workplace injuries may go unreported. For example, WCIS may not receive injury reports because injured workers or their physicians have not reported injuries to their employers. Alternatively, a claim may be filed and paid, but the employer, insurer, or TPA may neglect to report the claim information to WCIS.

Barriers to reporting can occur for different reasons, many of which are described in detail by Azaroff et al (2002). Workers and their physicians, for example, may not know that their injuries or illnesses are work-related. This is particularly important for chronic occupational diseases (Boden et al. 2006). They may know that a health problem is work-related, but they may not know that it is covered by workers' compensation. Even if workers know that their injuries or illnesses are compensable, they may think that they are too minor or the benefits are too small to be worth the effort and discomfort of dealing with the workers' compensation process. They may feel that a stigma is attached to getting workers' compensation benefits, or they may worry about losing their jobs if they file for benefits.

Managers may discourage claim filing because their job evaluations could be negatively affected. Employers may want to reduce claim filing because of its impact on workers' compensation costs or because injury costs can be a factor in competing for contracts. Less reporting can be facilitated by modified work programs (often with positive effects on return to work) or programs that provide group bonuses for no (reported) injuries. These programs can lead managers and fellow workers to apply pressure on injured workers not to report.

Employers and insurers can also reduce the number of claims filed by frequently contesting claims. This is a way of screening out unwarranted claims, but it can also discourage claim filing more generally and cause workers with legitimate claims to abandon them.

A substantial disparity between the number of injuries that are reported and the actual number that occur has several implications. First, if policymakers think that the number of workplace injuries and illnesses is smaller than it actually is, they may devote fewer resources to prevention. Second, reporting may be particularly incomplete for specific

conditions, groups of workers, and employer types. As a consequence, we may pay less attention to safety for those conditions, workers, and employers for which underreporting is the greatest. In addition, when compensable work-related injuries and illnesses are not filed as workers' compensation cases, benefits go unpaid, and the costs of these injuries may be shifted to workers and their families, to private health insurance, and to government disability and health insurance programs.

This study addresses the reporting of lost-time injuries to the California WCIS for injuries occurring during two time periods: January 1, 2003-December 31, 2003, and July 1, 2004-June 30, 2005. These periods are just before and just after the 2004 workers' compensation reform legislation. We chose these time periods to see if we could find a substantial change in reporting that might have been influenced by the 2004 legislation. The study also compares reporting in the California workers' compensation system with that in 6 other states: Minnesota; New Mexico; Oregon; Washington; Wisconsin; and West Virginia.

Before turning to the methods used in the study and the results, we present a short review of previous studies of workplace injury reporting.

Past Studies of the Reporting of Workplace Injuries and Illnesses

The most commonly used sources of workplace injury frequency are state workers' compensation data and the Survey of Occupational Injuries and Illnesses conducted annually by the Bureau of Labor Statistics, U.S. Department of Labor (BLS). Most state workers' compensation agencies collect first reports of injury, and many collect subsequent information on the duration of disability, benefit payments, and so on. Frequently, the states themselves or organizations involved in the process of setting workers' compensation rates publish information on the number and type of injuries. BLS collects information from a statistical sample of establishments on injuries that the Occupational Safety and Health Administration (OSHA) requires employers to record on the OSHA Form 300. Annually, BLS publishes estimates of the number of injuries and injury rates in the U.S. and by state. It also publishes more detailed information on cases involving at least one day away from work.

A series of studies beginning in the 1980s demonstrated that both nonfatal and fatal injuries are underreported. Using a simple comparison of the number of injuries reported to each system, Hanrahan (1987) reported an overall injury rate in Wisconsin that was 31% higher using data from the BLS survey than using workers' compensation data. Some or all of this disparity might have been caused by the fact that Wisconsin has a 3-day waiting period for income benefits (a threshold for entry into the workers' compensation database), and BLS does not. Hanrahan could not access BLS data at the injury level, so he could not account for this reporting difference.

Fine, Silverstein, Armstrong, Anderson, and Sugano (1986) conducted a study of injury reporting for upper extremity cumulative trauma disorders at 3 large automobile

manufacturing plants, two of which had recently designed improved cumulative-trauma surveillance systems within their occupational medicine departments. They compared four data sources: the OSHA injury logs used in the BLS survey, workers' compensation cases, records of medical absences longer than 3 days, and plant medical records. For acute traumatic injuries, OSHA injury logs and medical absence records generated similar incidence rates, workers' compensation rates were substantially lower, and rates derived from plant medical records were four times as high as from injury logs. For cumulative trauma disorders, medical absence and plant medical records produced incidence rates *nearly 10 times as high* as workers' compensation and *nearly 100 times as high* as injury logs. The increasing recognition of cumulative trauma disorders in the past 3 decades has undoubtedly improved reporting.[1] However the size of the disparity, even given this possibility, is disquieting.

A study of the automobile manufacturing industry in 1986-1989 compared incidence rates for musculoskeletal disorders using the OSHA log (the source of BLS survey data), workers' compensation, sickness and accident insurance, and self-reported symptoms, medical treatment, and lost time (Silverstein, Stetson, Keyserling, and Fine 1997). They found a great deal of variation from plant to plant and by calendar year in the relationship between incidence rates calculated from the OSHA 200 logs and from workers' compensation data. Indeed, in some cases, the OSHA incidence rate was much lower than the workers' compensation rate, and in other cases, it was considerably higher.

Several more recent studies have focused on the filing of workers' compensation claims. Biddle, Roberts, Rosenman, and Welch (1998) compare legally required physician reports of occupational diseases with workers' compensation claims in Michigan. The Michigan definition of occupational disease (as distinguished from occupational injury) includes many musculoskeletal disorders and, indeed, over 36% of reported diseases were for sprains, strains, or carpal tunnel syndrome. Biddle et al. found that less than half (and perhaps as little as 10%) of workers with known or suspected occupational diseases had filed workers' compensation claims. In a recent follow-up study focusing on musculoskeletal disorders, Biddle and Roberts (2001) concluded that, among those who missed more than 7 days from work (making them eligible for workers' compensation income benefits), less than 60% filed for these benefits. Many who did not file used other programs, like sick leave. Similarly, Morse, Dillon, Warren, Levenstein, and Warren (1998) conducted a random-digit-dial survey in Connecticut to determine the period prevalence of work-related upper extremity musculoskeletal disorders. They also asked respondents whether they had filed for workers' compensation. Of cases in which a medical provider had identified the condition as work-related, only 21% of respondents said that they had filed for workers' compensation benefits. Another study by Lakdawalla and Reville (2001) based on the National Longitudinal Survey of Youth indicates that only 55% of reported occupational injuries resulted in workers' compensation claims.

[1] Between 1982 and 1992, BLS reported an increase in reported cumulative trauma disorders from 22,600 to 332,100 cases (Panel on Musculoskeletal Disorders and the Workplace 2001, p. 37). Much of this increase may have been a consequence of more complete reporting.

Maier and Reinke (2004) use the Oregon Population Survey to estimate workers' compensation reporting. They estimate that 61% of injuries in 2002 and 54% of injuries in 2001 are reported. Using data from the 2002 Washington State Behavioral Risk Factor Surveillance System survey, Fan et al. (2006) examine workers' compensation reporting in that state. They estimate that only 52% of injured workers filed a workers' compensation claim.

Recent studies have used capture-recapture methods to generate better measures of the frequency of carpal tunnel syndrome (Maizlish, Rudolph, Dervin, and Sankaranarayan 1995) and upper extremity musculoskeletal disorders (Morse, Dillon, Warren, Hall, and Hovey 2000). Another recent study addresses the reporting of all types of non-fatal workplace injuries and illnesses by workers' compensation and the BLS survey using capture-recapture methods. Rosenman et al. (2006) compared injury reporting to the BLS survey and workers' compensation reporting in Michigan. They found only 32% ascertainment by the BLS. For workers' compensation, they found 66% ascertainment, at the low end of the range in our states.

Finally, a study by Boden and Ozonoff (2008) uses capture-recapture methods to estimate injury reporting in 6 states. Using very conservative assumptions, they estimate that workers' compensation systems in the 6 states studied missed between 7% and 35% of lost-time injuries. Under less conservative assumptions, between 16% and 48% of lost-time injuries were not reported.

Not even fatal injuries are completely recorded by the systems designed to capture them. Stout and Bell (1991) review studies of the completeness of fatal injury reporting, using OSHA fatality reports, workers' compensation records, death certificates, medical examiner records, and state health department records. Between 40% and 70% of occupational fatalities identified by at least one source were captured by workers' compensation, and between 21% and 42% were captured by OSHA fatality reports. (Of course, death certificates and medical examiner records cannot be used for surveillance of non-fatal injuries.) These studies led to a restructuring of the BLS fatality reporting system to the more comprehensive Census of Fatal Occupational Injuries.

Studies have shown not only that reporting is incomplete but also that it is affected by the incentives faced by those who report. Ruser and Smith (1988) studied the impact of OSHA's records-inspection policy of the early 1980s. They estimated that this policy, by which OSHA did not inspect workplaces with low injury rates, led to a decline in reporting of 5 to 14%. Boden and Gold (1984) examined reporting of coal dust exposures by mines subject to regulatory penalties based on samples collected by the mine operators. They found evidence of widespread underreporting, with most mines providing unrealistically low samples at least 15-30% of the time. Recently, Boden and Ruser (2001) have shown that changes in workers' compensation statutes in the 1990s to make claim filing more difficult caused a 10% reduction in those states of days-away-from-work injuries reported to BLS.

How do we Estimate the Total Number of Injuries?

Introduction

To address injury reporting in California, we use data from two sources. The first is the WCIS system, which is designed to include information on all workplace injuries, both those involving only medical payments and those for which indemnity benefits are paid. The second is data from California included in the BLS survey. The BLS survey is an annual, nationally representative survey of approximately 165,000 private industry establishments conducted by BLS. The BLS survey case-specific data is designed to include all cases where workers do not return to work for at least one full day after the day of the injury. Unlike WCIS, the BLS survey covers only a statistical sample of injuries in the state. BLS samples individual establishments, designing the sample to represent workplaces across the entire state.

To determine how many injuries go unreported, we estimate the total number of injuries, including those reported to neither BLS nor WCIS. The technical details of how we do this are in Appendix 1. We describe the basic idea behind the estimates in the rest of this section.

The method we use, called capture-recapture analysis, relies on assumptions about reporting to estimate the total number of injuries. A critical assumption is the relationship between reporting to the two systems. We begin by assuming that reporting to BLS is independent of reporting to WCIS. This means that the probability that an injury is reported to BLS does not depend on whether it is reported to WCIS. More concretely, suppose that 80% of injuries reported to WCIS are also reported to BLS. Then, independence implies that, for injuries that are not reported to WCIS, the same proportion, 80%, are reported to BLS. (See Appendix 1 for a more technical description.)

However, independence is almost certainly an unwarranted assumption. For many reasons, we expect that reporting to these two sources will be positively associated. (Statisticians call this positive source dependence.) This means that injuries reported to WCIS are more likely to be reported to BLS. There are some obvious reasons why this is probably the case. First, reporting to both systems is strongly influenced by employers. In some workplaces, the same individual may be responsible for filling out the BLS survey and for filing workers' compensation claims. Even if this is not the case, an employer may believe that a workplace injury that is not a compensable workers' compensation claim should not be reported to BLS. In addition, unless an injured worker reports an injury to a supervisor, it will very rarely enter either system.

Despite the likelihood that reporting to WCIS and BLS is not independent, this assumption provides a good starting point for a study of workers' compensation reporting. This is because, as we will explain shortly, assuming independence provides a lower-bound estimate of underreporting.

How capture-recapture methods work

We explain this method by looking at an example constructed with made-up numbers. Suppose that, in a given year, there were 200,000 injuries reported to WCIS and 240,000 reported to BLS. Provided we have some specific information about each of these cases, like the worker's name, date of birth, sex, the employer's name, the date of injury, and so on, we could link individual injury reports from the two datasets. If we did this, we might find out that 160,000 injuries were reported to both systems, 40,000 were reported only to WCIS, and 80,000 were reported only to BLS (Table 1). We would not know the number not reported to either system, so we would not know the total number of injuries.

Table 1. Hypothetical Example: Observed Cases Reported to WCIS and BLS

BLS	WCIS		
	(1) Not reported	(2) Reported	(3) Total
Not reported	?	40,000	
Reported	80,000	160,000	240,000
Total		200,000	

If we assume that reporting to BLS is independent of (unrelated to) reporting to WCIS, we can estimate of the number of injuries not reported to either. We know that that 80% (160,000/200,000) of injuries *reported* to WCIS are reported to BLS. Independent reporting means that the same proportion, 80% of those *not reported* to WCIS are reported to BLS. We know that 80,000 cases are not reported to WCIS but reported to the BLS. And 80,000 is 80% of 100,000. So, independence implies that 100,000 cases went unreported to WCIS and that 20,000 (100,000 minus 80,000) cases were not reported to either system (Table 2).

Table 2. Hypothetical Example: Estimate of Cases Reported to Neither WCIS nor BLS: Assuming Independent Reporting to BLS and WCIS

BLS	WCIS		
	(1) Not reported	(2) Reported	(3) Total
Not reported	**20,000**	40,000	
Reported	80,000	160,000	240,000
Total	**100,000**	200,000	

We can then fill in the rest of the table by adding columns (1) and (2) to get the total (Table 3). This gives us the overall estimate of 300,000 injuries. Of those, 240,000 (80%) are reported to BLS and 200,000 (67%) are reported to WCIS. This example illustrates

how capture-recapture analysis estimates the total number of injuries, including the number not reported to either system.

Table 3. Hypothetical Example: Estimate of Total Number of Injuries: Assuming Independent Reporting to BLS and WCIS

BLS	WCIS		
	(1) Not reported	(2) Reported	(3) Total
Not reported	**20,000**	40,000	**60,000**
Reported	80,000	160,000	240,000
Total	**100,000**	200,000	**300,000**

In Tables 2 and 3, the proportion of injuries not reported to BLS is 20%, whether or not they are reported to WCIS. However, BLS underreporting is almost certainly greater for injuries not reported to WCIS than for those that are reported. This implies that the number of unreported cases in column (1) is greater than 20,000. With the information available, we do not know whether that number is 20,100 or 200,000, but we are confident that it is more than 20,000.

We can estimate the total number of injuries on the assumption that there is a positive relationship between reporting to BLS and reporting to WCIS. In Table 4, like in Table 3, 80% of cases reported to WCIS are also reported to BLS. However, in this example, the proportion of injuries reported to BLS is only 50% (80,000) for those that are not reported to WCIS. In this case, the total number of cases not reported to WCIS must be 160,000.

Table 4. Hypothetical Example: Estimate of Total Number of Injuries: BLS Reporting more Likely for Cases Reported to WCIS

BLS	WCIS		
	(1) Not reported	(2) Reported	(3) Total
Not reported	**80,000**	40,000	**120,000**
Reported	80,000	160,000	240,000
Total	**160,000**	200,000	**360,000**

The observed numbers, in black, remain the same, but the number of estimated injuries not reported to either BLS or WCIS rises, as does the total number of estimated injuries. The estimated number of injuries not reported to WCIS increases from 100,000 to 160,000, and estimated underreporting increases from 33% (100,000/300,000, Table 3) to 44% (160,000/360,000, Table 4). Similarly, the number not reported to BLS rises from 60,000

to 120,000, and estimated BLS underreporting rises from 20% (60,000/300,000, Table 3) to 33% 120,000/360,000, Table 4).

Data

BLS provided us with complete detailed information from its annual Survey of Injuries and Illnesses on cases involving at least one day away from work in 2003-2005. In March 2006, DWC provided us with detailed information from WCIS from first and subsequent reports of injury for all 2003 injury date cases in this database. In November 2007, DWC provided parallel information on cases with injury dates from July 1, 2004, through June 30, 2005. Both datasets include: name, gender, and age of the injured worker; date of injury; days of lost work; days of restricted work; nature of injury; part of body affected; employer identifier; and employer name and address. WCIS data include detailed data on indemnity payments as well.

Industries, Establishments, and Injuries Covered by this Report

To conduct this study, we had to limit injuries to those reported to both systems. Unlike WCIS, the BLS survey does not cover all reportable injuries in the state. BLS samples individual establishments, where the sample is designed to represent workplaces in the entire state. In order to take this into account, the injuries we compared were limited to employers in the BLS sample. Even though this means that only a subsample of California workplaces were part of the study, the design of the BLS sample allows us to estimate reporting percentages for the entire state.

There was one problem unique to California that created a problem for linking employers in the BLS and WCIS data: although injury reports are supposed to include the EIN, many WCIS records lacked this information or had an invalid EIN. This did not prevent us from linking individual records, where we were able to use other information, but it presented a problem in determining whether unlinked cases in WCIS were for injuries in firms that were sampled by BLS. In all, 38% of 2003 cases and 25% of 2004-5 cases lacked a valid EIN. By using information from other reported cases with valid EINs, we were only able to reduce these percentages to 32% and 21% respectively. We eliminated from the analysis unlinked cases with invalid EINs. This means that the actual level of underreporting is greater than our estimates. We will return to this issue later in this report.

Establishments in mining and railroad transportation do not report directly to BLS for its annual survey, so their reporting patterns may differ from those of industries reporting directly to BLS. State and local government agencies are not surveyed for BLS national estimates, and data are not collected for the BLS survey by all states. Also, agricultural establishments with fewer than 11 employees are not surveyed for BLS national estimates. The water transportation industry is covered by its own workers' compensation system. Injuries to workers from temporary employment agencies may be reported to BLS by one

entity and to workers' compensation by another.[2] Finally, membership organizations have spotty coverage by workers' compensation programs. We exclude these industries from our analysis, reducing coverage in favor of analyzing industries with relatively uniform reporting requirements. This means that our reporting percentages are only for the private sector and do not reflect reporting in excluded industries.

The BLS system only captures individual case information for injuries involving at least one day away from work, not including the day the injury occurred. This means that medical-only cases that do not involve any lost time beyond the day of injury are not captured by the BLS survey. Limiting our analysis to only cases captured by both systems, we restrict our analysis to lost-time workers' compensation cases, identified by temporary disability benefits reported to WCIS, and we remove all unlinked BLS cases that do not exceed the waiting period. Because medical-only cases are probably less likely to be reported than lost-time cases for the same type of injury, this probably biases our estimates of underreporting toward zero.[3]

Linking BLS and WCIS Injuries

To conduct the capture-recapture analysis, the first step is to determine which injuries have been reported to both systems, which only to BLS, and which only to WCIS. This involves attempting to link each injury in one system with an injury in the other. Such a linkage could not be done by using an error-free common unique identifier, because no such identifier exists. Instead, it was necessary to use data elements common to both systems, understanding that matching on one or two characteristics, like name and date of birth, often will not produce a unique, correctly linked pair. We also recognized that the information in the records was often inaccurate. Names were misspelled, nicknames were used, month and day were switched, first and last name were reversed, to describe a few types of errors that occurred in both systems. Linking all the records would be virtually impossible if it all had to be done manually, because each possible BLS-WCIS pair would have to be looked at. This would involve over 15 billion possible pairs. Luckily, there are software programs that can link records from large datasets relatively easily. We used one of these programs, Linkpro 3.0 (© InfoSoft, Inc.). Using Linkpro 3.0, we linked workers' compensation claims to BLS-reported injuries using data elements common to both sources, including employer identifier, employer name, employer address, employer zip code or city, worker's first initial, worker's last name, sex, date of injury, and date of birth or age at injury. We first linked records deterministically, considering two injuries to be matched if identical on eight of these items. By inspecting linked injury records, we determined that all records linked in this way appeared to be true matches. The deterministic linkage procedure accounted for 80% to 85% of all linked cases.

---

[2] Excluding temporary employment agencies may still leave temporary employees reported by the employer to WCIS but not to BLS.

[3] However, medical-only injuries may be more likely to be traumatic injuries and lost-time injuries may be more likely to involve cumulative trauma. If cumulative trauma injuries are less likely to be reported, there may be a bias in the other direction.

We then needed to check to see how many injuries common to both systems were missed by this automated linkage procedure. To do this, we again used the Linkpro 3.0 software to rank BLS-WCIS pairs likely to refer to the same injury using a method called probabilistic record linkage. Within the potentially linked pairs that the program created, two coders determined whether cases are linked. Concordance between coders measured by Cohen's kappa was greater than 0.7. Where the coders differed, we applied a decision rule to determine whether a pair was linked.

About 285,000 reported lost-time injuries are represented by the analyzed WCIS samples, 156,000 from 2003, and 129,000 from July 1, 2004, through June 30, 2005. This is about 18% of all reported cases and about 88% of indemnity cases reported as of March 15, 2006. Roughly the same number of reported injuries is represented by the analyzed BLS sample. These numbers reflect private sector lost-time injuries excluding mining, railroad and water transportation, temporary employment agencies, membership organizations, and agricultural establishments with fewer than 11 employees.

Estimates of Underreporting

Estimates based on independent reporting

Using capture-recapture analysis, we estimated the total number of cases and the proportion of cases reported, assuming that reporting to BLS and to WCIS is independent. The estimates were initially done under the assumption that reporting rates do not vary by worker, employer, or injury characteristics. A second set of estimates was done allowing reporting rates to be affected by employment size categories, 1-digit industry code (NAICS), age, sex, job tenure, and categories for part of body injured and nature of injury. The estimates changed very little. We provide estimates for two time periods: injuries occurring in 2003 and those occurring from July 1, 2004, to June 30, 2005. We chose these periods to see if WCIS reporting rates changed after the passage of the 2004 workers' compensation reforms. The results are displayed in Table 5.

Table 5. Lower-Bound Estimates of Workplace Injury and Illness Underreporting in California: Lost-time Cases, 2003 and 2004-5, Independent Reporting

	2003	2004-5
Percent not reported to WCIS	25%	21%

Note: These percentages are statistically different at the 95% confidence level.

This very conservative estimate of workers' compensation underreporting for lost-time injuries in the private-sector sampled industries is 25% for 2003 and 21% for 2004-5.

The estimated underreporting rate in 2004-5 is somewhat lower than the rate in 2003, and it is. Thus, we have no evidence that the 2004 reforms led to less reporting of compensable lost-time injuries. In fact, reporting seems to have increased a little. This also may be related to better compliance with workers' compensation reporting regulations as claims administrators and insurers gained experience with WCIS, which began in 2000.

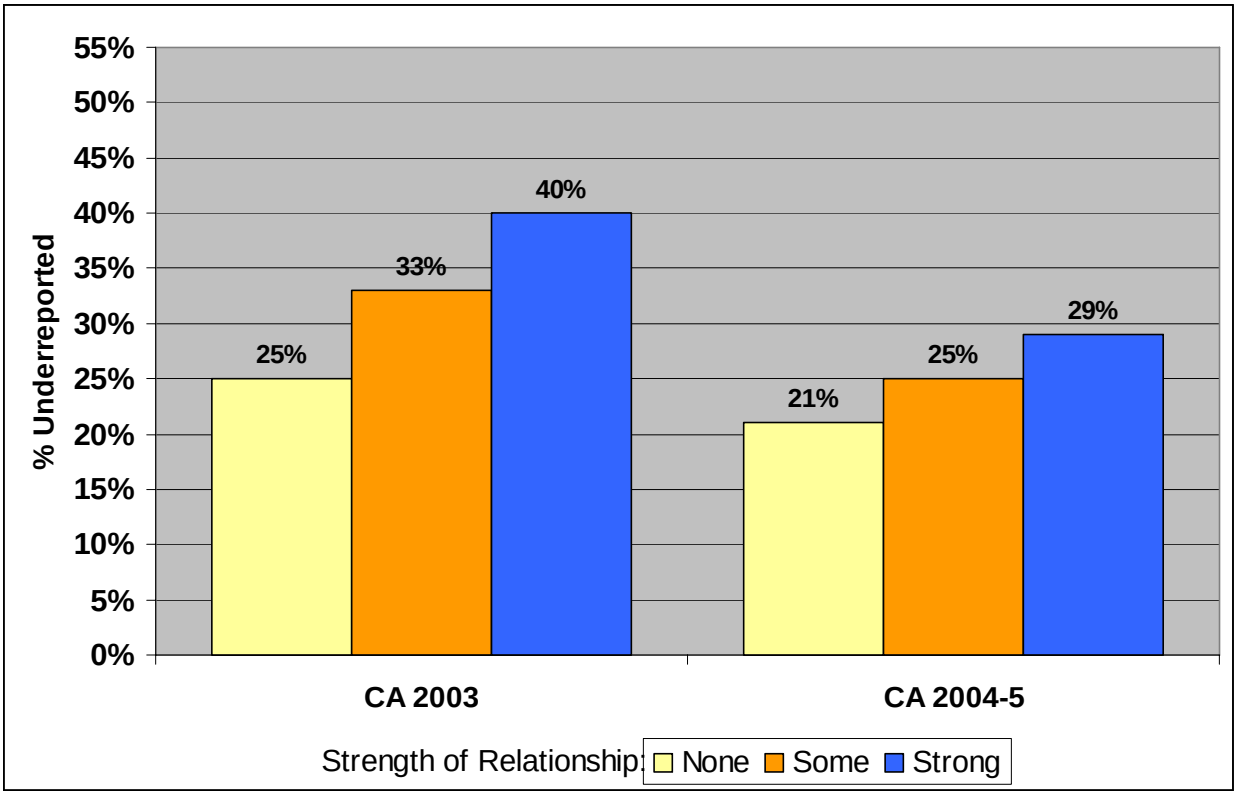
WCIS 2003 data provided to us had 797,084 cases[4], of which 173,587 were indemnity cases. The 2004-5 data had 756,709 cases, including 153,311 indemnity cases. If we apply these estimates to the data, WCIS did not receive reports on almost 60,000 indemnity cases for 2003 injuries and over 40,000 indemnity cases between July 1, 2004, and June 30, 2005. Underreporting is likely to be greater for cases not involving indemnity payments, so applying our estimates to all cases will lead to an even greater underestimate of the number of cases not reported to WCIS. Still, these estimates are very large, amounting to over 265,000 cases in the earlier period and over 200,000 cases in the later period. It is, however, encouraging that the estimated number of unreported cases declined between the two periods.

Estimates based on a positive relationship between BLS and WCIS reporting

The estimates in Table 5 are based on the assumption that reporting to WCIS is independent of reporting to BLS. However, we are confident that injuries reported to BLS are more likely to be reported to WCIS. Unfortunately, using information from these two sources, we cannot estimate how much reporting to the two sources is tied. Because we lack a measure of the strength of this relationship, we provide two additional estimates based on different scenarios representing the strength of the relationship between reporting to BLS and to WCIS.

[4] Many of the cases provided to us had no reported indemnity or medical payments. For example, as of April 24, 2008, more than 46% of cases in WCIS had no recorded payments.

Figure 1. Estimates of Underreporting for California: Sensitivity to the Relationship between BLS and WCIS Reporting



*Positive relationship means that BLS reporting is more likely for cases reported to WCIS. It is measured by the odds ratio, the ratio of the odds that an injury is reported to workers' compensation if it is reported to BLS to the odds that an injury is reported to workers' compensation if it is not reported to BLS. This ratio equals one for no relationship, 3 for "Some," and 5 for "Strong." Confidence intervals increase with odds ratios. For OR=5, 95% confidence intervals extend less than 5 percentage points from the point estimate. Full tables can be found in Appendix 2.

We can see in Figure 1 that the stronger the positive relationship between workers' compensation and BLS reporting, the greater the reduction the estimates of workers' compensation reporting. With some positive relationship between reporting in the two systems, the estimate of WCIS reporting is 67% in 2003 and 75% in 2004-5. With a stronger positive relationship, the estimate of WCIS reporting is 60% in 2003 and 71% in 2004-5. This translates into about 116,000 unreported indemnity cases in 2003 and 63,000 in July 1, 2004-June 30, 2005.

Sensitivity to Missing or Invalid Employer Identifiers

We have noted that 38% of EINs in 2003 and 25% of EINs in 2004-5 were missing or invalid. We were able to link WCIS and BLS cases despite the fact that we lacked EINs. This is because we had a substantial amount of personal and injury information (name, date of birth, date of injury, and so on) to use in linking cases. On the other hand, unlinked WCIS cases (cases reported to WCIS but not BLS) were only used in this analysis if they had an employer identifier in the BLS sample. Unlinked cases with invalid and missing EINs were discarded. So the number of unlinked WCIS cases is probably much larger than the number we used for our estimates.[5]

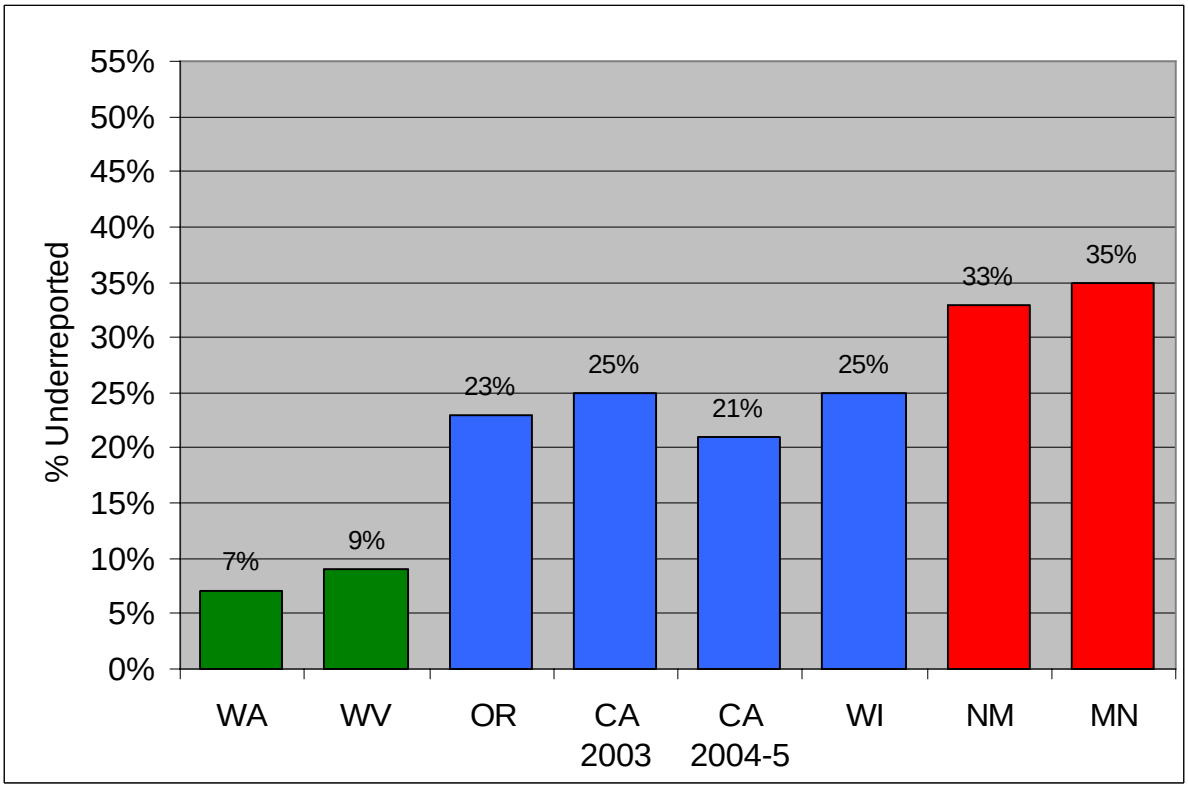
If we had been able to include the discarded unlinked WCIS cases in our analysis, it would have had only a small impact on the estimates of WCIS reporting rates, although it would have substantially reduced estimated BLS reporting rates. (For more details, see Appendix 1).

California Reporting Compared with Reporting in 6 Other States

In 2007, we completed a study of reporting in 6 other states: Minnesota; New Mexico; Oregon; Washington; West Virginia; and Wisconsin (Boden and Ozonoff 2008). This study focused on injuries that occurred during 1998-2001, a period several years earlier than the dates used to study California reporting. However, the methods in both studies are the same, and there was no clear trend in reporting from 1998 to 2001. As a result, we think that it is reasonable to compare reporting in California with reporting in the other states. This makes the comparison seem reasonable. (On the other hand, BLS changed the way it gathered its workplace injury data in 2002, and this means that some caution should be used in comparing reporting rates before and after 2002.)

[5] The insurer or TPA name and EIN is sometimes provided instead of the employer name and EIN. This also reduces the number of unlinked WCIS cases.

Figure 2. Estimates of Underreporting of Workplace Injury and Illness in 7 States: Lost-time Cases, Independent Reporting to BLS and WCIS

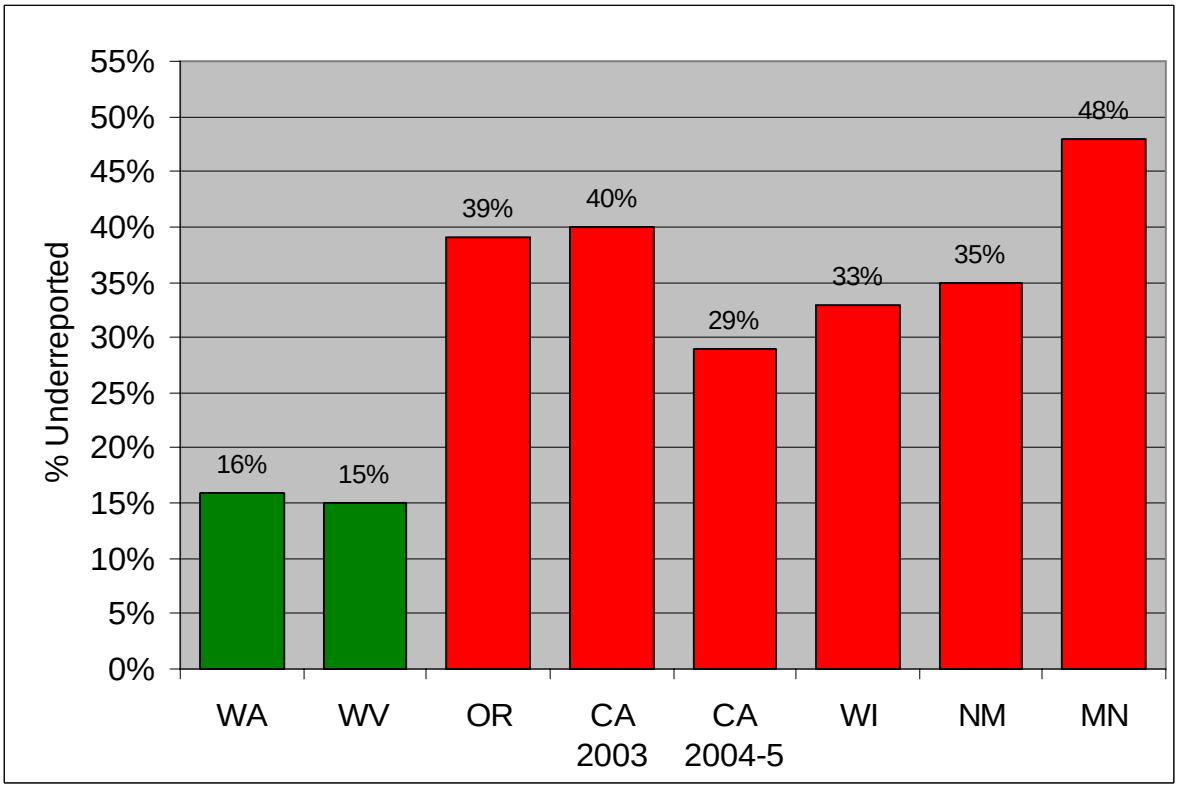


†New Mexico has a 7-day waiting period. The other states have 3-day waiting period.

Note: Underreporting in Washington, West Virginia, Oregon, Wisconsin, New Mexico, and Minnesota was estimated for injuries occurring in 1998-2001. Full tables can be found in Appendix 2.

Figure 2 shows that reporting in California is far from the best states studied (Washington and West Virginia), but also not among the worst of the states studied (New Mexico and Minnesota). California's ranking among the 7 states does not change when we estimate reporting based on the two alternate scenarios (odds ratios of 3 and 5) representing the strength of association between reporting to BLS and to WCIS. However, as noted above, once we allow for a positive relationship between BLS and workers' compensation reporting, estimated underreporting rates increase. This is shown in Figure 3.

Figure 3. Estimates of Underreporting for 7 States: Strong Relationship between BLS and WCIS Reporting



*A strong relationship means that BLS reporting is 5 times as likely for cases reported to WCIS than for those not reported to WCIS.
†New Mexico has a 7-day waiting period. The other states have 3-day waiting periods.

Note: Full tables can be found in Appendix 2.

WCIS Data Problems

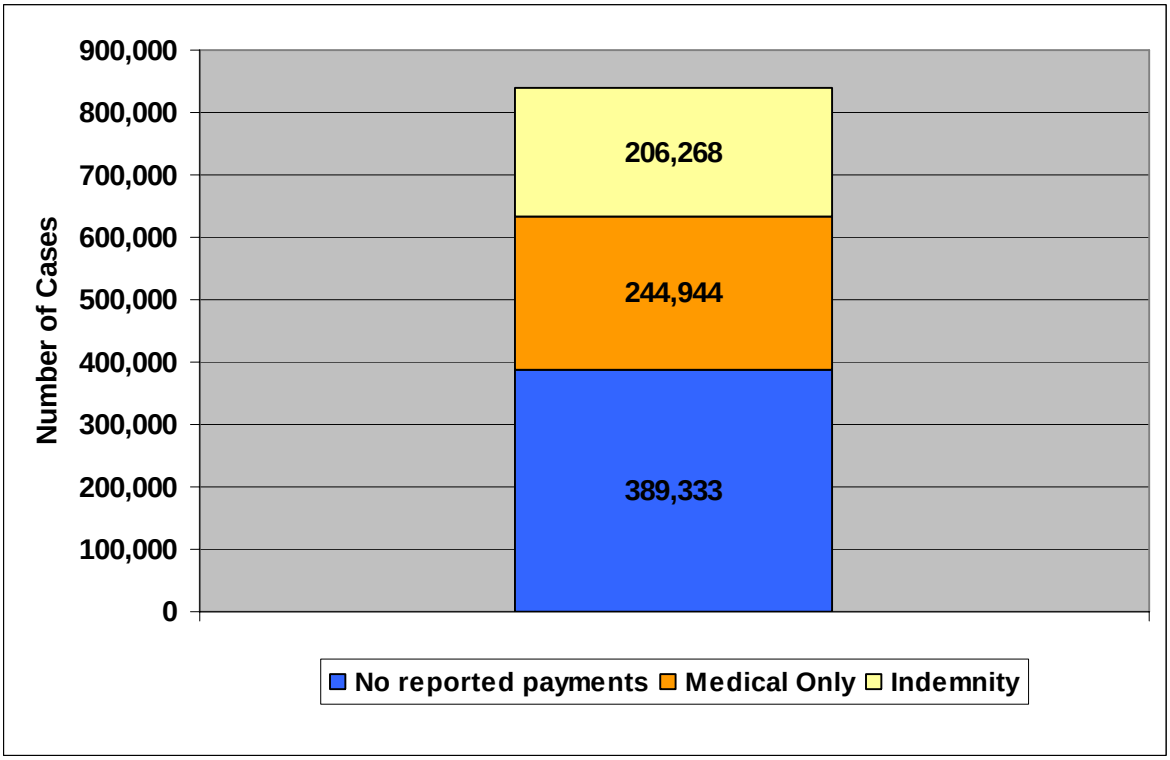
Cases lacking payment information

We have noted that many cases in the WCIS database have missing or invalid EINs. Even more important, almost half of WCIS cases lack information on benefit payments (Figure 4). Logically, these are either cases without medical or indemnity payments or cases for which medical or indemnity payments have been paid but not reported to WCIS.

It seems unlikely that there were almost 400,000 injuries in 2003 for which no benefits were paid. We infer that benefit payments were frequently made but not reported to WCIS. However, we cannot tell how many of these 389,533 cases were medical-only or indemnity cases. This appears to be a major data-quality problem, and one that has been recognized by the DWC for years (DWC 2005).

The lack of benefit payments in so many cases also raises a related question: Is benefit information complete on those cases for which benefits were reported with at least one Subsequent Report of Injury? There is currently no information available to answer this question.

Figure 4. WCIS Cases by Reported Benefit Category (2003 Injury Dates)*



*WCIS data as of April 24, 2008

Delayed reporting

Another issue that we observed during the conduct of this study is the late reporting of injuries. DWC provided us with information on 2003 injuries on March 15, 2006, 27-1/2 months after the end of the 2003 injury year. DWC again provided us with a snapshot of 2003 injuries on April 24, 2008. Between these two dates, the total number of reported cases increased by 5.5%, and the total number of reported lost-time cases increased by 18.8%. The increase in the number of reported cases was 43,461, and the increase in lost-time cases was 32,681. About 4,200 lost-time cases were newly reported. The rest were initially reported as medical-only cases but later reported as receiving indemnity benefits.[6]

[6] The July 1, 2004-June 30, 2005 cases were downloaded by DWC on September 12, 2007. This is 26-1/2 months after June 30, 2005. This means that the window for reporting was similar for the two time periods used in this study.

Although delayed injury reporting is an important issue, we cannot tell whether including injuries reported after March 15, 2006, would reduce our estimates of underreporting. We identified unreported WCIS cases by locating BLS cases that were not reported to WCIS. Late reporting to WCIS would only affect our estimates if the late-reported cases had been reported to BLS. However, employers provide their reports to BLS shortly after the end of the injury year, in this case shortly after the end of 2003. If employers had the information about the late-reported cases at the end of 2003 and reported to BLS but took over two years to deliver it to WCIS, then this would cause our estimates to understate the proportion of injuries reported to WCIS. It seems more likely that the parties reporting to WCIS received this information long after the end of 2003.

Alternatively, employers may initially have denied that the injuries were work-related but eventually have paid benefits. These injuries would probably not have been reported to BLS. For these reasons, we conclude that late reporting probably did not affect our estimates of WCIS underreporting. However, late reporting would increase the *actual level* of WCIS underreporting, and therefore would bias our estimates of underreporting toward zero.

Conclusions and Discussion

The most conservative estimate of reporting of workplace injuries in California suggests that 21% to 25% of lost-time injuries go unreported to WCIS. Reasonable alternate scenarios allow for the likelihood that reporting an injury to BLS increases the likelihood that it will be reported to WCIS. Under these circumstances, we estimate that only about 2/3 of injuries are reported to WCIS. This incomplete reporting places California in the middle of the pack for the 7 states we have studied.

There appears to have been an increase in reporting from injuries occurring in 2003 to injuries between July 2004 and June 2005. This suggests that the 2004 reforms probably did not lead to a decline in the reporting of injuries to WCIS. We do not know whether this increase is a random fluctuation or a stable change. It is possible that the increase is related to increased experience with the WCIS system. It is a good sign that the proportion of cases with missing or invalid EINs declined between the two reporting periods, but we can draw no firm conclusions from this.

Because we do not need the EIN to link injuries, the large number of missing EINs had at most a minor impact on our estimates of reporting to WCIS. Still, there are other reasons for encouraging better employer reporting of EINs. If EIN reporting were relatively complete, it would be easier for the Workers' Compensation Administration to use this information for employer-based research and evaluation.

Although it seems clear that there is substantial underreporting, this research cannot explain the reasons for underreporting. A different kind of study would be needed to help understand why underreporting occurs. For example, a study that interviewed employers

and workers about injuries reported to BLS but not to WCIS might help us understand the reasons for the discrepancy.

In addition to underreporting, WCIS suffers from inaccurate and incomplete information for those injuries that are reported. Valuable data elements, like the EIN, are frequently not reported or incorrectly reported. Even worse, almost half of all cases lack any benefit information. The DWC has identified many potential uses for WCIS, including evaluating benefit adequacy, analyzing the impact of legislation, analyzing medical costs, timeliness of payment, and so on. (See WCIS 2006.) However, the poor quality of WCIS data limits its value for these and other potential uses.

From a policy perspective, benefit payment is at least as important as injury reporting. We do not know how many workers receive benefits for injuries that go unreported to WCIS. It seems likely that benefits have been paid but not reported in many cases, but evidence about this is inadequate to support an estimate. On the other hand, and this is the more troublesome case, unreported injuries may be eligible for workers' compensation benefits but receive none. In this case, the unpaid workers' compensation benefits pose a burden to the injured workers and their families, health insurance programs, and public and private disability programs. This may have been what happened in the KFM example, described in the Executive Summary. KFM reportedly used an incentive plan that gave all members of a work group a cash bonus if they reported no restricted-duty or lost-time accident cases (Tucker 2005). This type of incentive plan leads to social pressures within work groups discouraging injury reporting. And, of course, when workers do not report injuries, they do not receive workers' compensation benefits.

Finally, we stress again that our estimates of reporting are designed to be higher than the actual level of reporting. This is in part because, when we had a choice of methods, we selected the one that was likely to produce the higher estimate of reporting. In addition, we limited this analysis to lost-time cases.

Recommendations

We have several specific recommendations that could help to identify reporting problems and reduce underreporting. We understand that it may be infeasible to adopt all of our recommendations, but we think that they are a useful starting place for planning. However, our first recommendation is more general.

- CHSWC, DWC, CDPH, DLSR, and Cal/OSHA should convene an interagency underreporting task force to develop a plan for improving WCIS reporting. This would include identifying late reporting, but also identifying employers, insurers, and TPAs that do not report compensated injuries. In doing so, these agencies should convene an interagency underreporting task force to develop a plan for improving WCIS reporting. This task force could include not only knowledgeable people from these agencies, but also people involved in other relevant activities,

like California's reporting to the BLS survey and planning for the California Trauma Registry.

Specific recommendations include:

- DWC could strengthen its efforts to identify problem areas in reporting of compensated injuries. This would include identifying late reporting, but also identifying employers, insurers, and TPAs that do not report compensated injuries. In doing so, DWC may identify problems in the way reporting systems work, in addition to identifying noncompliance with reporting requirements.

- DWC and Cal/OSHA could consider collaborating to identify employers who underreport injuries. Employers who engage in substantial underreporting to either system could be given substantial penalties, and the program and penalties could be publicized. DWC could also consider penalties for late reporting to WCIS. If current laws and regulations are inadequate to support such a program, this could be addressed.

- The DWC could begin an inquiry into the 40-50% of reported claims that lack information about benefit payments. DWC could draw a random sample of such cases with dates of injury at least 3 years in the past from a subset of claims administrators for insurers, TPAs, and self-insured employers. Initially the claims administrators might be chosen because they have a relatively high proportion of cases lacking benefit reports. The DWC could submit the sample to the trading partners and request up-to-date information on benefit payments and claim status. From this information and discussions with trading partners, the DWC may be able to diagnose systematic problems and fashion solutions.

- It may be possible to compare WCIRB data by reporting entity with WCIS data as a way of identifying reporting entities that substantially underreport compensated cases.

- California collects data on hospital and emergency room discharges and from ambulatory surgery clinics through MIRCAL. DIR might explore whether these data could be used to look for unreported workplace injuries and illnesses. The data contain diagnosis and Social Security Number of the patient and identify the expected source of payment. They do not identify the employer. If the WCIS data included state EDD account numbers (EANs), cross-matching with EDD wage files to determine the employer would be easier and more accurate than otherwise. We do not know if there are any legal issues precluding this use of MIRCAL data.

- CHSWC could explore linking other state occupational safety and health information systems with WCIS data to determine whether injuries and illnesses have been reported and compensated where appropriate.

- DIR could explore automating the doctor's first report of occupational injury or illness and requiring all doctor's first reports to be electronically transmitted. For example, reports could be filled out on the Internet and automatically transmitted to DIR. These reports could be compared with WCIS files to determine where underreporting occurs.

- CHSWC has recently funded a study of employer underreporting and misreporting of payroll to workers' compensation insurers (Neuhauser and Donovan 2007). This includes, among others, underreporting payroll, misclassifying the occupation of employees, and misclassifying employees as independent contractors. If an injured worker is misclassified as an independent contractor, the injury is not reported. Also, employers who underreport or misreport payroll may be more likely than others to suppress or simply not report workplace injuries and illnesses. If DWC or Cal/OSHA develops programs to detect WCIS underreporting and enforce reporting requirements, reporting at these employers might be examined first.

- DWC may want to consider rejecting reports of injury with invalid or incorrect EINs. These numbers can be valuable for potential uses of WCIS, including but not limited to the underreporting issue.

- DWC should consider adding the state EAN as a required field in the First Report of Injury. This would allow easier and more accurate linkage with EDD wage files and other state data collected from employers.

- California has recently added workers' compensation questions to the states' BRFSS survey. This could be used as another way of getting a handle on the extent of workers' compensation underreporting. Over time, it could be used to determine whether reporting is improving.

Appendix 1: Capture-Recapture Methods

Linking method

We linked reported injuries in BLS data with those in WCIS. We used deterministic record linkage first. In this step, a minimum number of identical fields in both datasets implied a definite link. For the remaining injuries, we used probabilistic record linkage, which is used by, among others, the U.S. Bureau of the Census. It is also the most common matching method used in linking medical records from different sources. Fellegi and Sunter (1969) and Copas and Hilton (1990), among others, present a formal theory for probabilistic record linkage, which operates by assigning weights to information fields in potentially matched pairs and then developing a combined weight for each pair. A large weight indicates a high probability that both of the pair's records refer to the same event. Jaro (1995) develops weights as follows:

$$\text{agreement weight} = \log_2[\text{sensitivity}/(1-\text{specificity})]$$

$$\text{disagreement weight} = \log_2[(1-\text{sensitivity})/\text{specificity}]$$

Sensitivity is defined as the probability that the field matches given that the pair refers to the same event. Specificity is defined as the probability that the field does not match given that the pair does not refer to the same event. To be a useful field, sensitivity must be greater than (1-specificity). If this is the case, agreement weights are positive and disagreement weights are negative. If several fields were used in the match, each would have its own agreement and disagreement weights. A composite weight for a pair is the sum of its agreement and disagreement weights. Note that agreement weights will always be positive, and disagreement weights will always be negative. Sensitivity and specificity can be calculated iteratively from the lists to be matched (Jaro 1995).

We first established two cut-off points. If a weight was above the high cutoff point, the pair was considered matched. If it was below the low cutoff point, the pair was considered unmatched. Pairs between the two were then hand-matched (on the assumption that hand-matching is the most accurate method available).

Capture-recapture Methods

Capture-recapture methods are used in epidemiology to estimate disease incidence or prevalence from multiple, overlapping, but incomplete sources. Typically, capture-recapture analyses identify all unique cases recorded by at least one source and then use log-linear or logistic models to estimate the number of cases unidentified by any source. The estimate of the missing cases typically is based on the assumption that the sources are independent samples of the target population. The idea behind these models is simple.

Table A1.1 shows hypothetical results of injury surveillance using two data sources, where X is the unknown number of missed cases. If capture by Source 1 and Source 2 is independent, then Source 1 will capture the same fraction of cases among those captured by Source 2 and those not captured by Source 2. That is,

$$N_1 / (N_1 + X) = N_3 / (N_3 + N_2) \text{ or } X = N_1 N_2 / N_3 \quad \text{(A1)}$$

This is the maximum-likelihood estimator of X under the assumptions of independence of ascertainment, uniform probability of ascertainment within both sources, and a closed population (Hook and Regal 1995). If the sources are (positively) dependent, then the probability of being in both sources is greater than the product of the probabilities of being only in one source or the other. Source 1 will capture a greater proportion of cases captured by Source 2 than those not captured by Source 2. That is,

$$N_1 / X < N_3 / N_2 \text{ or } X > N_1 N_2 / N_3 \quad \text{(A1a)}$$

Therefore, Equation (1) will underestimate X. Similarly, if the sources are negatively dependent, the probability of being in both is less than the product of the probabilities of being in either.

Table A1.1. Two-source Capture-Recapture Model

Source 1		Source 2	
	No	Yes	Total
No	X	N_2	$X + N_2$
Yes	N_1	N_3	$N_1 + N_3 = S_1$
Total	$X + N_1$	$N_2 + N_3 = S_2$	$X + N_1 + N_2 + N_3 = S$

To generate unbiased estimates of the total population, capture-recapture studies often make several assumptions in addition to the non-dependence of sources. These include: that the population remains closed (constant) over the observed period; and that each individual has the same probability of capture. Although the assumption of a closed population appears reasonable for non-fatal occupational injury surveillance, capture probabilities may vary by employer, worker, and injury characteristics. For this reason, we estimate the capture probability conditional on these characteristics. This allows us to see whether estimated injury incidence changes when reporting heterogeneity is accounted for.

We account for observed heterogeneity of capture probabilities using logistic regression. Several articles have been published that describe the favorable characteristics of logistic regression models in this context (e.g., Ahlo 1990 and Tilling and Sterne 1999).

We first use the multinomial logistic model used to estimate the probabilities of being in one of the observed cells, conditional on being observed. Thus, there are 3 probabilities to be estimated, p_1, p_2, and p_3, equal respectively to $N_1/ (N_1+ N_2,+ N_3)$, $N_2 / (N_1+ N_2,+ N_3)$, and $N_3 / (N_1+ N_2,+ N_3)$ in Table A1.1. The multinomial logistic model for two-source capture-recapture estimation without covariates is then:

$$\Pr(Y_i = j) = p_{ij} = e^{\beta_{0j}} / \sum_{k=1}^{3} e^{\beta_{0k}}, j = 1,2,3 \qquad \text{(A2)}$$

where Y_i indicates cell membership for the i^{th} observed subject. To identify the model, one of the parameters is set to zero, and the model is then estimated. The parameter estimates then can be used to generate an expected probability for each captured individual to be captured by Source 1, Source 2, or both sources. This also leads to an estimate for the probability of being missed by both sources.

To account for heterogeneity of capture probabilities according to individual, injury, and employer characteristics, the multinomial logistic model can be generalized to include a vector of covariates:

$$\Pr(Y_i = j) = p_{ij} = e^{\beta_{0j} + \beta_j' x_i} / \sum_{k=1}^{3} e^{\beta_{0k} + \beta_k' x_i}, j = 1,2,3 \qquad \text{(A3)}$$

The parameter estimates can be estimated through maximum likelihood and then be used to generate an expected probability for each individual to be captured by Source 1, Source 2, or both sources. Given the assumption of independent capture by both sources for each individual i, we can estimate p_{i0}, the probability of being missed for person i (Tilling and Sterne 1999).

The total population size S is then estimated from the observed sample as:

$$S = \sum_{i \in N_j}^{j=1,3} 1/(1 - p_{i0}) \qquad \text{(A4)}$$

Accounting for Establishment Sampling

BLS surveys a sample of establishments (business units at a single location) within each state. Workers' compensation reporting systems typically report by firm, not by establishment. As a consequence, we can identify firms in the workers' compensation data that report to BLS, but we cannot identify reporting and non-reporting establishments within those firms. For multi-establishment firms (e.g., fast food chains or manufacturers with several plant locations), this differential reporting presents a challenge: in such firms, an injury may be unrecorded by BLS because an establishment failed to report or because the establishment was not included in the BLS subsample. If we did not account for this,

many injuries in establishments not sampled by BLS would be treated as if they were underreported by BLS. This would lead to an underestimate of the proportion of injuries reported by BLS.
To address this, we use quarterly employment information from the BLS longitudinal database (LDB) of establishments and firms to derive for each firm the proportion of employment in establishments reporting to the BLS survey. We assume homogeneity of injury rates and reporting rates across all establishments within a firm. Under this assumption, we impute the number of workers' compensation injuries reported by establishments that are not in the BLS sample and adjust the weights of the unmatched workers' compensation cases accordingly to account for the expected number of unlinked workers' compensation injuries reported by unsampled establishments in multi-establishment firms.

Sensitivity to Source Dependence

Figures 1 and 3 display the impact on underreporting of source dependence, ranging from no source dependence to source dependence reflecting an odds ratio of 5 for reporting to the BLS. Table A1.2 shows the range of estimated impacts on reporting including odds ratios of 1.0, 3.0, and 5.0.

Table A1.2. Estimates of Underreporting for 7 States: Sensitivity to Positive Relationship between BLS and WCIS Reporting

	Odds ratio*	CA 2003	CA 2004-5	WA	WV	OR	WI	NM†	MN
Percent not reported to workers' compensation	1.0	25%	21%	7%	9%	23%	25%	33%	35%
	3.0	33%	25%	12%	12%	33%	33%	47%	43%
	5.0	40%	29%	16%	15%	39%	39%	55%	48%

*Positive relationship means that BLS reporting is more likely for cases reported to WCIS. 95% confidence intervals extend less than 5 percentage points from the point estimate.
†New Mexico has a 7-day waiting period. The other states have 3-day waiting periods.

Sensitivity to Missing Employer Identification Numbers

We can get a rough idea of the likely effect on our estimates of dropping unlinked workers' compensation cases when there was a missing or invalid EIN. Assume that the probability of a missing or invalid EIN is the same for employers in the BLS survey and for other employers. Then, for 2003 injuries, if there were 100 injuries reported to WCIS but not reported to BLS, 32 would have missing or invalid EINs. For the estimates in Figure 1, only 68 would be counted. What happens to our estimates if we increase the number of unlinked WCIS cases to account for this? We do this in Table A1.3, where we multiply the number of unlinked WCIS cases used in the estimates in Table A1.2 by 100/68.

Table A1.3. Estimates of Underreporting for California: Sensitivity to Missing and Invalid EINs

	Odds ratio*	2003 Estimate	2003 Additional WCIS Cases	2004-5 Estimate	2004-5 Additional WCIS Cases
Percent not reported to WCIS	OR=1	25%	25%	21%	21%
	OR=3	33%	35%	25%	26%
	OR=5	40%	43%	29%	31%

* We use the odds ratio as a measure of the relationship of reporting to WCIS and reporting to BLS. This ratio equals one for source independence and is greater than one for positive source dependence. An odds ratio of 1 means that reporting to BLS is just as likely for injuries reported to WCIS as for those that are not. An odds ratio of 3 means that injuries reported to WCIS have 3 times the odds of being reported to the BLS when compared to injuries that are not reported to WCIS.

Note: Full tables can be found in Appendix 2.

Appendix 2: Detailed Tables

Table A2.1. Capture-Recapture Estimates of Workplace Injury Reporting: Lost-Time Cases, No Source Dependence

		California 2003		
		Workers' Compensation		
		No Report	Report	Total
BLS	No Report	6%	19%	25%
	Report	18%	56%	75%
	Total	25%	75%	100%

		California 2004-5		
		Workers' Compensation		
		No Report	Report	Total
BLS	No Report	3%	10%	13%
	Report	18%	69%	87%
	Total	21%	79%	100%

Washington

		Workers' Compensation		
		No Report	Report	Total
BLS	No Report	2%	45%	48%
	Report	3%	50%	52%
	Total	5%	95%	100%

West Virginia

		Workers' Compensation		
		No Report	Report	Total
BLS	No Report	2%	23%	25%
	Report	6%	68%	75%
	Total	8%	92%	100%

Oregon

		Workers' Compensation		
		No Report	Report	Total
BLS	No Report	10%	35%	44%
	Report	12%	44%	56%
	Total	22%	78%	100%

Wisconsin

Workers' Compensation

		No Report	Report	Total
BLS	No Report	10%	27%	37%
	Report	17%	46%	63%
	Total	27%	73%	100%

New Mexico

Workers' Compensation

		No Report	Report	Total
BLS	No Report	17%	34%	52%
	Report	16%	32%	48%
	Total	34%	66%	100%

Minnesota

Workers' Compensation

		No Report	Report	Total
BLS	No Report	11%	21%	32%
	Report	24%	44%	68%
	Total	35%	65%	100%

Note: Rows and columns may not add correctly because of rounding. 95% confidence intervals extend less than 1.5 percentage points from the point estimates for all states but New Mexico and California, where they extend less than 4 percentage points. Data for states other than California are for 1998-2001 injuries.

Table A2.2. Capture-Recapture Estimates of Workplace Injury Reporting: Lost-Time Cases, Estimate Accounting for Missing and Invalid EINs

California 2003

		Workers' Compensation		
		No Report	Report	Total
BLS	No Report	8%	25%	33%
	Report	16%	50%	67%
	Total	25%	75%	100%

California 2004-5

		Workers' Compensation		
		No Report	Report	Total
BLS	No Report	3%	12%	16%
	Report	18%	66%	84%
	Total	21%	79%	100%

REFERENCES

Alho, JM. Logistic regression in capture-recapture models. *Biometrics*. 46:623-635. 1990.

Azaroff LS, Levenstein C, Wegman DH. Occupational injury and illness surveillance: Conceptual filters explain underreporting. Am J Public Health. 2002;92:1421–1429.

Belin TR, Rubin DB. A method for calculating false-match rates in record linkage. *Journal of the American Statistical Association* 90(430):694-707. 1995.

Biddle J, Roberts K, Rosenman D, Welch E. What percentage of workers with work-related illnesses receive workers' compensation benefits. *Journal of Occupational and Environmental Medicine* 40(4):325-31, 1998.

Biddle J, Roberts K. Claiming behavior in workers' compensation. J Risk Insurance. 2003;70:759–780.

Boden LI, Gold M. The accuracy of self-reported regulatory data: The case of coal mine dust. *American Journal of Industrial Medicine* 6:427-440, 1984.

Boden LI, Ozonoff A. Capture-Recapture Estimates of Nonfatal Workplace Injuries and Illnesses. Ann Epidemiol. In press, 2008.

Boden LI, Ruser J. Choice of Medical Care Provider, Workers' Compensation "Reforms," and Workplace Injuries. In review: *Review of Economics and Statistics.* 2002.

Brenner H. Uses and limitations of the capture-recapture method in disease monitoring with two dependent sources. *Epidemiology* 6:42-48. 1994.

Burton, John F. and Emily A. Spieler. Workers' Compensation and Older Workers. In: H. Allan Hunt, Ed. *Ensuring Health and Income Security for An Aging Workforce*. (Washington, DC: National Academy of Social Insurance, 2001) 41-84.

Cormack RM, Chang YF, Smith GS. Estimating deaths from industrial injury by capture-recapture: a cautionary tale. *International Journal of Epidemiology. 29(6):1053-9, 2000*

Copas J, Hilton F. Record linkage: Statistical models for matching computer records. *Journal of the Royal Statistical Society*, Ser. A. 153(3):287-320, 1990.

California Division of Workers' Compensation. WCIS reporting problems. http://www.dir.ca.gov/dwc/WCISenews/WCISen51.htm December 29, 2005.

California Division of Workers' Compensation. Uses of the WCIS database. http://www.dir.ca.gov/dwc/WCISenews/WCISen59.htm August 7, 2006.

Eisenberg WM, McDonald H. Evaluating workplace injury and illness records; testing a procedure. *Monthly Labor Review* 111(4):58-60. 1988.

Fan ZJ. Bonauto DK. Foley MP. Silverstein BA. Underreporting of work-related injury or illness to workers' compensation: individual and industry factors. *J Occup Environ Med.* 2006; 48(9):914-22.

Fellegi IP, Sunter AB. A theory for record linkage. *Journal of the America Statistical Association* 64(328):1183-1210. 1969.

Fine LJ, Silverstein BA, Armstrong TJ, Anderson CA, Sugano DS. Detection of cumulative trauma disorders of upper extremities in the workplace. *Journal of Occupational Medicine* 28(8):674-78, 1986.

Hanrahan, LP. A comparison of the BLS annual survey to workers' compensation for Wisconsin in 1984 and 1985. Appendix E of: Panel on Occupational Safety and Health Statistics, Commission on Behavioral and Social Sciences and Education. *Counting Injuries and Illnesses in the Workplace: Proposals for a Better System*. Washington DC: National Academy Press. 155-171. 1987.

Hook EB, Regal, RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiologic Reviews* 17(2):243-64. 1995.

International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation II: Applications in human diseases. *American Journal of Epidemiology. 142(10):1059-68, 1995.*

Jaro MA. Probabilistic linkage of large public health data files. *Statistics in Medicine.* 14(5-7):491-8, 1995.

Lakdawalla D, Reville RT. When and why do workers file for workers' compensation? Econ Inquiry. 2007;45: 286-303.

Maizlish N, Rudolf L, Dervin K, Sankaranarayan M, 1995. Surveillance and prevention of work-related carpal tunnel syndrome: an application of the sentinel events notification system for occupational risks. *American Journal of Industrial Medicine* 27:715-29.

Messiou E, Zaidman B. *Comparing Workers' Compensation Claims and OSHA Data Initiative Cases*. St. Paul, MN: Minnesota Department of Labor and Industry. 2005.

Morse T, C Dillon, N Warren, C Levenstein and A Warren. 1998. The Economic and social consequences of work-related musculoskeletal disorders: The Connecticut Upper-Extremity Surveillance Project (CUSP). *International Journal of Occupational and Environmental Health* 4:209-216.

Morse TF C Dillon, N Warren, C Hall, and D Hovey. 2001 Capture-recapture estimation of unreported work-related musculoskeletal disorders in Connecticut, *American Journal of Industrial Medicine*, 39(6):636-642.

Morse T, Dillon C, Kenta-Bibi E, Weber J, Diva U, Warren N, Grey M. Trends in work-related musculoskeletal disorder reports by year, type, and industrial sector: a capture-recapture analysis. *Am J Ind Med.* 2005*;* 48(1):40-9.

McCurdy SA, Schenker MB, and Samuels SJ. Reporting of occupational injury and illness in the semiconductor manufacturing industry. *American Journal of Public Health* 81(1):85-89. 1991.

Panel on Musculoskeletal Disorders and the Workplace, Commission on Behavioral and Social Sciences and Education. *Musculoskeletal Disorders and the Workplace: Low Back and Upper Extremities*. Washington DC: National Academy Press. 2001.

Rosenman KD, Reilly MJ, Henneberger PK. Estimating the total number of newly-recognized silicosis cases in the United States. *Am J Ind Med.* 2003; 44:141-147.

Ruser, JW, Smith RS. The effect of OSHA records-check inspections on reported occupational injuries in manufacturing establishments. *Journal of Risk and Uncertainty*. 1(4):415-435, 1988.

Seber GA, Juakau JT, Simmons D. Capture-recapture, epidemiology, and list mismatches: two lists. *Biometrics.* 56(4):1227-32, 2000

Seligman PJ, Sieber WK, Pedersen DH, Sundin DS, Frazier TM. Compliance with OSHA record-keeping requirements. *American Journal of Public Health. 78(9):1218-9, 1988*

Silverstein BA, Stetson DS, Keyserling WM, Fine LJ. Work-related musculoskeletal disorders: comparison of data sources for surveillance. *American Journal of Industrial Medicine* 31(5):600-8. 1997.

Stout N, Bell C. Effectiveness of source documents for identifying fatal occupational injuries: a synthesis of studies. *American Journal of Public Health.* 81(6):725-8, 1991.

Tilling K, Sterne JAC. Capture-recapture models including covariate effects. *American Journal of Epidemiology*. 149(4):392-400, 1999.

Tucker, J. State: Bridge builder fudged injury reports. *Oakland Tribune*. June 2, 2006. http://findarticles.com/p/articles/mi_qn4176/is_20060602/ai_n16454535