B. Collection of information employing statistical methods

The statistical methods used in the sample design of the survey are described in this section. The documents listed below are attached or available at the hyperlink provided. These documents are either referenced in this section or provide additional information.

Overview of the Survey of Occupational Injuries and Illnesses Sample Design and Estimation Methodology – Presented at the 2008 Joint Statistical Meetings (10/27/08) – https://www.bls.gov/osmr/research-papers/2008/pdf/st080120.pdf

Deriving Inputs for the Allocation of State Samples (04/01/10)

The growth in cases with Restricted Activity or Job Transfer (08/2011)

Methods Used To Calculate the Variances of the OSHS Case and Demographic Estimates (2/22/02)

Variance Estimation Requirements for Summary Totals and Rates for the Annual Survey of Occupational Injuries and Illnesses (6/23/05)

BLS Handbook of Methods – Survey of Occupational Injuries and Illnesses (11/03/2017) – https://www.bls.gov/opub/hom/soii/home.htm

Nonresponse Bias in the Survey of Occupational Injuries and Illnesses (10/2013) – https://www.bls.gov/osmr/research-papers/2013/pdf/st130170.pdf

Sample Allocation to Increase the Expected Number of Publishable Cells in the Survey of Occupational Injuries and Illnesses (10/2015) – https://www.bls.gov/osmr/research-papers/2015/pdf/st150070.pdf

Deep neural networks for worker injury autocoding (09/2017) – https://www.bls.gov/iif/deep-neural-networks.pdf

1. Description of universe and sample

Universe. The main source for the SOII sampling frame is the BLS Quarterly Census of Employment and Wages (QCEW) (*BLS Handbook of Methods*, *Quarterly Census of Employment and Wages* from https://www.bls.gov/cew/). The QCEW is a near quarterly census of employers collecting employment and wages by ownership, county, and six-digit North American Industry Classification System (NAICS) code. States have an option to either use the QCEW or supply public sector sampling frames for state and local government units. Business census files are utilized to create a sampling frame for Guam, whose establishment data are not available in QCEW. The number of states providing their own public sector frame are provided in Table 1:

Table 1: Number of states providing frames by ownership type

Year	State Frame	Local Frame	Private Frame
2017	6	3	1
2018	6	3	1
2019	6	3	1

The potential number of respondents (establishments) covered by the scope of the survey is approximately 8.1 million, although only about 800,000 employers keep records on a routine basis due to recordkeeping exemptions defined by OSHA for employers in low hazard industries and employers with fewer than 11 employees, or having no recordable cases. The occupational injury and illness data reported through the annual survey are based on records that employers in the following North American Industry Classification System (NAICS) industries maintain under the Occupational Safety and Health Act:

Table 2: NAICS Industry Sectors Covered by SOII

Sector	Description
11	Agriculture, forestry, fishing and hunting
21	Mining, quarrying, and oil and gas extraction
22	Utilities
23	Construction
31, 32, 33	Manufacturing
42	Wholesale trade
44,45	Retail trade
48,49	Transportation and warehousing
51	Information
52	Finance and insurance
53	Real estate and rental and leasing
54	Professional, scientific, and technical services
55	Management of companies and enterprises
56	Administrative and support and waste management and remediation services
61	Educational services
62	Health care and social assistance
71	Arts, entertainment, and recreation
72	Accommodation and food services
81	Other services (except public administration)

Excluded from the national survey collection are:

- Self-employed individuals;
- Farms with fewer than 11 employees (Sector 11);
- Employers regulated by other Federal safety and health laws;
- United States Postal Service and;
- Federal government agencies.

Mining and railroad industries are not covered as part of the sampling process. Injury and illness data from these industries are obtained directly from the Mine Safety and Health Administration and the Federal Railroad Administration, respectively, and used to produce state and national estimates.

Data collected for reference year 2008 and published in calendar year 2009 marked the first time state and local government agency data were collected and published for all participating states and for the nation as a whole. The SOII is a federal/state cooperative program, in which the federal government and participating states share the costs of participating state data collection activities. State participation in the survey may vary by year. Sample sizes are determined by the participating states based on budget constraints and independent samples are selected for each state annually. Data are collected by BLS regional offices for non-participating states.

For the 2019 survey, 41 states plus the District of Columbia plan to participate in the survey. For the remaining nine states which are referred to as Non-State Grantees (NSG), a smaller sample is selected to provide data which contribute to national estimates only. The nine NSG states for 2019 are:

Table 3: Non-State Grantees for the SOII, 2019

Colorado	Florida	Idaho
Mississippi	New Hampshire	North Dakota
Rhode Island	South Dakota	Oklahoma

Additionally, estimates are tabulated for three U.S. territories – Guam, Puerto Rico, and the Virgin Islands – but data from these territories are not included in the tabulation of national estimates.

Sample. The SOII utilizes a stratified probability sample design with strata defined by state, ownership, industry, and size class. The first characteristic enables all the state grantees participating in the survey to produce estimates at the state level. Ownership is defined into three categories: State government, local government, and private industry. There are varying degrees of industry stratification levels within each state. This is desirable because some industries are more prevalent in some states compared to others. Also, some industries can be relatively small in employment but have high injury and illness rates which make them likely to be designated for estimation. Thus, states determine which industries are most important in terms of publication and the extent of industry stratification is set independently within each state. BLS sets some minimal levels of desired industry publication to ensure sufficient coverage for national estimates. These industry classifications are defined using the North American Industry Classification System (NAICS, http://www.census.gov/eos/www/naics/) and are referred to as Target Estimation Industries (TEI).

Finally, establishments are classified into five size classes based on average annual employment and defined as follows:

Table 4: Establishment Size Classes for SOII

Size Class	Average Annual Employment
1	10 or fewer
2	11-49
3	50-249
4	250-999
5	1,000 or greater

After each establishment is assigned to its respective stratum, a systematic selection with equal probability is used to select a sample from each sampling cell (stratum). As mentioned earlier, a sampling cell is defined as state/ownership/TEI/size class. Prior to sample selection, units are sorted by employment within each stratum. This sorting ensures that the sampled establishments have varying numbers of employees. Full details of the survey design are provided in Section 2.

For survey year 2019, the sample size will be approximately 232,400 or 2.85 percent of the total 8.1 million establishments in state, local, and private ownerships.

Response rate. The survey is mandatory, with the exception of state and local government units in the following states:

Table 5: Mandatory Exceptions

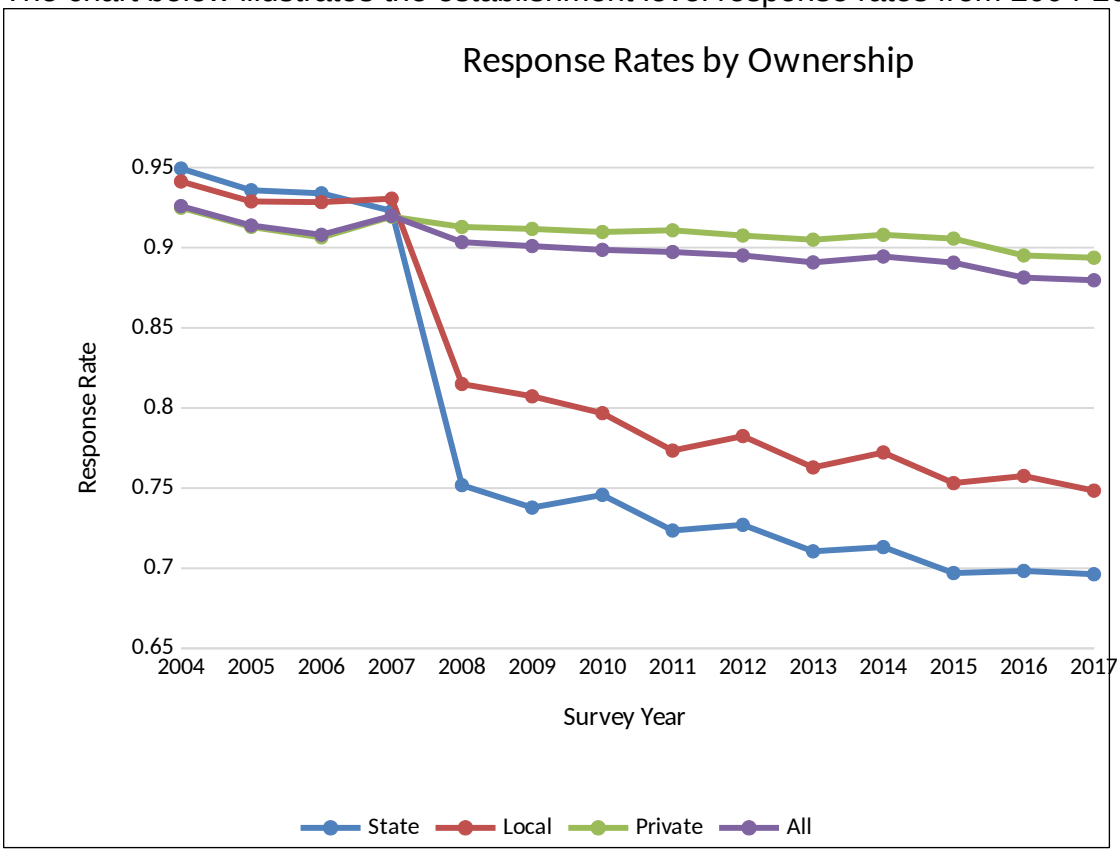
Alabama	Arkansas	Colorado
Delaware	District of Columbia	Florida
Georgia	Idaho	Illinois
Kansas	Louisiana	Mississippi
Missouri	Montana	Nebraska
New Hampshire	North Dakota	Pennsylvania
Rhode Island	South Dakota	Texas

Each year, establishments in the SOII are notified via mail or email of their requirement to participate. All non-respondents are sent up to two non-response mailings as a follow-up to the initial mailing. Some states choose to send a third or fourth non-response mailing to non-respondents late in the collection period. For survey year 2018, approximately two-thirds of the states will send an optional third non-response mailing to a majority of the non-respondents at that point in time, and less than five percent of the states will send a fourth non-response mailing. In addition, states may contact respondents via telephone for additional non-response follow-up. States may also use email follow-up for some respondents who have expressed a preference for communication via email. No systematic establishment level data on the number of telephone or email non-response follow-up contacts is captured.

As mentioned earlier, public sector establishments were included in the 2008 survey for **all** states, including those from which no public sector data had been collected in the past. In these states, public sector establishments have no mandate to provide data to the SOII; their participation is voluntary. For survey year 2008, the rates for both state and local government decreased, primarily due to the addition of the voluntary state and local government establishments.

In 2010, an in-depth response rate analysis was undertaken. Aggregate response rates in the SOII were shown to be above 90 percent due to the mandatory nature of the survey and the excellent efforts to obtain survey data by our state and regional partners. However, it was also shown that states with voluntary reporting status for the state and local governments had low response rates for the government units. In subsequent years, this study was updated to continually monitor the item and establishment non-response. As of the most recent update, there have been no significant changes.

The chart below illustrates the establishment level response rates from 2004-2017:



Although response rates for the SOII program have historically been high, the expansion of public sector collection in voluntary states resulted in a response rate of 75 percent in state government in 2008. Per OMB statistical guidelines, a nonresponse bias study was initiated and completed in 2013 (See “Nonresponse Bias in the Survey of Occupational Injuries and Illnesses” in the supporting documents). This work concluded that, in states where participation is voluntary, there is statistically significant

evidence to suggest that counts for establishments identified by a model as being 'likely' to respond are lower than establishments that were identified as 'unlikely' to respond. Similarly, the mean case rates for establishments identified by a model as being 'likely' to respond were higher than those identified as being 'unlikely' to respond. This apparent contradiction between the biases in the measures was explained by the changes in the estimates of the hours worked that are included in the rate estimate. Given these voluntary state/local units comprised 1.3 percent of the total survey, efforts to address these observed biases were deferred due to resource constraints.

Additional response efforts are being conducted to analyze response rates for several key data elements collected for each establishment in the survey. Data elements for NAICS industry, SOC occupation, source, nature, part, and event for each case with days away from work are coded by BLS regional staff and/or state partners. As such, these fields are always available for collected data. Other data elements such as ethnicity, whether the event occurred before/during/after the work shift, the time of the event, and the time the employee began work may be missing from collected data. BLS has initiated a response analysis effort for these other data elements to identify specific response rates and the characteristics of respondents versus non-respondents for these variables.

Regional offices are also working with states on collection practices to improve response for voluntary units.

BLS will continue to monitor the response rates in the next 3 years for all segments of the survey scope. BLS will update the analysis each year and make recommendations for improvements in the data collection process based on the results of our analysis. BLS will conduct a non-response bias analysis for groups of establishments with response rates below 80 percent and investigate potential non-response bias issues for any specific data elements with item response rates below 70 percent; BLS will also implement additional non-response bias studies. Details for these studies will be documented in subsequent clearance packages.

2. Statistical methodology

Survey design. The survey is based on probability survey design theory and methodology at both the national and state levels. This methodology provides a statistical foundation for drawing inference to the full universe being studied.

Research was done to determine what measure of size was most appropriate for the allocation module. Discussion with the Occupational Safety and Health Statistics (OSHS) Management Team narrowed the choices to the rates for Total Recordable Cases (TRC); Cases with Days Away from Work (DAFW); and Cases with Days Away from Work, Job Transfer, or Restriction (DART).

Rates from the 2003 SOII were studied for all 1251 TEIs for each of the above case categories. The average case rate, standard deviation (SD), and coefficient of variation

(CV) for each set of rates were calculated. The CV is the standard deviation divided by the estimate, which is commonly used to compare estimates in relative terms. The results are shown below:

Description	Ave. Rate	SD	CV
DAFW	1.5540	1.078	0.69
DART	3.0479	2.000	0.66
TRC	5.5300	3.229	0.58

Based on this information it was recommended that the TRC rate be used as the measure of size for the sample allocation process for the survey. The lower CV indicates that it is the most stable indicator.

Additionally, to fulfill the needs of users of the survey statistics, the sample provides industry estimates. A list of the industries for which estimates are required is compiled by the BLS after consultation with the principal Federal users. The sample is currently designed to generate national data for all targeted NAICS levels that meet publication standards.

Allocation procedure. The principal feature of the survey's probability sample design is its use of stratified random sampling with Neyman allocation. The characteristics used to stratify the units are state, ownership (whether private, state, or local government), industry code, and employment size class. Since these characteristics are highly correlated with the characteristics that the survey measures, stratified sampling provides a gain in precision and thus results in a smaller sample size.

Using Neyman allocation, optimal sample sizes are determined for each stratum within each State. Historical case data are applied to compute sampling errors used in the allocation process. Details about this process can be found in ***Deriving Inputs for the Allocation of State Samples*** (04/18/05).

The first simplifying assumption for allocation is that for each TEI $\times$ size class stratum *h*, the employment in each establishment is the same, which is denoted by E_h. BLS also ignores weighting adjustments. In addition, BLS assumes that the sampling of establishments in each stratum is simple random sample with replacement. (It is actually without replacement of course, but this is a common assumption to simplify the formulas.)

One consequence of these assumptions is that the estimate of the overall employment is constant and as a result the estimated incidence rate of recordable cases in the universe is the estimated number of recordable cases divided by this constant. Therefore, the optimal allocation for the total number of recordable cases and the incidence rate of recordable cases are the same. BLS will only consider the optimal allocation for the total number of recordable cases.

BLSintroduces the following notation. For sampling stratum *h* let:

N_h denote the number of frame units

n_h denote the number of sample units

$W_h = N_h / n_h$ denote the sample weight

$T_h = N_h E_h$ denote the total employment in stratum *h*

p_h denote the incident rate for total recordable cases

$\hat{Y}_h$ denote the unweighted sample number of recordable cases

Also let:

$\hat{Y}$ denote the estimated number of recordable cases in the entire universe.

Then

$$\hat{Y} = \sum_h W_h \hat{Y}_h = \sum_h \frac{N_h \hat{Y}_h}{n_h} \tag{1}$$

$$V(\hat{Y}) = \sum_h \frac{N_h^2 V(\hat{Y}_h)}{n_h^2} \tag{2}$$

where *V* denotes variance.

Now BLS will obtain $V(\hat{Y}_h)$ under two different assumptions. Assumption (a) is:

(a) All employees in stratum *h* have either 0 or 1 recordable cases and the probability that an employee has a recordable case is p_h .

In this case $\hat{Y}_h$ can be considered to have a binomial distribution with $n_h E_h$ trials and p_h the probability of success in each trial and consequently

$$V(\hat{Y}_h) = n_h E_h p_h (1 - p_h) \tag{3}$$

Assumption (b) is:

(b) The total recordable case rate for the n_h sample establishments in stratum *h* has a binomial distribution with n_h trials and p_h the probability of success in each trial. In that case

$$V(\hat{Y}_h)=n_h E_h^2 p_h(1-p_h) \tag{4}$$

Although BLS will derive the optimal allocations under both assumptions, BLS prefers assumption (b) since under assumption (a) the variance of the recordable case rate among establishments in stratum *h* BLS believes will be unrealistically small, particularly for strata with large E_h .

To derive the optimal allocation under assumption (a) we substitute (3) into (2) obtaining

$$V(\hat{Y})=\sum_h \frac{N_h^2 E_h p_h(1-p_h)}{n_h} \tag{5}$$

Viewing (5) as a function of the variables n_h and minimizing (5) with respect to these variables by means of the method of Lagrange multipliers from advanced calculus, BLS obtains that (5) is minimized when the n_h are proportional to

$$N_h(E_h * p_h * (1-p_h))^{.5} \tag{6}$$

As for the preferred assumption (b), to derive the optimal allocation, BLS similarly substitutes (4) into (2) obtaining

$$V(\hat{Y})=\sum_h \frac{N_h^2 E_h^2 p_h(1-p_h)}{n_h} \tag{7}$$

Minimizing (7) as BLS minimized (5), BLS obtains that (7) is minimized when the n_h are proportional to

$$N_h * E_h * (p_h(1-p_h))^{.5} = T_h(p_h(1-p_h))^{.5} \tag{8}$$

which is the preferred allocation.

Sample procedure. Once the sample is allocated, the process of selecting the specific units is done by applying a systematic selection with equal probability independently within each sampling cell. Because the frame is stratified by employment size within each TEI before sample selection, it was felt equal probability sampling was appropriate rather than a PPS selection. PPS selection is often applied to frames that aren't stratified by size so in this case, it was felt that no additional value would be gained by selecting the sample by PPS.

The survey is conducted by mail questionnaire through the BLS-Washington and regional offices and participating state statistical grant agencies. Survey participants are able to respond to the survey via the Internet, by phone, by fax, or by mail via a paper questionnaire. SOII respondents have been asked to provide a preferred method for notification of participation in future surveys. Starting with survey year 2016, employers who have expressed a preference of email notifications have been notified both of their requirement to maintain injury and illness records for the upcoming reference year and for notifying participants of their responsibility to report their data during collection. These email notifications are done in accordance with BLS policy on the use of email for data collection.

Estimation procedure. The survey's estimates of the number of injuries and illnesses for the population are based on the Horvitz-Thompson estimator, which is an unbiased estimator. The estimates of the incidence of injuries or illnesses per 100 full-time

workers are computed using a ratio estimator. The estimates of the incidence rates are calculated as

$$R = C\left(\frac{200{,}000}{\sum H}\right)$$

where:

C = number of injuries and illnesses
$\sum H$ = total hours worked by all employees during a calendar year
200,000 = base for 100 full-time equivalent workers (working 40 hours per week, 50 weeks per year).

The estimation system has several major components that are used to generate summary estimates. The first four components generate factors that are applied to each unit's original weight in order to determine a final weight for the unit. These factors were developed to handle various data collection issues. The original weight that each unit is assigned at the time the sample is drawn is multiplied by each of the factors calculated by the estimation system to obtain the final weight for each establishment. The following is a synopsis of these four components.

When a unit cannot be collected as assigned, it is assigned a **Reaggregation** factor. For example, if XYZ Company exists on the sample with 1,000 employees but the respondent reports for only one of two locations with 500 employees each, it is treated as a reaggregation situation. The Reaggregation factor is equal to the target (or sampled) employment for the establishment divided by the reported employment for collected establishments. It is calculated for each individual establishment.

In cases where a sampled unit is within scope of the survey but does not provide data, it is treated as a nonrespondent. Units within scope are considered viable units. This would include collected units as well as nonrespondents. The **Nonresponse** adjustment factor is the sum of the weighted viable employment within the sampling stratum divided by the sum of the weighted usable employment for an entire sampling stratum. The nonresponse adjustment factor is applied to each unit in a stratum.

In some cases, collected data is so extreme that it stands apart from the rest of the observations. For example, suppose in a dental office (which is historically a low incidence industry for injuries and illnesses), poisonous gas gets in the ventilation system which causes several employees to miss work for several days. This is a highly unusual circumstance for that industry. This situation would be deemed an outlier for estimation purposes and handled with the outlier adjustment. If any outliers are identified and approved by the national office, the system calculates an **Outlier** adjustment factor so that the outlier represents only itself. In addition, the system calculates outlier adjustment factors for all other non-outlier units in the sampling stratum. This ensures that the re-assigned weight is distributed equally amongst all units in the strata.

Benchmarking is done in an effort to account for the time lapse between the sampling frame used for selecting the sample and the latest available frame information. Thus, a factor is computed by dividing the target employment (latest available employment) for the sampling frame by the weighted reported employment for collected units.

The system calculates a final weight for each unit. The final weight is a product of the original weight and all four of the factors. All estimates are the sum of the weighted (final weight) characteristic of all the units in a stratum.

In 2010 a pilot study to measure rates of Days of Job Transfer or Restriction (DJTR) for selected industries was begun using data from the 2011 survey reference year. The first public release of the case and circumstances data for DJTR cases from this pilot occurred on April 25, 2013. BLS continues to analyze the results of this on-going pilot test to determine the value of the information and to assess how best to implement the collection of these data as well as days away from work cases in future survey years. Updates to this DJTR pilot study are continuing by changing the industries of interest. See the testing section below for details.

3. Statistical reliability

Survey sampling errors. The survey utilizes a full probability survey design that makes it possible to determine the reliability of the survey estimates. Standard errors are produced for all injury and illness counts and case and demographic data as well for all data directly collected by the survey.

The variance estimation procedures are described in detail in the attached documents mentioned earlier:

Methods Used To Calculate the Variances of the OSHS Case and Demographic Estimates (2/22/02)
Variance Estimation Requirements for Summary Totals and Rates for the Annual Survey of Occupational Injuries and Illnesses (6/23/05)

4. Testing procedures

The survey was first undertaken in 1972 with a sample size of approximately 650,000. Since then the BLS has made significant progress toward reducing respondent burden by employing various statistical survey design techniques; the present sample size is approximately 232,400. The BLS is continually researching methods that will reduce the respondent burden without jeopardizing the reliability of the estimates.

Responding to concerns of data users and recommendations of the National Academy of Sciences, in 1989, the BLS initiated its efforts to redesign the survey by conducting a series of pilot surveys to test alternative data collection forms and procedures. Successive phases of pilot testing continued through 1990 and 1991. Cognitive testing of that survey questionnaire with sample respondents was conducted at that time. The objective of these tests was to help develop forms and questions that respondents easily understand and can readily answer.

In survey year 2006, the SOII program conducted a one-year quality assurance (QA) study that had primarily a focus on addressing the magnitude of employer error in recording data from their OSHA forms to the different types of BLS collection forms and methods. The results showed no systematic under-reporting or over-reporting by employers. There was no strong dependence between error rates and collection methods.

Beginning in survey year 2007, the QA program introduced in 2006 was extended and modified to evaluate the quality of the data collected in terms of proper collection methods with the goal of minimizing curbstoning and collector adjustments without respondent contact. If improper collection methods or procedures were uncovered, they were corrected. A byproduct of this program was that each data collector would know that any form they have processed could be selected for the program.

In 2003, the BLS introduced the Internet Data Collection Facility (IDCF) as an alternative to paper collection of data. This system has edits built in which help minimize coding errors. The system is updated annually to incorporate improvements as a result of experience from previous years.

In 2008, extensive cognitive testing was completed on the IDCF collection system. In addition to being an overall review, this testing also provided detailed analysis of the site's usability and eye-tracking. The summary (Summary of Expert Review of SOII IDCF Web Pages) provided extensive feedback, as well as a rating system that addressed "short-term" (wording changes), "Mid-term" (changes that affect the order of pages (flow), but seemed simple to execute), and "long-term" (changes with skip patterns, or associated buttons that appear to be more complex and would require more testing). The implementation of these changes went through a prioritization processes that took into account BLS staff resources to implement.

In 2009, extensive cognitive testing was completed on the IDCF Adobe Fillable Form. Recommendations were provided (OSMR Review of the Revised SOII Adobe Form), and efforts were made to incorporate them in a timely manner.

In 2012, extensive follow-up cognitive testing was completed on the IDCF collection system. This testing showed (Results of the SOII Edits Usability Test) a vast improvement over previous studies, and noted limited issues in three main areas:

1) Respondents showed difficulty in understanding what they are supposed to enter in the 'total hours worked by all employees' field, and in using the optional worksheet that accompanies this field.
2) Respondents can be confused and/or frustrated by the way the information about the average hours worked per employee is derived and presented on the screen.
3) Respondents missed or had negative reactions to the error message that appeared on the detailed "cases with days away from work" reporting page.

Improvements have been made annually to IDCF to address issues based on the level of perceived need and available resource constraints.

In 2015, an option was added to the IDCF collection system allowing users to 'opt-in' to receive future communications from BLS via email. Extensive cognitive testing was performed on this additional module to ensure understanding and ease of use.

Since 2008, BLS has been conducting research concerning the completeness of estimates from the SOII. This multiyear research effort provided results in 2012 which were used to guide the selection of further research. (See section on Data Quality Research below.)

BLS is investigating methods for decreasing the number 'unpublishable' estimates. These methods aggregate multiple years of directly-collected annual survey data into 'multi-year' estimates. Aggregating annually collected data in this way will increase the

number of usable units contributing to an estimate, decreasing the likelihood the cells will be deemed ‘unpublishable’ due to too few usable units supporting estimates within the cell, or reliability (as measured by sampling variation) is too high.

BLS also utilizes statistical quality control techniques to maintain the system's high level of reliability.

Data Quality Research. BLS is conducting ongoing research into the completeness of the injury and illness counts from the Survey of Occupational Injuries and Illnesses (SOII). The purpose of this research is to better understand a potential undercount of occupational injuries and illnesses reported by the SOII and to investigate possible reasons behind it. BLS uses results from this research to actively address SOII data quality issues that are identified. Several articles and papers describing this research and actions taken are available at https://www.bls.gov/iif/data-quality.htm.

BLS continues to evaluate the results of the data quality research that has been completed, as well as to pursue further research. These efforts include evaluating reporting practices employed by establishments and testing the feasibility of collection of injury and illness data directly from workers.

Employer reporting practices were investigated by conducting a follow-back study of a subsample of respondents to the 2013 SOII. The results of this study are available at https://www.bls.gov/iif/national-respondent-recontact-survey-report.pdf. Additionally, a separate employer study was conducted to estimate the prevalence of compliant and noncompliant recordkeeping practices in four states. The results of this study are available at https://www.bls.gov/iif/four-state-data-report.pdf.

The feasibility of collecting injury and illness data directly from workers is currently being evaluated. BLS conducted a pilot test of a worker survey in 2017-2018 and is analyzing the results.

Computer-Assisted Coding. BLS is constantly looking for ways to minimize the impact of human error in data collection. Because much of the occupational data are provided in narrative form, BLS and its state partners traditionally had to manually translate these narratives into codes. While BLS has incrementally developed rules for identifying human coding errors, consistency remained a concern. In 2012, BLS began researching the concept of using computer learning algorithms to “autocode” free-form written case narratives from survey respondents. Initial results proved promising and indicated that computer-assisted coding was feasible.

BLS uses research output as part of the annual review of the codes state coders have assigned to occupation and case circumstances for more than a quarter million nonfatal injuries and illnesses. BLS will continue to develop and evaluate computer-assisted coding with the twin goals of improving consistency and freeing personnel for more complex assignments where staff expertise is critically needed.

For the 2014 SOII, BLS began automatically assigning occupation codes and found that it could successfully automatically assign occupation codes to about one-quarter of 2014 SOII cases. Autocoding was expanded for the 2015 SOII to include nature of injury or illness and part of body affected. With this expansion, SOII autocoded about 430,000 codes. A small portion of the autocoded values were withheld from the coders and were manually coded. The manually assigned codes were compared to the autocoder assigned values for quality assurance measurement purposes. The results of this experiment demonstrated that on average the autocoder assigned values had higher accuracy measures than the human coders, in line with other studies on coder accuracy.

For the 2016 SOII, autocoding was expanded to include event or exposure and source of the injury or illnesses. With this expansion, the autocoders assigned nearly 666,000 codes. And as a result of further improvements to the autocoder, nearly 750,000 codes were assigned automatically for the 2017 SOII.

BLS has continued to assess the quality and accuracy of the autocoders. For the 2018 SOII, the existing logistic regression autocoder was replaced with neural network autocoders which are expected to result in 25 percent fewer coding errors compared to the logistic regression autocoder and 40 percent fewer errors than human coders (https://www.bls.gov/iif/deep-neural-networks.pdf).

Days of Job Transfer or Restriction Testing. Beginning with the 2011 survey year, BLS began testing the collection of case and demographic data for injury and illness cases that required only days of job transfer or restriction (DJTR). The purpose of this on-going pilot study is to evaluate collection of these cases and to learn more about occupational injuries and illnesses that resulted in days of job transfer or work restriction. The results of the first three years of collection were successful and demonstrated that these data could be collected and processed accurately for a limited set of industries. Results from the DJTR study are available at http://www.bls.gov/iif/days-of-job-transfer-or-restriction.htm.

DJTR details were collected from a second set of industries covering the 2014-2016 SOII, results from which are available at https://www.bls.gov/opub/reports/job-transfer-or-work-restriction/2016/home.htm. A third set of industries were selected for collection of DJTR case details from the 2017-2019 SOII, which will be analyzed and published following the 2019 SOII. BLS anticipates the 2017-2019 DJTR study will be the last, after which results will be assessed to determine how best to implement the collection of these data along with days away from work cases in future survey years. BLS regards the collection of these cases with only job transfer or restriction as significant in its coverage of the American workforce.

To retain the level of case and demographic characteristics estimates published currently for cases with days away from work and publish similar estimates for cases with job transfer or restriction, a greater number of cases will need to be collected from

employers. BLS has maintained the subsampling process for employers to limit to 15 the number of cases each employer needs to submit. BLS is continuing to examine this issue to determine an optimal number of cases to collect for each type of case while limiting the burden on both the employer and on the participating State agencies.

OSHA Electronic Recordkeeping. The Occupational Safety and Health Administration (OSHA) amended its recordkeeping rules to add requirements for the electronic submission of injury and illness information from specified establishments. Beginning with calendar year 2016 data, establishments with 250 or more employees in all industries and establishments with 20-249 employees in certain industries with historically high rates of occupational injuries and illnesses were required to electronically submit to OSHA information from their OSHA 300A (Summary of Work-Related Injuries and Illnesses) by the end of 2017. The deadline for OSHA reporting of calendar year 2017 data was advanced to July 1, 2018 and for calendar years 2018 onward employers must submit data to OSHA by March 2 following the reference year.

The OSHA reporting rule did not add to or change any employer's obligation to complete and retain injury and illness records under OSHA's regulations for recording and reporting occupational injuries and illnesses. The rule also did not change any employer's obligation to complete the SOII. These data that employers are required to submit to OSHA are similar to those collected by the BLS injury and illness survey. OSHA requires establishment-specific data to target interventions such as inspections, consultations, and technical assistance.

OSHA further amended its recordkeeping rule in 2019 to remove the planned requirement for covered establishments to electronically submit to OSHA information from the OSHA Form 300 (Log of Work-Related Injuries and Illnesses) and OSHA Form 301 (Injury and Illness Incident Report). Covered establishments are required to electronically submit information only from the OSHA Form 300A.

BLS formed a working group with OSHA to assess data quality, including timeliness, accuracy, and public use of the collected data, as well as to align the collection with the BLS Survey of Occupational Injuries and Illnesses. BLS continues to research methods to evaluate ways to utilize OSHA administrative data with SOII data in order to reduce respondent burden. BLS modified its Internet Data Collection Facility for the 2018 SOII to add collection of OSHA ID from SOII respondents who are also required to report to OSHA. BLS hopes collection of this information will assist in matching establishments reporting to both OSHA and the SOII. OSHA amended its reporting rules in 2019 to begin collecting the federal Employer Identification Number (EIN) as early as 2020, which BLS hopes will further facilitate matching of OSHA and BLS data. BLS will launch a new webpage in 2019 on which results from on-going BLS research to evaluate methods for utilizing OSHA data in the SOII will be posted.

5. Statistical responsibility

The Statistical Methods Group, Chief, Jeffrey Gonzalez is responsible for the sample design which includes selection and estimation. His telephone number is 202-691-7517. The sample design of the survey conforms to professional statistical standards and to OMB Circular No. A46.