

B. Collection of information employing statistical methods.

The statistical methods used in the sample design of the survey are documented thoroughly in internal documents, referenced throughout this summary. These documents are listed below and can be obtained from the Statistical Methods Group by calling Gwyn Ferguson at 202-691-6941.

**SRS 2008 Users Guide** -- provides a detailed description of the sample design

**Overview of SRS Allocation Methodology** (02/03/05) -- provides an analysis describing why the current allocation methodology was chosen

**SOII Optimal Allocation** (02/10/05) -- describes the sample allocation process in detail

**SOII Estimation** (11/30/06) -- describes the estimation process

**OSH Variance Estimation Requirements** -- describes the process for calculating the sample error estimates for the summary estimates

**Methods Used to Calculate the Variances of the OSH Case and Demographic Estimates** -- describes the process for calculating the sampling error estimates for the Case and Demographic estimates

1. Description of universe and sample.

Universe. The potential number of respondents (establishments) covered by the scope of the survey is 6.5 million, although only about 1 million employers keep records on a routine basis due to recordkeeping exemptions for employers in low hazard industries and employers with less 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:

Sector 11 - Agriculture, Forestry, Fishing, and Hunting

Sector 21 - Mining

Sector 22 - Utilities

Sector 23 - Construction

Sectors 31, 32, and 33 - Manufacturing

Sector 42 - Wholesale Trade

Sectors 44 and 45 - Retail Trade

Sectors 48 and 49 - Transportation and Warehousing

Sector 51 - Information

Sector 52 - Finance and Insurance

Sector 53 - Real Estate and Rental and Leasing

Sector 54 - Professional, Scientific, and Technical Services

Sector 55 - Management of Companies and Enterprises

Sector 56 - Administrative and Support and Waste Management and Remediation Services

Sector 61 - Educational Services

Sector 62 - Health Care and Social Assistance

Sector 71 - Arts, Entertainment, and Recreation

Sector 72 - Accommodation and Food Services

Sector 81 - Other Services (except Public Administration)

Excluded from the national survey collection are self-employed individuals; farms with fewer than 11 employees; employers regulated by other Federal safety and health laws; and Federal government agencies. Survey year 2008 will mark the first time State and local governments agencies data will be collected for all States and published for all States and the nation as a whole.

Sample. A stratified probability sampling design is used for the survey, and the sample is selected by using a systematic sampling procedure with a random start for each stratum. Based on the survey's design criteria, a sample of 230,000 private and government respondents from a potential universe of 6.5 million establishments is required. For survey year 2008 and the collection of State and Local government data, the sample size will be 237,000. The units on the frame are stratified based on geography, industry, and employment. Since the survey is a Federal-State cooperative effort, the first characteristic used to stratify the units is the State; this enables all the State grantees participating in the survey to produce estimates at the State level. The units are further stratified by industry and employment size class. See **SRS 2008 Users Guide** for a detailed description of the sample design.

Response rate. The survey is a mandatory survey with an overall usable response rate of 92.0 percent based on the 2005 survey data (latest survey for which data are tabulated). The response rate for future surveys is expected to be about the same.

## 2. Statistical methodology.

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

While there were many characteristics upon which the national design could have been used to allocate the sample, the BLS elected to use the total recordable case rate. This was considered by BLS to be one of the most important characteristics. The analysis in choosing this characteristic can be found in **Overview of SRS Allocation Methodology** (02/03/05).

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.

Sample procedure. The principal feature of the survey probability sample design is its use of stratified random sampling with a Neyman allocation. The characteristics used to stratify the units are the State, 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. The sample allocation process is described in detail in ***SOII Optimal Allocation*** (02/10/05).

As mentioned before, a probability sample is selected by using a systematic sampling procedure with a random start for each stratum.

The survey will be conducted by mail questionnaire through the BLS-Washington and Regional Offices and participating State statistical grant agencies.

Estimation procedure. The survey's estimates of the population total are based on the Horvitz-Thompson estimator, which is an unbiased estimator. The precision of the estimates is further improved by using the ratio estimator that utilizes available auxiliary information to improve the estimate. The estimates of the incidence rates are calculated as:  $N/EH \times 200,000$ , where:

N = number of injuries and illnesses  
EH = 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 process is illustrated in a slideshow presentation titled ***SOII Estimation*** (11/30/06).

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.

All State statistical grant agencies participating in the survey produce industry estimates at the State level. To enable these States to produce reliable estimates, their samples are supplemented with additional units. These supplemental units also are used in the national estimates.

### 3. Statistical reliability.

Response rates and nonresponse adjustment. The survey is a mandatory survey, with the exception of selected voluntary units in State and Local government. The survey achieves an overall usable response rate of 92.0 percent. The following techniques are employed to help maximize survey response:

- o A follow-up mailing to nonrespondents in February.
- o A follow-up mailing to nonrespondents in April.
- o A telephone or mail follow-up of key nonrespondents in May.

This high level of response greatly aids in protecting the survey estimates from nonresponse bias. The data for the remaining nonrespondents are imputed from the respondent data using a weighting cell adjustment technique.

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

Variance estimators are applied to the survey data to produce estimates of relative standard errors. A Taylor Series approximation is used in calculating standard errors on injury and illness counts. A model based estimator is used in calculating estimates of standard errors for case and demographic data. Specifications for the calculations of these sampling errors are found in ***OSH Variance Estimation Requirements*** for the summary estimates and in ***Methods Used to Calculate the Variances of the OSH Case and Demographic Estimates*** for the Case and Demographic estimates of standard errors.

#### 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 230,000. The BLS is continually researching for 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.

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

5. Statistical responsibility.

The Statistical Methods Group, Chief, Gwyn Ferguson is responsible for the sample design which includes selection and estimation. Her telephone number is 202-691-6941. The sample design of the survey conforms to professional statistical standards and to OMB Circular No. A46.