

Supporting Statement  
Report on Occupational Employment and Wages

**B. Statistical Methodology**

1. Universe and Sample Size Summary

The following table shows the estimated number of universe units, sampled units, and responding units for all in-scope NAICS by Fiscal Year for the regular OES survey.

Survey year	Mandatory / voluntary	Ownership	NAICS Coverage	Estimated Responding Units	Estimated burden hours
FY 2007	Voluntary	private	1133, 1151, 1152, 21-81 (exc. 814)	285,099	213,824
	Voluntary	government	State and local government	13,361	10,021
	Mandatory	private	1133, 1151, 1152, 21-81 (exc. 814)	16,356	12,267
	Mandatory	government	State and local government	1,084	813
	Total			315,900	236,925

Survey year	Mandatory / voluntary	Ownership	NAICS Coverage	Estimated Responding Units	Estimated burden hours
FY 2008	Voluntary	private	1133, 1151, 1152, 21-81 (exc. 814)	285,099	213,824
	Voluntary	government	State and local government	13,361	10,021
	Mandatory	private	1133, 1151, 1152, 21-81 (exc. 814)	16,356	12,267
	Mandatory	government	State and local government	1,084	813
	Total			315,900	236,925

Survey year	Mandatory / voluntary	Ownership	NAICS Coverage	Estimated Responding Units	Estimated burden hours
FY 2009	Voluntary	private	1133, 1151, 1152, 21-81 (exc. 814)	285,099	213,824
	Voluntary	government	State and local government	13,361	10,021
	Mandatory	private	1133, 1151, 1152, 21-81 (exc. 814)	16,356	12,267
	Mandatory	government	State and local	1,084	813

		government	
Total	315,900	236,925	

A goal of the OES survey is that each State achieves an 80 percent response rate. The overall response rate for the 2005 survey was approximately 78 percent based on units. (See section 3(a) for more information about this topic.)

The universe size in 2006 was approximately 6.5 million establishments. We expect the OES universe to continue growing by about 100,000 units per year.

## 2. Collection Procedures

### 2(a) Sample Design

The Enhanced Quarterly Unemployment Insurance (EQUI) file is the survey’s primary sampling frame. The Division of Business Establishment Systems maintains this file. The EQUI file will be supplemented by the Federal Railroad Administration’s list of railroads because railroad workers are not covered by regular Unemployment Insurance.

Units on the sampling frame are stratified by Metropolitan Statistical Area (MSA) and Balance of State code, by four-or-five-digit NAICS industry code, and by size of establishment based on employment. A variation of Neymann allocation procedure is used to allocate the sample to each State-MSA/4-5digit NAICS cell. Within each cell, the sample is selected using probability proportional to estimated employment size. Each panel sample is designed to represent the frame. Consequently, each sampled establishment is assigned a sampling weight equal to the reciprocal of its probability of selection in the sample.

Recent changes in the sampling procedures introduced for collection efficiencies may result in a small downward bias in employment estimates in some industries for some areas. This bias is estimated to be between 0.1 and 0.2 percent of total employment. This may occur in cases where the single panel allocation can be rounded to either 0 or 1. In the case when the allocation is 0 for all 6 panels contributing to the 6-panel estimate, the estimated employment for the cell is 0, while in all other cases, the expected employment is correct. This bias is mitigated at higher levels of aggregation through benchmarking. BLS will research this bias to determine whether changes in sampling procedures are warranted.

Occupational employment data from prior survey rounds are used by BLS-Washington to produce sample allocations that result in relative standard errors (RSE) on mean wages of 10 to 20 percent for the typical occupations in each MSA/ four-or-five-digit industry cell. Estimates for typical occupations at higher aggregate levels of area and industry will have substantially smaller relative standard errors.

Each year, semiannual panels of about 202,000 establishments each are selected for the May and November surveys.

### 2(b) Estimation Process

Estimates of occupational employment and wage rates are produced using data from the current semiannual panel’s survey and survey data from the five semiannual panels prior to the current panel (a total sample of

about 1,200,000 establishments). Data from several panels are combined in order to reduce the sampling error of the estimates at detailed levels (MSA by 4-5 digit NAICS). Combining samples from six panels increases the population counts at the total level by six fold. However, not all cells have sample in them in each panel because many of the detailed cells are sparsely populated. To avoid a multiple count of up to six fold, the sampling weight for each establishment in a "State-MSA/4-5 digit sampling cell" is reduced by a factor  $1/d$ ; where,  $d$  is a counter indicating if the sampling cell has establishments from 1, 2, 3, 4, 5, or 6 panels. The adjusted sampling weight is used to calculate the employment and wage estimates. Additional information about estimation procedures for employment and wages can be found in the attached technical note, which will be published on the BLS web site in May or June of 2007. (Please see attachment V, Survey Methods and Reliability Statement for the May 2005 Occupational Employment Statistics Survey)

The OES survey calculates estimates of occupational employment for each MSA/ four-or-five-digit industry cell within a State or for the Nation. The estimation process begins with an edit procedure to identify and correct inconsistent or incomplete data on the data file. The procedure also identifies and makes adjustments to atypical reporting units. Afterwards, a hot-deck nearest-neighbor imputation is used to impute occupational staffing patterns for the nonresponding units. Next, a mean of cell imputation procedure is used to impute for missing occupational wage employment data. After the data are edited and imputed, the  $d$ -weighting as described in the previous paragraph is processed. Finally, the weighted sampled employment totals are ratio adjusted, or benchmarked, to known employment totals. These known employment totals are extracted from the Bureau's EQUI files. The sampling weight of each unit is multiplied by the benchmark factor to produce a final weight value for the unit. Estimates of occupational employment at the MSA/ four-or-five-digit industry level or at the MSA/ three-digit industry level are obtained by summing the lower-level cell estimates. Each occupational employment estimate has a standard error that is calculated using a random group jackknife variance estimator.

Mean wage and median wage estimates are calculated for each occupation within an MSA/ four-or-five-digit industry cell (and these are summed to higher levels). OES wage rate data are collected in broad wage bands instead of exact data points, so the mean wage rate for each wage band is obtained externally from the Bureau's National Compensation Survey (NCS). The inflation factor from the Employment Cost Index is used to update wage data collected in past panels to be used in current wage estimates. To approximate median wage rates, a distribution is assumed for each wage interval and simple linear interpolation between the endpoints of the OES wage intervals is used. Background on median wage estimators used in OES can be found on <http://www.bls.gov/ore/pdf/st990160.pdf>.

Because wage rate data are collected in broad wage bands instead of exact data points, the standard error for each mean wage rate estimate is calculated using a components model. This model accounts for the variability of both the observed and unobserved components of wage rate data. A traditional ratio variance estimator is used to account for the variability observed in the collected OES wage data. Since the mean wage rate for each wage band is obtained externally from the Bureau's National Compensation Survey (NCS), there are unobserved components of wage rate variance that are modeled. Detailed wage data collected by the NCS are used to estimate the variability of the unobserved components.

## 2(c) Accuracy

Before occupational employment estimates are released to the public, they must first be screened to ensure that they do not violate the Bureau of Labor Statistics' (BLS) confidentiality pledge. A promise is made by the Bureau to each sampled unit that the BLS will not release its employment data to the public in a manner that would allow others to identify the unit. If an occupational employment estimate fails confidentiality screening, the estimate is suppressed.

After confidentiality screening, the estimates (including confidentiality collapsed estimates) are screened a second time to ensure that they satisfy the Bureau's publishability standards. Among them, employment estimates must have an associated relative standard error of 50 percent or less, there must be at least two responses (i.e., not imputed) for the occupation, and the employment estimate must be 10 or greater.

#### 2(d) Problems

The OES survey, in order to produce wage rate and employment estimates at detailed geographic levels, combines data across a three-year time period (six semiannual panels). Special sampling procedures are in place to allocate the sample, to limit the inclusion of units to once in a three-year time period, and to combine the data to produce estimates. Among these special procedures are procedures to update, or "age" the previous years wage data to reflect the current time period. Background on procedures for wage updating can be found on: <http://www.bls.gov/ore/pdf/st000080.pdf>. BLS continues to conduct research to evaluate the effectiveness of the updating process, and to improve it where possible. Collecting all of the certainty units each year would allow these data to be used in the validation of the updating process; we are evaluating several collection options with respect to these units.

#### 2(e) Frequency

The OES program employs the following sampling methodology. Occupational employment and wage-range data are collected for all nonagricultural industries over a six-panel semiannual cycle, with data collected in May and November of each year. In each panel, one sixth of a 1,200,000 establishment sample is selected for nonagricultural industries, where overlap with the prior five panels' samples is minimal. Thus, establishments will be included in the sample at most once every three years.

### 3. Methods of Maximizing Response

#### 3(a) Response

The OES survey is a voluntary survey with an overall response rate of approximately 78 percent based on units. Each State is responsible for collecting the questionnaire forms mailed to the sample units selected in their State. Every effort is made to maximize response rates to achieve the 80-percent goal by:

- Surveying sampled units at most once every three years (once every six panels);
- Providing each sampled unit with a cover letter explaining the importance of the survey and the need for voluntary cooperation;
- Giving each sampled unit the Bureau's pledge of confidentiality;
- Sending each nonresponding unit two to three additional mailings after the initial mail-out (if necessary); the BLS also recommends that the States obtain specific contact names for each sampled firm;
- Contacting key nonresponding units in each MSA/ four-or-five -digit industry or MSA/ three-digit industry cell by telephone;
- Contacting critical nonresponding units through personal visits (if necessary);
- Using status reports and control files to identify MSA/industry cells with low response rates;
- Requesting that States consider making initial personal visits to firms identified as requiring special attention; and,
- Stressing to respondents that assistance is available to help them complete the survey form.

- Providing a link on the survey form to OES data on the BLS Internet website to demonstrate program usefulness.
- Using a respondent web page that provides detailed information about responding to the OES survey, including state contact information for those needing assistance
- Increasing the use of electronic and telephone collection in order to allow the respondent to provide information in a way that is most convenient to them.
- Conducting a pilot response rate improvement test that included changes to the state letter and an information brochure to determine whether response rates can be improved using these methods.

### 3(b) Nonresponse Adjustment

A hot deck (nearest neighbor) imputation procedure is used to impute for unit nonresponse. This type of nonresponse occurs when a unit reports no employment data. In this procedure, units in the sample are stratified into ‘Year / State / 5-digit industry’ cells. Within each cell, the “nearest” donor (i.e., responding unit) is selected for each nonrespondent; an algorithm is used, which minimizes the possibility of selecting the same donor twice. The nonrespondent’s sampling frame employment and the donor’s summed total employment are used within a cell to match donors with nonrespondents. Once a donor and nonrespondent are matched, the occupational employment totals of the nonrespondent are computed using the proportional distribution of the donor’s occupational employment totals. In the event that a donor is not available at the ‘Year / State / 5-digit industry level, the procedure advances to succeeding higher-level cells until a donor is found. Background on nearest neighbor imputation procedures for OES can be found on <http://www.bls.gov/ore/pdf/st950180.pdf>.

A variation of the mean imputation procedure is used to impute for item nonresponse. This type of nonresponse occurs when a unit reports the total-employment for an occupation but not the corresponding wage-employment. Units where occupational employment was imputed in the previous step are also included as nonrespondents in this procedure. In this procedure, units in the sample are stratified into ‘Year / MSA / 4-digit industry / Size class’ cells. A wage-employment distribution is then calculated for those occupations missing wage-employment based on the usable data in the cell. Missing wage-employment is then imputed by using that distribution to prorate the reported total-employment across the wage intervals.

### 3(c) Reliability

A probability based sample design is used to develop the OES survey. This design allows the Bureau to control and measure the sampling error of the occupational employment and wage rate estimates. Relative standard error estimates are used to measure sampling error. A random group jackknife variance estimator is used to estimate the relative standard errors for the occupational employment estimates. A variation of the stratified ratio estimator is used to estimate the relative standard errors for the mean wage estimates.

Background on the variance estimator used in OES can be found on [http://www.amstat.org/Sections/Srms/Proceedings/papers/1997\\_081.pdf](http://www.amstat.org/Sections/Srms/Proceedings/papers/1997_081.pdf).

Nonsampling errors, unlike sampling error, are not easily measurable but can be controlled through various quality control procedures. One such procedure is the requirement that all States use data processing software provided by the BLS national office. This standardization and automation of survey operations should reduce several sources of nonsampling error. State and BLS staff use automated and manual data screening procedures that help identify data that was misreported, or keypunched incorrectly.

#### 4. Developmental Tests

The OES survey is a constantly evolving data collection process. Pretest studies of occupations for selected industries were conducted in the 1960s and 1970s prior to the first OES survey in 1971. Pilot studies were conducted in 1989 and 1990 to determine the feasibility of collecting occupational wage-employment data through the OES survey. And over the years, improvements have been made to this survey through better machinery, better operational procedures, and improved sampling and estimation methodology.

The OES survey maintains ongoing efforts to reduce respondent burden. These efforts include periodic evaluations of the effectiveness of the materials mailed to each sampled unit. These evaluations encompass most aspects of the data collection environment. A continuous feedback system is in process which asks for improvements in the survey questionnaire, the content of the form, the instructions on how to fill out the form, and other issues related to respondent burden.

Also contributing to our goal of increasing quality while not increasing respondent burden, the BLS is in the process of testing a pilot initiative that provides the respondent with additional materials and an improved solicitation letter in order to increase response rate.

OES occasionally conducts developmental tests to measure affects on data quality and response rates. These tests typically are expected to reduce response burden; increases in response burden, if any, would be expected to be minimal.

#### 5. Statistical Responsibility

Ms. Shail Butani, Chief, Statistical Methods Division, Office of Employment and Unemployment Statistics, Bureau of Labor Statistics, is responsible for the statistical aspects of the OES survey. Her telephone number is (202) 691-6347.