

**SUPPORTING STATEMENT FOR
REPORT ON THE OCCUPATIONAL EMPLOYMENT AND WAGE STATISTICS PROGRAM
OMB CONTROL NO. 1220-0042**

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

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 nearly ten million business establishments of which about 7.9 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. Data for Federal Government employees covered by the Unemployment Compensation for Federal Employees program (UCFE), the US Postal Service, and the Tennessee Valley Authority 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 OEWS 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	Agriculture
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 (OEWS designation)

Sample Size--The sample size is approximately 1.1 million establishments over a 3-year period. The sample is divided into six panels over three years with two semi-annual samples of about 187,000 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 OEWS survey:

Table 1: Universe and Sample Size Summary

Survey	NAICS Coverage	Responding Units	Sample Units	Universe Units
FY 2019 (November 2018 & May 2019 panels)	All in-scope NAICS	252,826	379,880	7,908,634
FY 2020 (November 2019 & May 2020 panels)		223,089	369,727	8,065,022
FY 2021 (November 2020 & May 2021 panels)		233,264	377,265	7,926,258
6-panel Totals		709,179	1,126,872	7,926,256

Stratification--Units on the sampling frame are stratified by State/Metropolitan Statistical Area (MSA) and Balance of State, and by three-, 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. Describe the procedures for the collection of information including:

- Statistical methodology for stratification and sample selection,
- Estimation procedure,
- Degree of accuracy needed for the purpose described in the justification,
- Unusual problems requiring specialized sampling procedures, and
- Any use of periodic (less frequent than annual) data collection cycles to reduce burden.

2(a) Sample Design

Allocation method--A variation of Neyman allocation procedure called a Power Allocation (Bankier, 1988)ⁱ is used to allocate the non-certainty sample to each State-/area/3-4-5-6-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/area/3-4-5-6 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, OEWS 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. Further discussion on the strengths and weaknesses of this approach are presented in Lawley et. al., 2007ⁱⁱ and Piccone, 2009ⁱⁱⁱ.

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/ three-, 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 187,000 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 OEWS sample allocation and selection procedures can be found in the 2021 OEWS Technical Note^{iv} and Survey Methods and Reliability Statement for May 2021^v.

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 1,122,000 establishments). Data from several panels are combined in order to reduce the sampling error of the estimates at detailed levels (MSA by 3-4-5-6 digit NAICS). Combining samples from six panels increases the sample counts.

Model-Based Estimation

The Model-Based Estimation using 3 years of data (MB3), a product of a long-term research project, was first used in official production for the 2021 OEWS estimates, which were published in March 2022. Testing indicates that the accuracy and reliability of the MB3 estimates improved over the former approach. In September 2019, BLS published 2016 data using the new estimation method as a research series, and a Monthly Labor Review article describing the method with comparisons to the old methods. Additional research series for 2017 through 2020 have since been published.

The MB3 method takes advantage of the fact that BLS observes key determinants of occupational staffing patterns and wages for all units in a target population. In particular, the QCEW provides data on the detailed industry, ownership status, geographic location, and size for every establishment whose workers are covered by state unemployment insurance laws. OEWS sample information is used to model wage distributions and industry/area/size/ownership/time wage adjustments. The estimation system includes redesigned components for model fitting, unit matching, and variance estimation. Further details of the method are presented in the Monthly Labor Review article, Model-based estimates for the Occupational Employment Statistics Program^{vi} and the Survey Methods and Reliability Statement for MB3 Research Estimates of the Occupational Employment and Wage Statistics Survey (MB3 Survey Methods Statement)^{vii}.

Estimates

Occupational employment and wage estimates are computed using observed data and predicted data for the population of about 9 million units. Predicted data are created for each unobserved unit of the population, so estimates are computed using full-population expressions.

Occupational Employment Estimates

Estimates of occupational employment totals are computed by summing all employment counts of a given occupation over the modeled population data. Estimates are made over area, industry, and ownership. For occupation o , where unit i is any establishment in cell c , the occupational employment estimate is:

$$\hat{X}_{o,c} = \sum_{i \in o,c} x_{i,o}$$

Hourly wage rate estimates

Mean hourly wage is calculated as the total hourly wages for an occupation divided by its total modeled population employment. Wage rate information is available for every individual federal employee and most state and private sector employees. Other wage data are in wage interval ranges and are converted to local hourly wages for each employee. These local hourly wages are predicted using adjusted estimates of local interval means, and then treated as point data. Mean wage is calculated as a sum of the hourly wage for each employee in a cell divided by the total number of employees in a cell. Employees E in a given occupation and wage interval at a single establishment will all have the same predicted wage w . For establishments i , wage ranges r , and occupation o in cell c , the computation is as follows:

$$\hat{w}_{c,o} = \frac{\sum_{i \in c,o} \sum_r x_{iro} \cdot w_{iro}}{\sum_{i \in c,o} x_{iro}}$$

Percentile wage rate estimates are computed directly from the predicted population using the empirical distribution function with averaging, which is implemented in many statistical packages.

Annual wage rate estimates

These estimates are calculated by multiplying mean or percentile hourly wage rate estimates by a “year-round, full time” figure of 2,080 hours (52 weeks x 40 hours) per year for most occupations. These estimates, however, may not represent mean annual pay should the workers work more or less than 2,080 hours per year.

Alternatively, some workers are paid based on an annual basis but do not work the usual 2,080 hours per year. For these workers, survey respondents report annual wages. Since the survey does not collect the actual number of hours worked, hourly wage rates cannot be derived from annual wage rates with any reasonable degree of confidence. Only annual wages are reported for some occupations.

More information on MB3 can be found in the [MB3 Survey Methods Statement](#).

2(c) Reliability

A probability based sample design is used to develop the OEWS 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. Variances for both mean wage estimates and occupational employment estimates are computed using the “bootstrap” replication technique. Background on the variance estimator used in OEWS can be found in the MB3 Survey Methods Statement.

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 OEWS 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. 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 1,122,000 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. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

3(a) Maximizing Response

A goal of the OEWS survey is that each State achieves an 80 percent response rate. The overall response rate for the 2021 survey was approximately 67 percent based on units.

Each State is responsible for collecting the occupational employment and wage data from their 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 OEWS 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 pre-notification letters to establishments before they are contacted to provide data.
- 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/three-, four-or-five-digit industry or MSA/three-digit industry cell by telephone.
- Contacting critical employers through personal visits (if necessary).
- Including information that explain the many uses of the OEWS data.
- Using status reports and control files to identify MSA/industry cells with low response rates.
- Stressing to respondents that assistance is available to help them complete the survey form.
- Providing a link on the survey instruments to OEWS data on the BLS Internet website to demonstrate program usefulness.
- Using a respondent web page that provides detailed information about responding to the OEWS 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.
- Coordinating collection for multi-unit establishments.
- Initial and follow-up email blasts to units that provided an email address.
- Conducting periodic Response Analysis or Nonresponse Analysis surveys to learn what motivates employers to respond.

3(b) Nonresponse Adjustment

Under MB3, there are three types of unobserved units, including nonrespondents, non-sampled units, and sampled units with an “unstable” response. An “imputation,” or prediction, procedure is used to impute for all unobserved units including nonresponse. A staffing pattern and wage distribution are imputed to all three types of unobserved units based on up to 10 “nearest neighbor” responses based on panel/State/MSA/employment size, and 6-digit industry cells. Within each cell, the 10 “nearest donors” (i.e., responding units) are selected for each nonrespondent. The nonrespondents’ sampling frame employment and the donors’ summed total employment are used within a cell to match donors with nonrespondents. Once donors and nonrespondents are matched, the occupational employment totals of the nonrespondent are computed using the proportional distribution of the donors’ occupational employment totals.

3(c) Non-Response Bias Research

Extensive research has been previously done to assess whether the non-respondents to the OEWS survey differ systematically in some important respect from the respondents of the survey and would thus bias

OEWS estimates. A study of comparison of average hourly occupational wages from OEWS 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 (Barkume, 2006)^{viii}. 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, resources permitting.

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 or wage estimate fails confidentiality screening, the estimate is suppressed.

3(e) Publishability

After confidentiality screening, the 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 direct match responses for the occupation, and the employment estimate must be 30 or greater for BLS publication. Records with employment estimates between 10 and 30 are released to states. For wage estimates, the relative standard errors must be 30 percent or less.

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information

The OEWS program regularly evaluates its processes and methods and occasionally updates its collection materials as a result of this ongoing research. Current and future plans are outlined below. Additional details regarding testing or improvements will be submitted through the non-substantive change request process.

Test of Email Timing and Format

OEWS uses a combination of emails and letters for non-response follow-ups. Currently, the mailings are sent out on fixed schedule, every four weeks. BLS would like to explore the effectiveness of varying this schedule, to see if sending the email closer to the letter mailing leads to improved response

rates or electronic reporting. The content of the emails would not vary from the OMB approved materials, but the frequency or gap between contacts would change depending on the treatment group. For states that volunteer to participate, we would randomly assign respondents to treatment groups. The groups would receive the emails on different schedules. Analysis of the final response rates for each group would be analyzed to identify the most effective schedule to follow in the future.

In conjunction with timing, email blast formats must be periodically evaluated to ensure maximization of character counts and other response factors. With evaluation and comments from BLS and state office staff, and respondents, OEWS may decide to alter the email blast templates accordingly. A test plan and proposed template will be submitted BLS through the non-substantive change process.

Adjusting solicitation materials

OEWS solicitation materials are designed to be dynamic. Solicitation materials are under constant evaluation to ensure the proper information is being conveyed to respondents, to ensure that respondents are properly assisted with submitting their data, and to ensure that the data being collected is that of the highest quality. To fulfill our promise of quality data and to control respondent burden, OEWS may need to make periodic changes to our solicitation materials. All changes will be subject to approval through the non-substantive process. OEWS will submit detailed test and change plans, which will include templates and design, to OMB when necessary.

Asking employers for additional data elements

BLS would like to ask employers to report data items that many already report without solicitation. Often employers include additional data items such as the number of hours the person is paid for. Many employers already provide many data elements in their electronic OEWS report that we do not ask for. These data elements include information that is requested by customers, but cannot be provided by OEWS 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. A small-scale test successfully collected extra data elements. These results showed that extra elements can be collected from respondents. While the test was limited in time and scope, the response rates mirrored those of regular OEWS data collection. Also, 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. BLS would like to continue this research.

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

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. 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 OEWS analyst sorts through the

reports, and matches them to the sampled units, saving the respondent the burden of doing so. The OEWS analysts ignore the unsolicited establishments. Some of the volunteered establishments might be included in the OEWS 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 OEWS 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 OEWS 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 OEWS survey. We would like to explore the possibility of asking selected employers to provide this data in their OEWS report to address any bias that may be the result of self-selection to report this data.

5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze person(s) who will actually collect and/or analyze the information for the agency.

Mr. Edwin Robison, Chief, Statistical Methods Division of the Office of Employment and Unemployment Statistics, is responsible for the statistical aspects of the OEWS program. Additionally, BLS seeks consultation with other outside experts on an as needed basis.

6. References

- ⁱ Bankier, Michael D. (1988). Power Allocations: Determining Sample Sizes for Subnational Areas. *American Statistician*, Vol. 42, pp. 174-177.
- ⁱⁱ Lawley, Ernest; Stetser, Marie; and Valaitis, Eduardas, (2007). Alternative Allocation Designs for a Highly Stratified Establishment Survey, 2007 Joint Statistical Meetings.
- ⁱⁱⁱ Piccone, David and Stetser, Marie, (2009). National Sample Reallocation for the Occupational Employment Statistics Survey, 2009 Joint Statistical Meetings.
- ^{iv} Technical Notes for May 2021 OES Estimates, Bureau of Labor Statistics, 2021 (http://www.bls.gov/oes/current/oes_tec.htm).
- ^v Survey Methods and Reliability Statement for May 2021 (https://www.bls.gov/oes/methods_21.pdf)
- ^{vi} Dey, Matthew; Piccone, David; and Miller, Stephen (August 2019). Model-based estimates for the Occupational Employment Statistics program. *Monthly Labor Review* (<https://www.bls.gov/oes/mb3-methods.pdf>).
- ^{vii} Survey Methods and Reliability Statement for MB3 Research Estimates of the Occupational Employment and Wage Statistics Survey (<https://www.bls.gov/oes/mb3-methods.pdf>).
- ^{viii} Barkume, Tony; Dey, Matthew; Ernst, Larry; Gittleman, Maury; and Polivka, Anne (November 2006). Comparing OES and NCS Wage Estimates. BLS internal report. (final report).