

Supporting Statement

Report on Occupational Employment and Wages

B. Collection of Information Employing Statistical Methods

1(a) Respondent Universe

The universe for this survey consists of the Quarterly Contribution Reports (QCR) filed by employers subject to State Unemployment Insurance (UI) laws. The U.S. Bureau of Labor Statistics (BLS) receives these QCR for the Quarterly Census of Employment and Wages (QCEW) Program from the 50 States, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. The QCEW data, which are compiled for each calendar quarter, provide a comprehensive business name and address file with employment, wage, detailed geography (i.e., county), and industry information at the six-digit North American Industry Classification System (NAICS) level. This information is provided for over nine million business establishments of which about 6.7 million are in the scope of this survey. The final data is stored in a Longitudinal Data Base (LDB), which is then used as a sampling frame for sample selection. Similar data for Federal Government employees covered by the Unemployment Compensation for Federal Employees program (UCFE) are also included. Other data used for sampling include the universe of railroad establishments obtained from Federal Railroad Administration and a universe of establishments in the U.S. territory of Guam obtained from the Government of Guam, Department of Labor.

1(b) Sample

Scope--The OES measures occupational employment and wage rates of wage and salary workers in nonfarm establishments in the 50 States and the District of Columbia. Guam, Puerto Rico, and the Virgin Islands are also surveyed, but their data are not included in national estimates. The survey covers the following NAICS industry sectors:

11	Logging (1133), support activities for crop production (1151), and support activities for animal production (1152) <i>only</i>
21	Mining
22	Utilities
23	Construction
31-33	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 [private households (814) are excluded]
 99 Federal, State, and local government (OES designation)

Sample Size--The sample size is approximately 1.2 million establishments over a 3-year period. The sample is divided into six panels over three years with two semi-annual samples of about 200,000 establishments selected each year. 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:

Table 1: Universe and Sample Size Summary

Survey	NAICS Coverage	Responding Units	Sample Units	Universe Units
FY 2010 (November 2009 & May 2010 panels)	All in-scope NAICS	301,000	404,000	6,700,000
FY 2011 (November 2010 & May 2011 panels)		300,000	403,000	6,700,000
FY 2012 (November 2011 & May 2012 panels)		298,000	402,000	6,700,000
6-panel Totals		899,000	1,209,000	6,700,000

Stratification--Units on the sampling frame are stratified by State/Metropolitan Statistical Area (MSA), Metropolitan Division, and Balance of State, and by four-, five-, or six-digit NAICS industry code. The frame is further stratified into certainty and non-certainty portions for sample selection. Certainty units include Federal and State governments, hospitals, railroads, and large establishments. These are sampled with probability equal to one every 3-year cycle because of their occupational employment coverage and economic significance. All remaining establishments are non-certainty establishments and are selected with probability less than one but greater than zero.

2(a) Sample Design

Allocation method--A variation of Neyman allocation procedure called a Power Allocation (Bankier, 1988) is used to allocate the remaining non-certainty sample to each State-MSA/4-5-6-digit NAICS stratum. The allocation methodology balances employment size of areas and industries with the variability of the occupational employment in each industry. The use of the power allocation shifts sample away from very large areas and industries to medium and smaller ones allowing more comparable estimates for smaller domains. The power allocation is calculated as follows:

$$n_h = n \cdot \frac{\sqrt{X_h} \cdot S_h}{\sum_{h=1}^H (\sqrt{X_h} \cdot S_h)}$$

Where,

h = the stratum defined as defined as State-MSA/4-5-6 digit NAICS industry

n_h = non-certainty sample allocated to stratum h

n = national sample size less the number of certainty units

X_h = non-certainty frame employment in stratum h

S_h = average occupational variability within stratum h

Additionally, OES ensures that a minimum sample size is allocated to each sample stratum such that the final sample allocation for each stratum is equal to the maximum of the minimum allocation and the power allocation.

Sample Selection--Within each stratum, the sample is selected using probability proportional to estimated employment size with large units being selected with certainty. Each semi-annual panel sample is designed to represent the frame. Every attempt is made to ensure that private and local government establishments are only selected once every three years. Consequently, each sampled establishment is assigned a sampling weight equal to the reciprocal of its probability of selection in the sample. Note: Censuses of federal and state government are collected annually.

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-, five-, or six-digit industry cell. Mean wage estimates for typical occupations at higher aggregate levels of area and industry will have substantially smaller relative standard errors.

Frequency of Sampling--Each year, semiannual panels of about 200,000 establishments each are selected for the May and November reference periods.

Sampling Issues--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.

More detailed information about OES sample allocation and selection procedures can be found in Chapter 3 of the BLS Handbook of Methods (<http://www.bls.gov/opub/hom/homch3.htm>).

2(b) Estimation Procedures

Annual estimates of occupational employment and wage rates are produced using data from the current May 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 4-5-6 digit NAICS). Combining samples from six panels increases the sample 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/4-5-6 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 six-panel combined sample weight is used to calculate the employment and wage estimates.

Employment Estimation

The OES survey publishes estimates of occupational employment for four-, five-, or six-digit industry cells within States and the Nation and cross-industry cells for MSAs. The estimation process begins with an edit procedure to identify and correct inconsistent or incomplete data on the 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 Enhanced Quarterly Unemployment Insurance (EQUI) files, which are the raw state micro-data used to create the longitudinal links for the QCEW. The sampling weight of each unit is multiplied by the benchmark factor to produce a final weight value for the unit. The following equation is used to calculate occupational employment estimates for an estimation cell defined by geographic area, industry group, and size class:

$$\hat{X}_{ho} = \sum_{i \in h} (w_i \text{ BMF}_i x_{io})$$

Where,

o = occupation;

h = estimation cell;

w_i = six-panel combined sample weight for establishment i ;

BMF_i = final benchmark factor for establishment i ;

x_{io} = reported employment for occupation o in establishment i ;

\hat{X}_{ho} = estimated employment for occupation o in cell h

Each occupational employment estimate has a standard error that is calculated using a random group jackknife variance estimator.

Wage Estimation

Mean wage and median wage estimates are calculated for each occupation within an MSA/four-, five-, or six-digit industry cell (and these are summed to higher levels). Traditionally OES wage rate data are collected in broad wage bands instead of exact data points. In 2009, OES began incorporating exact point data for federal and state governments for wage and percentile estimates.

For data that is collected in wage bands, 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, OES assumes a uniform distribution within wage intervals and uses a simple linear interpolation between the endpoints of the wage intervals. Background on median wage estimators used in OES can be found on <http://www.bls.gov/ore/pdf/st990160.pdf>.

Because most 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.

As mentioned above, OES began incorporating federal and state point data for both wage and percentile estimates in 2009. State governments report some of their data as exact wage rates and some in interval bands. OES receives a census of occupational employment and exact wage information from these reporters each year so these do not contribute to the sampling error for the survey. Furthermore, these reporters are currently out-of-scope for the NCS and would not introduce bias by using exact wages in the estimator instead of the NCS means.

For more detailed information about estimation procedures for employment and wages please see Bureau of Labor Statistics' Handbook of Methods, Chapter 3 and Appendix M of the OES State Operations Manual (<http://199.221.111.170/program/oes/documentation/StOpsApps%20Dec2012.pdf>).

2(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.

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.

2(d) Special Procedures

In order to produce wage rate and employment estimates at detailed geographic levels, the OES 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 methods 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) Data Collection Cycles

Occupational employment and wage-range data are collected for all nonagricultural industries over a six-panel semiannual cycle, with data collected for a reference period of 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(a) Maximizing Response

A goal of the OES survey is that each State achieves an 80 percent response rate. The overall response rate for the 2011 survey was approximately 77 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). With the research to improve time series capabilities of the OES survey this constraint may be relaxed for some employers.
- Conducting extensive address refinement to ensure that the survey form reaches the correct establishment in a timely manner.
- Providing each sampled unit with a cover letter explaining the importance of the survey and the need for voluntary cooperation.
- Giving each private sector sample 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).
- Including brochures that explain the many uses of the OES data.
- 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.
- 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.
- Providing email and online data submission options.
- Advertising electronic submission options to certain large establishments.
- Coordinating collection for multi-unit establishments.

Response Burden--The following table shows the estimated response burden by Mandatory/Voluntary and Ownership:

Table 2: Response Burden

Survey year	Mandatory / voluntary	Ownership	NAICS Coverage	Estimated Responding Units	Estimated burden hours
FY 2013	Voluntary	private	1133, 1151, 1152, 21-81 (exc. 814)	274,231	205,673
	Voluntary	government	State and local government	5,855	4,391
	Mandatory	private	1133, 1151, 1152, 21-81 (exc. 814)	29,352	22,014
	Mandatory	government	State and local government	630	472
	Total			310,068	232,550
Survey year	Mandatory / voluntary	Ownership	NAICS Coverage	Estimated Responding Units	Estimated burden hours
FY 2014	Voluntary	private	1133, 1151, 1152, 21-81 (exc. 814)	274,231	205,673
	Voluntary	government	State and local government	5,855	4,391
	Mandatory	private	1133, 1151, 1152, 21-81 (exc. 814)	29,352	22,014
	Mandatory	government	State and local government	630	472
	Total			310,068	232,550
Survey year	Mandatory / voluntary	Ownership	NAICS Coverage	Estimated Responding Units	Estimated burden hours
FY 2015	Voluntary	private	1133, 1151, 1152, 21-81 (exc. 814)	274,231	205,673
	Voluntary	government	State and local government	5,855	4,391
	Mandatory	private	1133, 1151, 1152, 21-81 (exc. 814)	29,352	22,014
	Mandatory	government	State and local government	630	472
	Total			310,068	232,550

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 or 6-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 or 6-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) Non-Response Bias Research

Extensive research has been previously done to assess whether the non-respondents to the OES survey differ systematically in some important respect from the respondents of the survey and would thus bias OES estimates. A study of comparison of average hourly occupational wages from OES survey were compared to those of another of the Bureau’s programs, the National Compensation Survey (NCS), used for Employment Cost Index and for President’s Pay Agent in setting locality pays. The comparisons were done at the national level and at various MSA levels. This research showed that the mean hourly wages for 70% of the occupations were not statistically different between the two surveys and an additional 10% were not economically different; that is, the two mean occupational wages were within 10% of each other. Many of the remaining 20% of the occupations had wage discrepancies that could be explained by conceptual differences and coding inconsistencies, etc. between the two programs. An important finding from this report was the differences in mean occupational wages between the two surveys were in both directions with a 40-60% split. These findings were the same at the National and MSA levels (BLS internal report “Comparing OES and NCS Wage Estimates”, November 2006 (final report), by Tony Barkume, Matt Dey, Larry Ernst, Maury Gittleman, and Anne Polivka). The results from this study show that the nearest neighbor imputation method and the wage imputation method as described in this document are performing a reasonable adjustment for missing data. Future plans include repeating this type and other similar analysis of non-response bias.

3(d) Confidentiality

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 private sector sample 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.

3(e) Publishability

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. For wage estimates, the relative standard errors must be 30 percent or less.

4. Developmental Tests

The OES data collection process continues to evolve over time. 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 a 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.

OES occasionally conducts developmental tests to measure effects on data quality and response rates. These tests typically are used to evaluate procedures designed to reduce response burden; increases in response burden, if any, would be expected to be minimal. If funding is available, OES would like to conduct a Response Analysis Survey (RAS). The purpose of the RAS would be to re-contact respondents to ask questions about the OES survey and look for ways to improve the OES survey and reduce respondent burden.

Future Research:

Subject to funding availability, OES would like to conduct tests to investigate the viability of collecting additional information from employers, which could potentially improve the quality of occupational wages, reduce response burden on employers, allow for further analysis of factors affecting occupational wages, and lead to expansion of OES data products. BLS would like to ask employers to report data items that many already report without solicitation. Currently some employers provide electronic files of employees with actual wage rates rather than wages reported in intervals because it requires less time and effort. Often employers include additional data items such as the number of hours the person is paid for. Multiunit companies often report occupational wage data for all establishments rather than just providing data for the requested sampled establishments. Since so many employers provide this information without being asked, BLS would like to explore how employers would respond if we specifically asked employers to provide these data. BLS is planning to develop more detailed proposals on testing the collection of actual wage rates, additional data items, and additional establishments. This research may use a split panel design, and may lead to reduced burden hours. Sample designs, survey instruments, and survey methods need to be detailed.

a. Using wage rate data in combination with interval data.

Many employers find it easier to provide wage rate data rather than converting their data to wage intervals. Many employers already do this and BLS or the SWAs convert the reported data to intervals. We would like to provide employers the option of filling out a survey form with wage rates, or providing their data in weblite by wage rate. In addition to adapting to respondents preferred method of reporting, this data is valuable in conducting research on methodology improvements.

Examination of wage rate data already provided by respondents on a volunteer basis has suggested that improvements in methodology can be obtained by capturing the wage rate data from employers. This is especially true when wages are clustered, such as for minimum wage workers, or for union workers within a particular industry or area. In addition, wages for highly paid occupations are affected by assigning them to predefined intervals. In order to study this more fully, and incorporate it into our methodology, BLS is considering targeting specific employers to report their data by wage rate rather than wage range. In the meantime, BLS will be exploring ways to incorporate the data from employers that choose to report that way.

b. Asking employers for additional data elements.

Many employers already provide many data elements in their electronic OES report that we do not ask for. These data elements include information that is requested by customers, but cannot be provided by OES or other BLS surveys. For example, establishments report data items along with the occupation and wages such as: part-time or full-time status, hours, whether or not employees are exempt from the Fair Labor Standards Act, gender, age, EEO category, union status, specific job title, department, and others. While some of these occupational characteristics are available from other BLS sources, none are available for states and all areas, and in the case of demographic data, they cannot be associated with a particular employer's industry or size, and are not available for many occupations. Examination of the data already provided shows a wealth of information that will be the subject of an upcoming BLS article. A Response Analysis Survey (RAS) conducted in 2011-12 showed that most employers are willing to provide additional data like hours worked and part-time/full-time status.

c. Asking employers to report for all their establishments instead of randomly selected establishments.

Many employers provide comprehensive electronic data files or data dumps containing payroll data for all of their establishments every year, rather than providing data for just the sampled establishment. The OES analyst sorts through the reports, and matches them to the sampled units, saving the respondent the burden of doing so. The OES analysts ignore the unsolicited establishments. Some of the volunteered establishments might be included in the OES sample a different year, and the newer data will be solicited. For the units that aren't in the 6-panel sample used for estimates, their inclusion might help local area estimates. Capturing newer data for units that are in older panels might improve the currency of the data. While OES is not a time series, there are many customers that would like to use it this way. Capturing data for some employers that report electronically every panel might facilitate the time series qualities of OES data. BLS is interested in testing ways to improve time series. Asking more multi-unit reporters to report all their data, rather than selected sample units is one way to do so.

Two categories included in the data dumps have already been proven useful in quality control. For example, hours worked data provided by some airlines helped to improve wage estimates for pilots and flight attendants. Wage rate data has shown the necessity to use wage rate data rather than intervals for the US Post office, where even nationwide, occupational wages are clustered. The job titles provided in the data dumps have helped to find job titles that are coded in the wrong occupations, or paid employees, such as students, who should not be in the scope of the OES survey. We would like to explore the possibility of asking selected employers to provide this data in their OES report to address any bias that may be the result of self selection to report this data.

5. **Statistical and Analytical Responsibility**

Mr. Dave Byun, Acting Chief, Statistical Methods Division of the Office of Employment and Unemployment Statistics, is responsible for the statistical aspects of the OES program. Mr. Byun can be

reached on 202-691-6359. Additionally, BLS seeks consultation with other outside experts on an as needed basis.

6. References

Bankier, Michael D. (1988). Power Allocations: Determining Sample Sizes for Subnational Areas. *American Statistician*, Vol. 42, pp. 174-177.

Bureau of Labor Statistics' Handbook of Methods, Chapter 3, Bureau of Labor Statistics, 2008 (<http://www.bls.gov/opub/hom/>).

Lawley, Ernest, Stetser, Marie, and Valaitis, Eduardas. (2007) *Alternative Allocation Designs for a Highly Stratified Establishment Survey*. 2007 Joint Statistical Meetings.

Response Analysis Survey Results for the Occupational Employment Statistics Survey, Bureau of Labor Statistics, Internal Document, 2013

Piccone, David and Stetser, Marie. (2009) *National Sample Reallocation for the Occupational Employment Statistics Survey*, 2009 Joint Statistical Meetings.

Technical Notes for May 2012 OES Estimates, Bureau of Labor Statistics, 2012 (http://www.bls.gov/oes/current/oes_tec.htm).