

# Appendix A. Scope and Method of Survey

## Scope

The Survey of Occupational Injuries and Illnesses provides annual estimates of the frequency (incidence rates) and number of occupational injuries and illnesses based on logs kept by private industry employers. These logs reflect the year's injury and illness incidents and the employer's understanding of which cases were work related, under current recordkeeping guidelines provided by the U. S. Department of Labor. The number of injuries and illnesses reported in any given year may be influenced by changes in the level of economic activity, working conditions and work practices, number of hours worked, and worker experience and training.

Data published in this bulletin cover the 50 States plus the District of Columbia; data from the U.S. territories are not included.

The survey includes the following industries: agriculture, forestry, fishing, and hunting, North American Industry Classification System (NAICS) code 11; oil and gas extraction, NAICS 211; utilities NAICS 22; construction, NAICS 23; manufacturing, NAICS 31-33; wholesale trade, NAICS 42; retail trade, NAICS 44-45; transportation and warehousing, NAICS 48-49; information, NAICS 51; finance and insurance, NAICS 52; real estate and rental leasing, NAICS 53; professional, scientific, and technical services, NAICS 54; management of companies and enterprises, NAICS 55; administrative support and waste management and remediation services, NAICS 56; educational services, NAICS 61; health care and social assistance, NAICS 62; arts, entertainment, and recreation, NAICS 71; accommodation and food services, NAICS 72; other services, NAICS 81; and public administration, NAICS 92. Excluded from the survey are self-employed individuals; farms with fewer than 11 employees; employers regulated by other Federal safety and health laws, and Federal, State, and local government agencies.

Data conforming to definitions of recordable occupational injuries and illnesses for coal, metal, and nonmetal mining, and for railroad transportation are provided by the Mine Safety and Health Administration of the Department of Labor and the Federal Railroad Administration of the Department of Transportation.

The Occupational Safety and Health Administration is responsible for the collection and compilation of comparable data for Federal agencies. Although State and local government agencies are not surveyed for national estimates, several States have legislation that enables them to collect data for this sector.

The 2006 survey sample was composed of approximately 190,900 sample units. Original and follow-up mailings resulted in a response rate of 95 percent. From the selected establishments, approximately 296,082 injuries and illnesses with days away from work were used to obtain demographic and detailed case characteristic information.

## Sample

A two-stage sample selection process generates the survey estimates. The first stage involves selecting establishments. The second stage involves selecting the sample of cases involving days away from work which is derived from the sampled establishments.

Because the survey is a Federal-State cooperative program and the data must meet the needs of participating State agencies, an independent sample is selected for each State. The sample is selected to represent all private industries in the States and territories. The sample size for the survey is dependent upon (1) the characteristics for which estimates are needed, (2) the industries for which estimates are desired, (3) the characteristics of the population being sampled, (4) the target reliability of the estimates, and (5) the survey design employed.

For the establishment selection process, the total number of cases with days away from work, job transfer, or restriction is used as the base for the sample design. The sample design is based on cases with days away from work because it is considered the most important characteristic of the sample.

The important features of the sample design are its use of stratified random sampling with a Neyman allocation and a ratio estimator. The characteristics used to stratify the establishments are the North American Industry Classification System (NAICS) code and employment. Because these characteristics are highly correlated with an establishment's number and rate of reported injuries and illnesses, stratified sampling provides greater precision and, thus, results in a smaller sample size than simple random sampling. The Neyman allocation produces the minimum sample size that will provide an estimate with a fixed sampling variance. The precision of the sample is further improved, hence permitting a reduction in

sample size, by using the ratio estimator, which in turn uses employment data that are correlated with the characteristics that are to be measured.

The sample of cases involving days away from work is derived from the sampled establishments. For each sample unit, a predicted number of cases involving days away from work is calculated based on the establishment's industry and employment size. If the predicted number of cases is greater than 30, the establishment is requested to provide detailed demographic and case characteristics, and the number of days away from work for cases occurring at specified time intervals (for example, certain days of the month, only certain months). For establishments with a predicted caseload less than or equal to 30, the employer is requested to provide information for all cases involving days away from work occurring in the survey year. Either way, employers are permitted to attach supplementary forms that answer questions about individual days-away-from-work cases.

### **Estimating procedures**

*Sample unit weighting.* By means of a weighting procedure, sample units represent all units in their size class for a particular industry. The weight is determined by the inverse of the sampling ratio for the industry/employment-size class from which the unit was selected. Because a small proportion of survey forms are not returned, weights of responding employers in a sampling cell are adjusted to account for the nonrespondents. The respondents are then shifted into the estimating cell determined by the employment reported. Data for each unit are multiplied by the appropriate weight and nonresponse adjustment factor. The products are then aggregated to obtain a total for the estimating cell.

*Industry benchmarking.* Because the universe file that provides the sampling frame is not current to the reference year of the survey, it is necessary to adjust the data to reflect current employment levels. In the survey, all estimates of totals are adjusted by the benchmark factor at the industry level.

*Weighting cases involving days away from work.* Each case involving days away from work is weighted by the sample unit weight with which it is associated. In addition, each case is weighted to adjust for case subsampling and case nonresponse for those establishments that did not provide information on all cases with days away from work that occurred in their establishment in the survey year.

### **Federal-State cooperation**

To eliminate duplicate reporting by respondents and to ensure maximum comparability of estimates, survey respondents complete one reporting form, which is then used for national and State estimates.

### **Industrial classification**

Reporting units are classified into industries by their principal product or activity based on the 2002 edition of the North American Industry Classification System (NAICS). The data are tabulated according to this NAICS scheme.

### **Reliability of estimates**

The survey estimates of occupational injuries and illnesses are based on a scientifically selected probability sample, rather than a census of the entire population. These sample-based estimates may differ from the results obtained from a census of the population. The sample used was one of many possible samples, each of which could have produced different estimates. The variation in the sample estimates across all possible samples that could have been drawn is measured by the relative standard error. The relative standard error is used to calculate a "confidence interval" around a sample estimate.

The 95-percent confidence interval is the interval centered at the sample estimate and includes all values within two times the estimate's standard error. If several different samples were selected to estimate the population value (for example, injury and illness incidence rates), the 95-percent confidence interval would include the true population value approximately 95 percent of the time.

For example, the total injury and illness case incidence rate of 5.9 for construction (NAICS 23) in 2006 (table 1) had an estimated relative standard error of 1.5 percent. Hence, we are 95-percent confident that the interval between 5.7 and 6.1 (or  $5.9 \pm (1.5 \times 5.9 \times 0.02)$ ) includes the true rate for total cases for construction.

The relative standard errors for the rate of injuries and illnesses are given in table A-1, and for the number of injuries and illnesses in table A-2. Tables A-3 and A-4 provide the same information for new cases of occupational illnesses.

Because it would be too costly to calculate exact standard errors for all case and demographic estimates, models are used to approximate the standard errors on these estimates. These approximate values indicate the general magnitude of an estimate's standard error rather than its precise value. Detailed instructions for using each model are presented below.

## Notation

The following is the notation used in the relative standard error model approximations.

$\hat{C}_i$  = total number of cases with days away from work (DAFW) in the industry of interest.

$\hat{R}_i$  = incidence rate on the number of cases with DAFW in the industry of interest.

$F_i$  = factor used for the industry of interest. These values, for all published industries, are listed in table A-5.

$\hat{R}_{ic}$  = incidence rate on the number of cases with DAFW in the industry of interest for a specified characteristic.

$\hat{X}_{ic}$  = total number of estimated cases involving DAFW in the industry of interest for a specified characteristic.

$\hat{p}_{ic}$  = proportion of estimated cases with DAFW in the industry of interest for a specified characteristic.

$\hat{p}_{i \times d}$  = proportion of estimated cases involving DAFW that resulted in a specified classification of days given that the case was in the industry of interest (row percentage).

$\hat{p}_{icd}$  = proportion of estimated cases with DAFW in an industry for a specified characteristic that resulted in a specified classification of DAFW.

$\hat{p}_{ic \times d}$  = proportion of estimated cases involving DAFW in an industry that resulted in a specified classification of DAFW given that the case with DAFW had a specified characteristic (row percentage).

%RSE = percent relative standard error.

RSE = relative standard error =  $\frac{\%RSE}{100}$ .

*Totals.* This procedure shows the process for approximating the percent relative standard error (%RSE) on all estimates of totals.

The %RSE( $\hat{X}_{ic}$ ) can be computed with the following formula:

1)  $RSE(\hat{C}_i)$  (from table A-2 or A-4)

2)  $\hat{p}_{ic}$

3)  $F_i$  (from table A-5)

$$\%RSE(\hat{X}_{ic}) = 100 \times \sqrt{\left(RSE(\hat{C}_i)\right)^2 + \left[\left(\frac{1 - \hat{p}_{ic}}{\hat{p}_{ic}}\right) \times F_i\right] \times \left(1 + \left(RSE(\hat{C}_i)\right)^2\right)}$$

*Example 1:*

The %RSE on the estimated total number of DAFW cases that occurred in private industry to persons aged 55 to 64 (i = private and c = 55-64 years old) can be computed with the following:

$$RSE(\hat{C}_i) = \frac{\%RSE(\hat{C}_i)}{100} = \frac{1}{100} = 0.01 \quad (\text{from table A-2})$$

$$\hat{p}_{ic} = \frac{\hat{X}_{ic}}{\hat{C}_i} = \frac{134,030}{1,183,500} = 0.113$$

$$F_i = 0.000002532 \quad (\text{from table A-5})$$

Now substitute back into the equation:

$$\%RSE(\hat{X}_{ic}) = 100 \times \sqrt{(0.01)^2 + \left[ \left( \frac{(1-0.113)}{0.113} \times 0.000002532 \right) \times (1 + (0.01)^2) \right]}$$

$$\%RSE(\hat{X}_{ic}) = 1.1\%$$

*Rates.* This procedure shows the process for approximating the percent relative standard error on all estimates of incidence rates.

The %RSE( $\hat{R}_{ic}$ ) can be computed with the following:

1)  $RSE(\hat{R}_i)$  (from table A-1 or A-3)

2)  $\hat{p}_{ic}$

3)  $F_i$  (from table A-5)

$$\%RSE(\hat{R}_{ic}) = 100 \times \sqrt{\left( RSE(\hat{R}_i) \right)^2 + \left[ \left( \frac{(1-\hat{p}_{ic})}{\hat{p}_{ic}} \times F_i \right) \times (1 + (RSE(\hat{R}_i))^2) \right]}$$

*Example 2:*

The %RSE on the estimated incidence rate on DAFW cases in the construction industry whose source was chemicals and chemical products (i = construction and c = chemicals and chemical products) also can be computed as follows:

$$RSE(\hat{R}_1) = \frac{\%RSE(\hat{R}_1)}{100} = \frac{2}{100} = 0.02 \quad (\text{from table A-1})$$

$$\hat{p}_{1c} = \frac{\hat{X}_{1c}}{\hat{C}_1} = \frac{1,730}{153,180} = 0.011$$

$$F_1 = 0.000035762 \quad (\text{from table A-5})$$

Now substitute back into the equation:

$$\%RSE(\hat{R}_c) = 100 \times \sqrt{(0.02)^2 + \left[ \left( \frac{(1-0.011)}{0.011} \times 0.000035762 \right) \times (1+(0.02)^2) \right]}$$

$$\%RSE(\hat{R}_c) = 6.1\%$$

*Distribution of DAFW cases by days lost for private industry.* This procedure shows the method for approximating the %RSE on an estimate of distribution of DAFW cases by days lost for private industry (from tables 24-29). It also can be used to compute the percent relative standard error on an estimate of distribution of DAFW cases in private industry for a specified characteristic.

The %RSE(  $\hat{p}_{i_{cxd}}$  ) can be computed with the following formula:

1)  $\hat{p}_{i_{cd}}$

2)  $\hat{p}_{ic}$

3)  $F_i$  (from table A-5)

$$\%RSE(\hat{p}_{i_{cxd}}) = 100 \times \sqrt{F_i} \times \sqrt{\frac{(1 - \hat{p}_{i_{cd}})}{\hat{p}_{i_{cd}}} + \frac{(1 - \hat{p}_{ic})}{\hat{p}_{ic}}}$$

*Example 3:*

The %RSE on the proportion of DAFW cases in private industry that occurred to males which resulted in three to five DAFW (i= private, c = males, and d = 3-5 days) can be found as follows:

$$\hat{p}_{i_{cd}} = \frac{(\hat{p}_{i_{cd}}) \times \hat{X}_{ic}}{\hat{C}_i} = \frac{(0.184 \times 775,900)}{1,183,500} = 0.121$$

$$\hat{p}_{ic} = \frac{\hat{X}_{ic}}{\hat{C}_i} = \frac{775,900}{1,183,500} = 0.656$$

$$F_i = 0.00000253 \quad (\text{from table A-5})$$

Now substitute into the equation:

$$\% RSE(\hat{p}_{i \times d}) = 100 \times \sqrt{0.000002532} \times \sqrt{\frac{(1-0.121)}{0.121} + \frac{(1-0.656)}{0.656}}$$

$$\% RSE(\hat{p}_{i \times d}) = 0.4\%$$

*Example 4:*

The %RSE on the proportion of DAFW cases that occurred in the construction industry which resulted in 3 to 5 DAFW (i = construction and d = 3-5 days) can be computed as follows:

$$\hat{p}_{i \times d} = 0.172$$

$$F_i = 0.000035762 \quad (\text{from table A-5})$$

Now substitute those values into the equation:

$$\% RSE(\hat{p}_{i \times d}) = 100 \times \sqrt{0.000035762} \times \sqrt{\frac{(1-0.172)}{0.172}}$$

$$\% RSE(\hat{p}_{i \times d}) = 1.3\%$$

*Example 5:*

The %RSE on the proportion of DAFW cases that occurred in the construction industry which resulted from falls to a lower level (i = construction and d = falls to lower level) can be computed with the following formula:

$$\hat{p}_{i \times d} = 0.119$$

$$F_i = 0.000051648 \quad (\text{from table A-5})$$

Now substitute those values into the equation:

$$\% RSE(\hat{p}_{i \times d}) = 100 \times \sqrt{0.000051648} \times \sqrt{\frac{(1-0.119)}{0.119}}$$

$$\% RSE(\hat{p}_{i \times d}) = 1.6\%$$

**Nonsampling error**

Although not measured, nonsampling errors will always occur when statistics are gathered. The inability to obtain information about all cases in the sample, mistakes in recording or coding the data, and definitional difficulties are general examples of nonsampling error in the survey. BLS has implemented quality assurance procedures to reduce nonsampling error in the survey, including a rigorous training program for State coders, mechanical edits that identify questionable entries, and a continuing effort to encourage survey participants to respond fully and accurately to all survey elements.

**Publication guidelines**

The estimating procedure generates occupational injury and illness estimates for approximately 1200 NAICS codes. This publication, however, excludes estimates for industry codes if one of the following situations occurred:

- Annual average employment for the industry in 2006 was fewer than 10,000. However, data for an industry with an annual average employment of fewer than 10,000 were published if the majority of the employment was reported in the survey.

- The relative standard error on total cases with days away from work, job transfer, or restriction for the industry exceeded a specified limit.

- The benchmark factor for the industry was less than .70 or greater than 1.49.

- Publication might disclose confidential information.

Data for an unpublished industry are included in the total for the broader industry level of which it is a part. Also, selected items of data were suppressed for publishable industries if the sampling error for the estimate exceeded a specified limit, typically 65 percent at the national level.

For the case characteristics and demographic data, a sliding scale methodology is used based on the relative standard error of the estimate. Estimates are rounded to the nearest tenth and are suppressed if one of the following situations occurred:

- The number of cases is fewer than 15.
- The number of cases is 15 or greater and the relative standard error for the estimate exceeds a specified limit.