

# **Allegations of Undercounting in the BLS Survey of Occupational Injuries and Illnesses** October 2010

John W. Ruser, Ph.D.<sup>1</sup>

US Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Washington, DC 20212

## **Abstract**

Recent external studies find that the Bureau of Labor Statistics (BLS) Survey of Occupational Injuries and Illnesses (SOII), based on employers' OSHA logs, substantially underestimates the US total number of workplace injuries and illnesses. This paper describes four dimensions of the undercount allegations and summarizes the recent studies, which have largely compared aggregate or case-level data from various sources, augmented in the case-level studies with capture-recapture analysis. The paper provides an overview of on-going research conducted and funded by BLS to better understand the nature of the undercount findings. These studies include counting work-related amputations and carpal tunnel syndrome cases with multiple data sources; matching SOII to workers' compensation records; and, interviewing employers about reporting practices on OSHA logs and to workers' compensation.

**Key Words:** workplace injuries and illnesses; capture-recapture; data matching

## **1. Introduction and background**

National estimates of nonfatal workplace injuries and illnesses are produced by the Bureau of Labor Statistics (BLS) Survey of Occupational Injuries and Illnesses (SOII), a comprehensive statistical program covering private industry and State and local government. The survey information is unique and of great value to the safety and health community in allocating prevention resources among several hundred diverse industries and occupations, across which workers' risks of injury and illness vary widely. For injuries and illnesses with days away from work, the survey provides details concerning the characteristics of injured workers and the circumstances of the injury and illness cases.<sup>2</sup> These data are critical to designing prevention strategies to protect workers. The survey also produces estimates by State for those States that participate with the BLS in conducting the survey, providing a uniform set of injury and illness surveillance information across States.

Survey data for SOII are collected from responding employers, who draw information from Occupational Safety and Health Administration (OSHA) logs and supplementary materials maintained by these employers throughout the year. Employers included in the

---

<sup>1</sup> This paper represents the views of the author and may not represent the views or policies of the Bureau of Labor Statistics or the US Department of Labor.

<sup>2</sup> Injuries and illnesses with days away from work are those with a loss of one or more workdays after the day of occurrence of the injury or illness.

survey may or may not otherwise be required by OSHA to keep these records. However, according to regulations, they are required to keep records and to respond to SOII when notified by the BLS that they will be included in the survey. To ensure that workplace injuries and illnesses are recorded throughout the year, BLS pre-notifies employers of their inclusion in the SOII prior to the start of the survey year. This pre-notification makes it more likely that employers will have recordkeeping in mind at the start of the collection year, and that they will have the tools to record injuries and illnesses correctly. The SOII is separate from other systems for recording workplace injuries and illnesses, including workers' compensation, trauma registries and other administrative and survey data sources.

The SOII and the OSHA recordkeeping system underpinning it have come under strong criticism in Congress and elsewhere for failing to completely record workplace injuries and illnesses. For example, a majority staff report issued in June 2008 by the Committee on Education and Labor of the U.S. House of Representatives concluded that "extensive evidence from academic studies, media reports and worker testimony shows that work-related injuries and illnesses in the United States are chronically and even grossly underreported." (US House of Representatives 2008)

The basis of the criticism is a number of recent non-BLS research studies that have compared SOII data to other data sources (some involving workers' compensation case data). The comparisons, both of macro-level estimates and of individual injury and illness case data, generally conclude that the SOII substantially undercounts all workplace injuries and illnesses. Estimates of the undercount range widely from 20 percent to 70 percent of all cases in some research. However, other research and analysis concludes that the size of the undercount is small. Interestingly, studies that find an undercount conclude that both SOII and other data sources each miss injury and illness cases, leading to the conclusion that no single source of data can completely enumerate all cases.

While the undercount studies provide some evidence to explain differences in coverage of cases in the SOII and other data, additional information about differences in the data systems for capturing workplace injuries and illnesses is necessary to understand the SOII undercount. To this end, BLS is conducting intramural research and supporting extramural research to identify factors that are associated with SOII completeness. Three projects are being undertaken: analysis of SOII case data (microdata) matched to workers' compensation cases, interviews of employers about OSHA recordkeeping and workers' compensation filing, and enumeration of work-related amputations and carpal tunnel syndrome cases using multiple data sources (SOII, workers' compensation, hospitalizations, emergency department visits, and others).

This article discusses the SOII undercount issue. After providing a brief overview of the SOII and some alternative data systems, it describes four different dimensions of the potential undercount and summarizes some of the key papers in the undercount literature. Finally, the article summarizes research that BLS is conducting and supporting that is aimed at addressing the undercount issue.

## **2. Data collection**

SOII is a Federal-State program in which employers' reports are collected annually from over 200,000 private-industry and State and local government establishments. Data are

collected starting in January after the end of the survey reference year. Employers who respond to the SOII questionnaire provide information on the number of workplace injuries and illnesses by copying the data from their Occupational Safety and Health Administration (OSHA) records to the SOII. Employers also provide the number of employee hours worked (needed in the calculation of incidence rates) as well as the establishments' average employment.

Besides reporting injury and illness counts, survey respondents are asked to provide additional information for cases involving at least 1 day away from work beyond the day of injury or onset of illness. Employers answer several questions about these cases, including the demographics of the worker disabled, the nature of the disabling condition, and the event and source producing the condition. This information is largely obtained from supplementary OSHA recordkeeping forms.

To minimize the burden on many larger employers, sampled establishments that are projected to have numerous cases involving days away from work are instructed to report on a sample of those cases. These employers are assigned a range of dates and are instructed to provide information only on the cases with days away from work for which the date of injury or onset falls within the assigned range of dates.

Injuries and illnesses logged by employers conform to definitions and recordkeeping guidelines set by the Occupational Safety and Health Administration, U.S. Department of Labor. (See box.) Under these guidelines, nonfatal cases are recordable if they are work-related illnesses or injuries that involve lost worktime, medical treatment other than first aid, the restriction of work, loss of consciousness, a transfer to another job, or other specific conditions. Employers keep counts of injuries separate from counts of illnesses. They also identify whether each injury or illness involved any days away from work, days of restricted work activity, or both that occurred after the day of injury or onset of illness.

All employers with 11 or more employees in OSHA-designated high-hazard industries are required by OSHA regulation 29 Code of Federal Regulations (CFR) 1904 to maintain logs throughout the year and to complete a summary based on the log at the end of the year. Other employers also are required to maintain logs according to OSHA regulation 29 CFR 1904.42 in the event that those employers are asked to participate in the SOII. BLS draws a sample of employers for SOII from both OSHA-designated high hazard industries and other industries.

The SOII excludes all work-related fatalities, for which data are collected in the BLS Census of Fatal Occupational Injuries. The SOII also excludes nonfatal work injuries and illnesses incurred by the self-employed, workers on farms with 10 or fewer employees, private household workers and Federal government workers. These groups of workers are outside OSHA jurisdiction. The SOII receives occupational injury and illness data from the U.S. Department of Labor Mine Safety and Health Administration for establishments in the coal, metal, and nonmetal mining industries and data from the U.S. Department of Transportation Federal Railroad Administration for railroad incidents.

Over 75 percent of SOII responses are now collected electronically, largely by means of the BLS Internet Data Collection Facility. Electronic responses are also obtained from an Adobe fillable form and electronic file submissions.

### OSHA case recordability criteria

OSHA guidelines for recording cases are codified in 29 CFR (Code of Federal Regulations) 1904. In general, recordable cases include new work-related cases of injuries and illnesses or the significant work-related aggravation of preexisting non-work-related conditions. Cases are recordable if they result in

- death
- loss of consciousness
- days away from work
- restricted work activity or job transfer
- medical treatment (beyond first aid)
- significant work-related injuries or illnesses that are diagnosed by a physician or other licensed health care professional, including cancer, chronic irreversible disease, a fractured or cracked bone, and a punctured eardrum

Cases also are recordable if they meet additional criteria for special cases; cases that qualify include those involving needlesticks and “~~sharp~~” injuries, occupational hearing loss, and tuberculosis. The regulations provide definitions of many key concepts, explaining how to determine whether a case is work related, what is a new case, what is involved in a significant aggravation of a preexisting condition, what is restricted work, and so forth.

Occupational injuries, such as sprains, cuts, and fractures, account for the vast majority of all cases that employers log and report to the BLS survey. Occupational illnesses are new cases recognized, diagnosed, and reported during the year. Overwhelmingly, illness cases that are reported are easier to relate directly to workplace activity (for example, contact dermatitis) as opposed to long-term latent illnesses, such as cancers.

The SOII provides estimates that are based on a scientifically selected sample of establishments, some of which represent only themselves but most of which also represent other employers of like industry and workforce size that were not chosen in a given survey year. As such, the estimates are subject to sampling error and the BLS publishes estimates of those sampling errors. The data are also subject to non-sampling errors that are not measured. These errors include the unavailability of characteristic data for some cases, mistakes in recording or coding the data, and definitional difficulties. To minimize non-sampling errors, the Bureau conducts a rigorous training program for survey coders and continues to encourage survey participants to respond fully and accurately to all survey elements.

An important advantage of SOII is that it is a large system that provides the most complete occupational injury and illness counts for the Nation and does so consistently across States. But, there are other data systems that provide estimates of occupational injuries and illnesses. While it is beyond the scope of this paper to discuss other systems in detail, a brief summary of some of them is necessary, because it is comparisons between SOII and the other systems that provide the basis for the undercount estimates.

One data system against which SOII data have been compared extensively is workers' compensation insurance. Each State has its own workers' compensation system to cover injured and ill workers. The systems vary somewhat but have the same general

characteristics. With the exception of Texas, all States mandate coverage of nearly all private-sector workers. Some States exempt from coverage workers in very small companies, certain agricultural workers, and some other categories of workers. All State laws require that employers cover nearly 100 percent of an injured or ill worker's medical expenses and further require that workers who are off work longer than a specified —waiting period” are paid cash benefits related to lost earnings. States differ in the durations of their waiting periods, which range from 3 days to 7 days, and also differ to a small extent regarding which cases are compensable. Recently, for example, a number of States passed legislation requiring that work be a major or predominant cause of the disability. Some states eliminated compensation for the aggravation of a preexisting condition or for a condition related to the aging process.

Despite the fact that there is a workers' compensation system in each State, national estimates of occupational injuries and illnesses are difficult to derive from workers' compensation records due to lack of comparability across States. For example, some workers' compensation databases can provide estimates only of cases for which workers are off work for longer than the particular State's waiting period. Thus, there are differences in scope between workers' compensation and SOII data with which researchers must contend in trying to reconcile estimates between the two systems.

Another data system against which SOII estimates have been compared is the National Health Interview Survey (NHIS), the principal source of information on the health of the civilian non-institutionalized population of the United States and one of the major data collection programs of the National Center for Health Statistics. The NHIS is an annual cross-sectional household interview survey of about 35,000 households and 87,500 people. Among many questions it asks are whether an injury occurred while the individual was working at a paid job, what type of medical care was sought, the external cause or nature of the injury, what the person was doing when the injury occurred, the date and place the injury occurred, and whether the person missed days of work.

There are a variety of advantages and disadvantages of the NHIS for estimating workplace injuries and illnesses. An advantage is that the scope of the survey is broader than that of SOII, encompassing all civilian workers, including Federal workers and the self-employed. Further, Leigh and colleagues (Leigh, Marcin and Miller 2004) argue that economic incentives for workers not to report injuries in the NHIS are weak or nonexistent. However, the sample of injury episodes collected by the NHIS is quite small (fewer than 2,000), so the NHIS cannot publish the amount of detail that the SOII can. Further, the NHIS relies on proxy respondents—that is, individuals who respond to questions on behalf of other household members and who may not be aware of some work injuries and illnesses. In addition, workers tend to forget less severe injury episodes, so —recall bias” is a problem for injury cases that occurred further away from the time of the interview. (Warner et al. 2005) Beginning with 2004, data were collected on injury episodes occurring within 3 months of the interview. However, the National Center for Health Statistics tabulates data only for injury cases that occurred within 5 weeks of the interview. Finally, any comparison of NHIS and SOII estimates is complicated by the fact that cases in NHIS are not necessarily OSHA recordable (as defined in the box above).

Other data sources used to track workplace injuries and illnesses and to compare against SOII include data from hospital discharges and emergency room visits. Three of the 19 occupational health indicators identified by the Council of State and Territorial

Epidemiologists (CSTE) are based on the National Hospital Discharge Survey. As noted by CSTE,

[S]tate hospital discharge data are useful for surveillance of serious health conditions. While these state data sets do not include explicit information about “work-relatedness” of the health conditions for which a patient is hospitalized, they do include information about the payer for the hospital stay. The designation of workers' compensation as primary payer is a good proxy for the work-relatedness of hospitalized injuries.<sup>3</sup>

Another source of hospital data that can track workplace injuries and illnesses is the National Electronic Injury Surveillance System (NEISS). Data for this source are collected for the National Institute for Occupational Safety and Health from a small sample of U.S. hospital emergency departments. In each hospital, a staff member reads the emergency room charts and identifies work-related cases.

An advantage of using data from hospitals is that all workers are potentially in scope, as opposed to the more limited scope of SOII. However, the cases that appear in hospitals are typically more severe than a typical OSHA-recordable case. Hospitalizations account for only a small percentage of all workplace injuries and illnesses—3 percent, according to CSTE. Identifying cases by means of the payer implies that the cases counted by hospital discharge data may or may not be OSHA recordable. In the case of the CSTE indicators, these cases would be workers' compensation claims. Similarly, NEISS data pertain only to cases treated in emergency departments, while the scope of the OSHA-recordable cases counted by SOII is both broader and potentially different. Finally, the relatively small sample size of the NEISS limits the availability of detailed estimates. All of the data sources just described should be viewed as providing estimates that are complementary to the SOII.

### **3. Dimensions of the SOII undercount**

Some have viewed SOII with misgivings over its failure to count all workplace injuries and illnesses. Their comments can be classified into four separate categories: underrecording of illnesses, incomplete scope in the coverage of workers, incomplete capture of injury and illness cases that are reported in other systems, and unreported cases.

#### **3.1 Underrecording of illnesses**

It is well known and acknowledged by BLS that SOII does not capture all occupational illnesses. In the BLS Handbook of Methods, BLS notes that

Each year, the SOII measures the number of new work-related illness cases which are recognized, diagnosed, and reported. But some conditions, such as long-term latent illnesses caused by exposure to carcinogens, often are difficult to relate to the workplace and are not adequately recognized and reported and are believed to be understated in the SOII. In contrast, the overwhelming majority of

---

<sup>3</sup> Quoted from [www.cste.org/dnn/ProgramsandActivities/OccupationalHealth/OccupationalHealthIndicators/Indicator2/tabid/97/Default.aspx](http://www.cste.org/dnn/ProgramsandActivities/OccupationalHealth/OccupationalHealthIndicators/Indicator2/tabid/97/Default.aspx) (visited July 16, 2008).

the reported new illnesses are those which are easier to directly relate to workplace activity. (US Bureau of Labor Statistics 2010)

A central problem is that many work-related illnesses take years to develop and may be difficult to attribute to the workplace. Thus, a recording mechanism based on employer records, as is SOII, will generally fail to capture these illnesses.

### **3.2 SOII scope restrictions**

Because of restrictions on the scope of the workers covered, SOII does not count all nonfatal workplace injuries and illnesses incurred by U.S. workers. Specifically, SOII currently excludes the self-employed; farms with fewer than 11 employees; private households; and, Federal government agencies.<sup>4</sup> To partially address this shortcoming, together with OSHA, BLS is exploring ways to collect data for Federal agencies. Currently, Federal agencies are required to record their workplace injuries and illnesses on OSHA logs, but they are not required to report these data to OSHA.

Collecting data on other workers who are outside the scope of SOII (the self-employed, private household workers, and workers on small farms) is problematic, because these workers are also outside of the scope of the Occupational Safety and Health Act of 1970 and therefore are not required to record injuries and illnesses on the OSHA logs that form the basis for the SOII. In addition, sample frames are not available to BLS to capture data on these workers. Some different approaches might be pursued to collect such data, including through a household survey such as the NHIS or from various other sources, such as insurance claims, emergency room visits, and hospital discharges.

Absent the collection of data through methods such as those just described, some researchers have generated estimates for out-of-scope workers. Estimates for some groups of workers are obtained from alternative data sources that are adjusted to conform to the OSHA-recordability concept underlying SOII. In other cases, estimates are produced by extrapolating from the known injury or fatality data on other groups of workers. Leigh, Marcin, and Miller (2004) estimated that in 1999 1.76 million injuries were incurred by out-of-scope workers, in addition to 5.335 million injuries reported in SOII. This out-of-scope estimate included injuries and illnesses to State and local government workers for whom SOII did not produce national estimates in 1999 (but does now).

### **3.3 Incomplete SOII capture of cases that are reported in other systems**

An important strand of the undercount literature argues that SOII fails to capture cases that are within the scope of the survey, but that are captured in other work-related injury and illness data systems. At least three approaches have been taken to establish the completeness of SOII and the OSHA logs underlying it: OSHA audits of employer records, aggregate comparisons of SOII estimates with estimates generated from other data systems, and microlevel matches of cases in SOII with cases in other data systems.

OSHA conducts onsite audits of employer injury and illness records to verify the overall accuracy of source records and to estimate the extent of employer compliance with

---

<sup>4</sup> SOII produced national estimates for State and local government worker injuries and illnesses starting with reference year 2008. This was done to partially address the criticism of restricted scope.

OSHA recordkeeping requirements. Annually, OSHA draws a small sample of establishments that have responded to its Data Initiative.<sup>5</sup> For calendar year 2004, OSHA found that 95.7 percent of establishments had “accurate” recordkeeping (at or above the 95-percent threshold) for total recordable injury and illness cases and that 95.3 percent of establishments had “accurate” recordkeeping for cases with days away from work, work restrictions, or transfers (DART). Among the recordable cases identified by auditors, 10.0 percent were not recorded, 6.4 percent were DART cases recorded as less severe non-DART cases, and 0.9 percent were non-DART cases recorded as more severe DART cases. In 2004, recordkeeping accuracy, according to the 95-percent criterion, was not statistically significantly different from previous years’ accuracy. (Eastern Research Group 2007)

One additional issue uncovered by the OSHA audits is a small amount of overrecording. The audits found instances where employers recorded non-OSHA recordable cases. These were almost exclusively non-DART cases and, as a result, were out of the scope of the microdata studies of underreporting to be discussed below. Overrecording of these minor cases increases the count of total workplace injuries and illnesses and partially compensates for the effects of any undercounting of more severe cases.

Aggregate studies of the undercount involve comparing estimates from SOII with estimates produced from other data systems. To the extent that these other data systems have different scopes from that of SOII, the estimates need to be adjusted to comparable scopes. As an example of an aggregate comparison, Leigh and colleagues (Leigh, Marcin and Miller 2004) compared SOII estimates with those from the National Health Interview Survey (NHIS). The SOII estimate of 6.3 million injuries and illnesses in 1994 was 28.2 percent below the NHIS estimate of 8.8 million injuries and illnesses. Leigh and colleagues note that economic incentives for workers not to report in SOII might be weak or nonexistent in the NHIS, explaining at least part of the estimated undercount.

Not all aggregate comparisons of estimates conclude that the SOII undercounts injuries and illnesses. Oleinick and Zaidman (2004) compare counts of workers’ compensation cases with counts of days-away-from-work cases in SOII for Minnesota over the period from 1992 to 2000. For cases lasting 4 or more days away from work—the cases for which data were available in the Minnesota workers’ compensation data set—Oleinick and Zaidman conclude that there is 92- to 97-percent concordance between the two estimates of injury and illness counts and that the BLS survey has “high sensitivity” for workplace injuries with 4 or more days away from work.

There are some limitations of aggregate comparisons. Most fundamentally, even if an estimate from another data system is close to the SOII estimate, it does not mean that underreporting is absent in SOII (or in the other system). It is possible that SOII captures some cases that are not in the other system, while the other system captures some cases that are not in SOII. In this instance, there is underreporting in both SOII and the other system. Indeed, that is what appears to occur in the microlevel studies described below.

---

<sup>5</sup> The OSHA Data Initiative (ODI) collects occupational injury and illness data and information on number of workers employed and number of hours worked from establishments in portions of the private sector and from some state and local government agencies. These data allow OSHA to calculate occupational injury and illness rates and to focus its efforts on individual workplaces with ongoing serious safety and health problems.

However, Oleinick and Zaidman dismissed the possibility that offsetting biases resulted in the close concordance that they found between SOII and workers' compensation. Another limitation of aggregate comparisons is that it may be difficult to ensure that the estimates from SOII and the other data system are for cases within the same scope. Researchers must make careful adjustments to ensure scope comparability. The Oleinick and Zaidman research is an example of a study in which adjustments draw the SOII and workers' compensation counts together.

To address these purported limitations of aggregate comparisons, recent studies have matched individual cases in SOII with cases in other systems. These studies attempt to restrict the data in SOII and other systems to the same scope and then to match cases on a variety of characteristics, including those of the worker, employer, and case. The studies are able to document the number of cases that are in another system but not in SOII, the number that are in SOII but not in the other system, and the number that are in both SOII and the other system.

Rosenman and colleagues (2006) match case-level SOII data to workers' compensation cases for the State of Michigan in 1999, 2000, and 2001. The scope of the data-set comparison was restricted to cases with more than 7 days away from work, because only cases surpassing a 7-day benefit waiting period are captured in the Michigan database. The researchers estimated that, on average each year from 1999 to 2001, 79,379 injury and illness cases were reported in only SOII, in only workers' compensation, or in both systems (that is, in the combination of the two data sets). Of these more than 79,000 cases, SOII captured 30,800, or 38.8 percent, whereas workers' compensation captured 62,264, or 78.4 percent. Focusing on specific types of injuries, the researchers found that SOII was more likely to capture certain injuries that are easier to observe and relate to the workplace, such as surface and open wounds, burns, and traumatic injuries to bones. In contrast, SOII was less likely to capture traumatic injuries to muscles, tendons, and the like, which include sprains and strains. These injuries are quite frequent both in SOII and in workers' compensation.

Subsequent analysis by Boden and Ozonoff (2008) provides undercount estimates that are considerably smaller than those of Rosenman and colleagues. Boden and Ozonoff matched SOII and workers' compensation data from 1998 to 2001 for six States: Minnesota, New Mexico, Oregon, Washington, West Virginia, and Wisconsin. The undercount estimates differ by State, but they indicate that on average SOII may be picking up about 69 percent of the injuries and illnesses appearing in the combination of the SOII and workers' compensation data sets. SOII did best at capturing cases in West Virginia (79.4 percent) and worst in the State of Washington (55.7 percent). Similarly, the researchers find widely varying estimates of the extent to which workers' compensation captures injuries that appear in the combination of both data sets—from 72.4 percent in Minnesota to 96.9 percent in Washington State.

It is difficult to gauge the reason for the difference in the findings of Rosenman and colleagues, on the one hand, and Boden and Ozonoff, on the other. The difference may be due to differences in the methodologies used, or it may be due to State-by-State variation. However, the Rosenman SOII-capture estimate of 38.8 percent is lower than the results found by Boden and Ozonoff for any State, suggesting that differences in methodology play a role.

### 3.4 Unreported cases

Cases that are unreported in multiple data systems constitute another group of undercounted cases. In the context of the BLS survey, this means that cases not reported in SOII may also not be reported elsewhere. Applying some assumptions, it is possible to estimate the number of such cases by means of a technique called “capture-recapture.” This technique was first applied to the estimation of animal populations in the wild, but it has been adapted to generate improved estimates in a wide variety of situations, such as drug use, homelessness, infectious diseases, and occupational injury and illness.

Capture-recapture uses probability theory and multiple overlapping, but incomplete, data sources to make inferences about the size of a partially unobserved population. Whereas the most straightforward application of the capture-recapture method uses basic probability theory, more sophisticated analyses rely on multivariate models. The latter analyses identify all unique cases recorded in at least one source and then use log-linear or logistic models to estimate the number of cases unrecorded by any source. Capture-recapture is a natural extension of the matching of data sources described in the previous section.

After matching individual cases in SOII and workers’ compensation data for Michigan from 1999 to 2001, Rosenman and colleagues (2006) used capture-recapture to estimate that although the data sources together included a total of 79,379 cases on average each year, an additional 15,654 were not captured in either data system. The latter cases bring the annual average total of cases to 95,033. Thus, 16.5 percent of cases went unreported. Further, when the SOII estimate of 30,800 was compared with the total, including unreported cases, Rosenman and colleagues estimated that SOII captured only 32.4 percent of all cases.

Boden and Ozonoff (2008) applied capture-recapture to the data for the six States in their study. They found that cases unrecorded in either SOII or workers’ compensation ranged from 13 percent of all cases in Minnesota and New Mexico to 3 percent in Washington State and West Virginia. The researchers’ estimate of the total SOII undercount after utilizing capture-recapture was smaller than that of Rosenman and colleagues, ranging from 46 percent in Washington to 22 percent in West Virginia. Boden and Ozonoff estimated that SOII missed about 40 percent of all cases across the six States; that is, SOII obtained about 60 percent of all cases. As previously mentioned, capture-recapture has been used for a variety of purposes. In an interesting non-U.S. example, Van Charante and Mulder (1998) found that employers reported only 35.6 percent of injuries to the government in the Netherlands.

Capture-recapture is a sophisticated technique for making inferences about unreported cases. However, the methodology does rely on some assumptions to generate results. One important assumption, termed “source independence,” is that the recording of cases in one system is independent of the recording of cases in another system. In fact, sources could be positively or negatively source dependent, meaning that a case recorded in one system is, respectively, more likely or less likely to be recorded in another system. For a variety of reasons, Boden and Ozonoff expect that SOII and workers’ compensation are positively source dependent. Some of these reasons are that the same person might record a case in both systems; if a worker does not report a case, it is not likely to be recorded in either system; and if an employer does not think a case is compensable, then he or she might erroneously believe that it also is not OSHA recordable.

If two sources are positively source dependent, then the estimate of the number of cases not captured in either system is biased downward; that is, underreporting is greater than when the number of cases is estimated assuming independence. Without data from a third source or without additional assumptions, it is not possible to estimate the extent of source dependence. Still, Boden and Ozonoff conduct a sensitivity analysis by estimating the undercount under a couple of positive dependence scenarios. Assuming different values for the odds ratio that a case is reported in SOII, given that it is reported in workers' compensation, they show that the estimated coverage of both SOII and workers' compensation drops with positive source dependence—sometimes substantially. However, they concede that they do not know what the correct odds ratio is (although they believe it is greater than one), leaving the source dependence issue unresolved.

#### **4. Current research**

In order to better understand the SOII undercount issue, BLS is conducting its own intramural research and supporting extramural research by State grantees and a contractor. Internal BLS research to date includes analysis of matched workers' compensation to SOII microdata and pilot interviews of employers regarding OSHA recordkeeping and workers' compensation claiming practices.

The aims of the analysis of the matched SOII-workers' compensation data are three-fold: to determine where SOII underreporting is greatest, according to industry, type of employer and type of case; to generate and test some hypotheses as to why SOII matches better to workers' compensation for some case types; and, to determine whether there are any methodological problems with the matching that might create an upward bias in the estimated undercount.

Details of some initial research appear in Nestoriak and Pierce (2009) and are discussed further in a paper at this conference titled —Using Capture-Recapture Analysis to Identify Factors Associated with Differential Reporting of Workplace Injuries and Illnesses.” (Boden, Nestoriak and Pierce 2010) In brief, the researchers analyzed Boden's matched microdata for Wisconsin. They calculated SOII capture rates, that is, the proportion of workers' compensation cases that appeared in the SOII microdata. They found that SOII capture rates are lower for: multi-establishment companies as compared to companies with only a single establishment in the State; carpal tunnel syndrome, inflammation, other cumulative, and hearing loss cases (cases that are more likely to be contested and that are harder to verify and/or relate to the workplace); workers' compensation cases filed after the year of onset or injury (cases that are more likely to be contested or to manifest themselves after the SOII survey is fielded); and, workers' compensation cases with zero temporary total disability days.

The issue of multi-establishment companies in a State is crucial. The SOII is a sample survey that obtains data only for some establishments in the State. In contrast, workers' compensation cases are reported for all establishments and tend to be reported on a company basis, with limited information available in the workers' compensation files about the particular establishment where a case occurs. The lack of establishment information in the workers' compensation data makes it difficult to determine whether a particular workers' compensation case should have a corresponding case in SOII. To make valid comparisons between the SOII and workers' compensation data, an adjustment must be made for the SOII sampling. Boden and Ozonoff's solution to this

problem is to use a universe file of establishments (the Quarterly Census of Employment and Wages) to determine the fraction of the total employment at the affiliated firm covered by the sampled BLS establishments. This fraction is then used to lower the weight applied to the workers' compensation unlinked cases (because these cases may be from an establishment not sampled for SOII). Although this solution makes good use of the available information, it does introduce additional non-sampling error.

Boden, Nestoriak and Pierce's study suggests that some of the undercount may stem from the fact that, in order to be timely, the SOII is fielded soon after the end of the reference year. The SOII will miss some cases that are not identified until after the SOII is conducted. This will explain why the SOII does not fully capture cases such as carpal tunnel syndrome and other cumulative cases that might be recognized later. It also would explain the lower capture rate of cases that tend to be contested (and therefore not initially recorded) by employers.

The other research that has been conducted by BLS involved interviewing employers about their OSHA recordkeeping and workers' compensation filing practices. BLS researcher Phipps conducted 26 in-person interviews of employers in the greater Washington, DC area in 2008. The interviews were based on a loosely structured questionnaire. The research was qualitative in nature and no generalizations about the population were possible. Details of this research are discussed in a paper by Phipps titled "Employer Interviews: Exploring Differences in Reporting Work Injuries and Illnesses in the Survey of Occupational Injuries and Illnesses and State Workers' Compensation Claims" presented at this conference.

Consistent with the results of Boden, Nestoriak and Pierce, Phipps found that half of the interviewed establishments said that some cases were not in SOII due to late reporting. Further, most establishments using temporary help agency workers did not include the injuries of those workers on their OSHA log, even when it was possible that they should according to OSHA regulations.<sup>6</sup> Phipps also found that contested and denied cases were handled differently on OSHA logs by the various interviewees (that is, interviewees differed as to whether the cases were kept or removed). Finally, Phipps found that different staff in a company might respond to SOII and handle workers' compensation claims, so that there might be some discrepancy in the recording and reporting of some cases.

The 2009 Omnibus budget bill passed by the US Congress included additional funding for the SOII to expand the amount and type of research that the BLS is undertaking regarding a potential undercount. The BLS is working internally and with State partners to strengthen its research efforts in three areas, with results expected by 2012. Additional details follow.

#### **4.1 Matching Research**

The funding has allowed the BLS to accelerate its current research with matched SOII-workers' compensation data to understand what types of cases are captured in each data system. The BLS has acquired and matched to SOII workers' compensation data for

---

<sup>6</sup> OSHA requires that the entity directly supervising the worker should record an injury, regardless of who pays the worker's wages and benefits and regardless of who pays for workers' compensation insurance.

Kentucky and hopes to conduct a similar analysis of Louisiana data. A contractor from Boston University (BU) is analyzing California data; while a State partner (Washington) is analyzing data for that State. The research will help identify potential methodological biases in previous research, better define the scope of the apparent underreporting, and might identify strategies to make the SOII estimates more complete.

#### **4.2 Employer Interviews**

The BLS and its BU and Washington State research partners will interview a sample of employers about factors that affect the recording of cases on OSHA logs and the filing of workers' compensation claims. The sampling will use the results from the matching activities described above to focus interviews on employers whose workers' compensation and SOII data match closely or differ greatly. The interviews will focus on certain areas where recordkeeping might be difficult or unclear.

Work to develop and test a standard interview protocol began in 2009. Establishments will be selected for interview in 2010 and 2011. Interviews will begin no later than 2011, once sufficient matching research is completed.

#### **4.3 Multiple Data-Source Pilot**

The work of Rosenman and colleagues, Boden and Ozonoff, and others suggests that no single data source can measure the total number of workplace injuries and illnesses. Using multiple data sources can improve completeness of coverage by including workers and cases that are outside the scope of any particular data source and by covering cases that, for a variety of possible reasons, do not appear in a particular data set. This, in fact, is the rationale for the multiple-data-source BLS Census of Fatal Occupational Injuries.

The BLS is working with three State partners (Washington, California, and Massachusetts) to pilot the use of multiple data sources to count two types of injury or illness: one acute—amputations—and one with a more gradual onset—carpal tunnel syndrome. The core data sources include the SOII, workers' compensation, emergency department visits, and hospitalizations. State partners are exploring the use of other data sets as well.

The pilot will provide information on the feasibility and cost of implementing a multiple-data source system for measuring the total number of workplace injuries and illnesses in the United States. It will also provide estimates of the contributions of various data sources to the total counts of amputations and carpal tunnel syndrome cases in each of the three study States.

The BLS and the States are finalizing common research protocols for the various studies in 2010. Evaluation of the data sources that have been collected thus far are being used to refine the protocols. Some of the issues that have been discussed include

- Case definitions: Inclusions (evidence of work-relatedness) and exclusions (Diabetic amputations, for example)
- Case selection criteria examples: ICD codes, OIICS codes, CPT, DRG, narrative text
- Scope criteria
- Study periods
- Complications of sequelae

Details of the status of this research are discussed in a paper by Wuellner et al. titled “Work-Related Injury and Illness Surveillance through Multiple Data Sources, a Pilot Project” presented at this conference.

## 5. Conclusions

SOII is designed to estimate the number and rate of OSHA-recordable cases of workplace injuries and illnesses. It covers most, but not all, sectors of the U.S. economy. This means that SOII does not capture some workplace injuries and illnesses that appear in other data systems, because of differences in the scope of cases captured and sectors covered. SOII (and other data systems) also may be limited in completeness by incentives that affect worker and employer reporting of workplace injuries and illnesses. Further, with an employer-based system for counting workplace injuries and illnesses such as SOII, it is difficult to measure long-latent occupational illnesses. For all of these reasons, SOII does not measure the total number of workplace injuries and illnesses.

However, SOII has advantages over other data systems. It quickly and efficiently produces detailed estimates that are consistent in definition across all States and industries. For cases with days away from work, it provides rich information about the occupation and demographics of injured or ill workers and about case characteristics such as the number of days away from work, the nature of the case (for example, a fracture), the body part affected, the event (a fall, for instance), the source (the floor, for example) and the timing of the incident. In comparison with SOII, many other data systems are not consistent across States (workers’ compensation is a prime example); cannot produce detailed estimates by State, industry, and case characteristics (NHIS is an example); do not exist for all States; or are very expensive to collect. These other systems also may have major scope limitations (for example, they measure only hospitalizations) or may be affected by various reporting incentives.

Some recent studies conclude that both SOII and other data systems undercount cases of workplace injuries and illnesses. Explanations other than willful underreporting—such as differences in the cases captured by various data systems and methodological aspects of the undercount research— may account for this finding. Some have argued that the gold standard for producing estimates of the total number of workplace injuries and illnesses is a multiple data source system. Indeed, BLS has implemented such an approach in collecting fatalities in the Census of Fatal Occupational Injuries (CFOI) program. However, in 2008, there were 3.7 million OSHA-recordable nonfatal workplace injuries and illnesses in private industry according to SOII, in comparison with 5,214 workplace injury fatalities counted by CFOI in all sectors of the U.S. economy. The vastly greater number of nonfatal injuries and illnesses suggests that it would be quite costly to implement a multiple data source system uniformly across all States for all nonfatal occupational injuries and illnesses.

Within the constraints of its mission as a statistical agency, BLS will continue to work to ensure that SOII accurately measures within-scope workplace injuries and illnesses. As described in this article, BLS is undertaking and supporting research designed to understand and explain differences between its estimates and those of other systems. This research may suggest some ways that the SOII might be modified to provide a more complete workplace injury and illness count.

## References

- Boden, Leslie I., Nicole Nestoriak and Brooks Pierce, —Using Capture-Recapture Analysis to Identify Factors Associated with Differential Reporting of Workplace Injuries and Illnesses.” In 2010 JSM Proceedings, Statistical Computing Section. Alexandria, VA: American Statistical Association.
- Boden, Leslie I. and Al Ozonoff, —Capture-Recapture Estimates of Nonfatal Workplace Injuries and Illnesses,” *Annals of Epidemiology*, June, 2008, pp. 500-6.
- Eastern Research Group and National Opinion Research Center, —OSHA Data Initiative Collection Quality Control: Analysis of Audits on CY 2004 Employer Injury and Illness Recordkeeping,” Final Report, June 2007.
- Leigh, J. Paul, James P. Marcin, and Ted R. Miller, —An estimate of the U.S. Government's undercount of nonfatal occupational injuries,” *Journal of Occupational and Environmental Medicine*, 46, January, 2004, pp. 10–18.
- Nestoriak, Nicole and Brooks Pierce, —Comparing Workers' Compensation claims with establishments' responses to the SOII,” *Monthly Labor Review*, May 2009, pp. 57–64.
- Oleinick, Arthur and Brian Zaidman, —Methodological Issues in the Use of Workers' Compensation Databases for the Study of Work Injuries with Days Away from Work. I. Sensitivity of Case Ascertainment,” *American Journal of Industrial Medicine*, 45, March, 2004, pp. 260–74.
- Phipps, Polly, —Employer Interviews: Exploring Differences in Reporting Work Injuries and Illnesses in the Survey of Occupational Injuries and Illnesses and State Workers' Compensation Claims.” In 2010 JSM Proceedings, Statistical Computing Section. Alexandria, VA: American Statistical Association.
- Rosenman, Kenneth D., Alice Kalush, Mary Jo Reilly, Joseph C. Gardiner, Matthew Reeves and Zhewui Luo, —How Much Work-Related Injury and Illness is Missed by the Current National Surveillance System,” *Journal of Occupational and Environmental Medicine*, 42, April, 2006, pp. 357-65.
- Van Charante, Anton W. Moll and Paul G. Mulder, —Reporting of Industrial Accidents in the Netherlands,” *American Journal of Epidemiology*, 148, July 15, 1998, pp. 182-190.
- US Bureau of Labor Statistics, Handbook of Methods 2010, available at [http://www.bls.gov/opub/hom/homch9\\_a.htm](http://www.bls.gov/opub/hom/homch9_a.htm).
- US House of Representatives, —Hidden Tragedy: Underreporting of Workplace Injuries and Illnesses,” A Majority Staff Report by the Committee of Education and Labor, June 2008.
- Warner, M., N. Schenker, M. A. Heinen, and L.A. Fingerhut, —The effects of recall on reporting injury and poisoning episodes in the National Health Interview Survey,” *Injury Prevention*, 11, October, 2005, pp. 282–7.
- Wuellner, Sara, Dave Bonauto, Darrin Adams, Cody Spann and Nicole Reister, —Work-Related Injury and Illness Surveillance through Multiple Data Sources, a Pilot Project.” In 2010 JSM Proceedings, Statistical Computing Section. Alexandria, VA: American Statistical Association.