SUPPORTING STATEMENT FOR
QUARTERLY CENSUS OF EMPLOYMENT AND WAGES
BUSINESS SUPPLEMENT

OMB CONTROL NO. 1220-0198

B. COLLECTIONS 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.

1a. Universe

The Bureau of Labor Statistics (BLS) plans to conduct surveys under the Quarterly Census of Employment and Wages Business Supplement (QBS) on a sample of establishments included in the BLS Quarterly Census of Employment and Wages (QCEW). The universe of respondents to the QCEW are the 50 States, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. The primary source of data for these 53 entities are the Quarterly Contribution Reports (QCRs) submitted to State Workforce Agencies (SWAs) by employers subject to State Unemployment Insurance (UI) laws. The QCEW data, which are compiled for each calendar quarter, provide a comprehensive business name and address file with employment and wage information by industry, at the six-digit North American Industry Classification System (NAICS) level, and at the national, State, Metropolitan Statistical Area (MSA), and county levels for employers subject to State UI laws. Similar data for Federal Government employees covered by the Unemployment Compensation for Federal Employees program (UCFE) also are included.

The scope of the QCEW Business Supplement is:

- Establishments in-scope of the Annual Refiling Survey (ARS).
- Establishments with one to three employees that are excluded from the ARS.
- Establishments that are in low-change NAICS industries that are excluded from the ARS.

The QBS sample for the second survey under this collection, the 2022 Business Response Survey (BRS), will be restricted to those eligible for the 2023 ARS using the 7th and 8th digit of the EIN within the range of 67-99 from the QCEW. The QBS sampling frame of approximately 2.2 million establishments includes this 1/3 of the ARS, plus all other in-scope establishments within this EIN range.

1b. Sample size

For the 2022 BRS, BLS plans to select a sample of approximately 320,000 establishments. The objective of the large sample is to produce statistics at detailed levels including by size class, state, industry, and some state-industry, state-size combinations. Production of statistics at any detailed levels will depend on the ability to pass disclosure requirements to ensure confidentiality of the establishments responding to the survey.

	Sample	**Estimated Responses (25%)**
QBS Sample for the 2022 BRS	320,000	80,000

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.**

2a. Sample Design and Selection Procedures

The sample design and selection process involves the following steps:

➢ Identify the research goals
➢ Identify the QBS universe and relevant sub-populations (e.g. ARS-eligible establishments)
➢ Calculate sample sufficiency based on the research goals and universe definition
➢ Design sample to achieve sufficiency
➢ Select sample

Research Goals

The survey analysis breakouts initially identified by researchers:

➢ State by Industry Goods-Producing/Services-Producing Categorization
{53 states * 2 industry classifications =104 estimation cells}
➢ State by Broad Industry Size
{53 states * 4 broad industry size categories = 208 estimation cells}
➢ Modified NAICS Sector by Broad Industry Size Class
{16 modified NAICS sectors * 4 broad industry size categories = 64 estimation cells}
➢ Narrow Industry Size Class
{9 narrow industry size classes = 9 estimation cells}

In prior versions of the BRS, BLS has used the fifty-two "states" which include the District of Columbia and Puerto Rico but excluded the U.S. Virgin Islands (USVI). For the 2022 BRS, BLS will aim to add USVI to include all 53 "states" in QCEW.

Modified NAICS sectors and goods-producing/services-producing categories of interest are defined according to the following table:

Modified NAICS Sector	**Goods- or Services-Producing**	**Description**
11-21	G	Agriculture, Forestry, Fishing, and Hunting and Mining, Quarrying, and Oil
21	G	Mining, Quarrying, and Oil and Gas Extraction

22	G	Utilities
23	G	Construction
31-33	G	Manufacturing
42	S	Wholesale Trade
44-45	S	Retail Trade
48-49	S	Transportation and Warehousing
51	S	Information
52-53	S	Finance and Insurance and Real Estate and Rental and Leasing
54-56	S	Professional, Scientific, and Technical Services; Management of Companies and Enterprises; and Administrative and Support and Waste Management and Remediation Services
61	S	Educational Services
62 MOD	S	Health Care and Social Work (Excluding Services for the Elderly and Persons with Disabilities Establishments with less than two employees)
624120 MOD	S	Services for the Elderly and Persons with Disabilities Establishments with less than two employees
72	S	Accommodation and Food Services
81	S	Other Services (except Public Administration)

Note that NAICS sector 92 (Public Administration) establishments are excluded.

Industry sizes are first categorized into nine narrow size classes, then further grouped into four broader size classes, as shown below:

Narrow Industry Size Class	Broad Industry Size Class
1-4	
5-9	1-19
10-19	
20-49	
50-99	20-99
100-249	
250-499	100-499
500-999	
1000+	500+

QBS Universe and Sub-Populations

The QBS universe of inference will be based on QCEW 2021 Q4 establishments subject to the following constraints:

➢ Private sector only; not public administration
➢ Average monthly employment > 0 (average taken over the past year)
➢ NAICS 814110 (private households) and 491110 (postal service) excluded
➢ NAICS 999999 (NAICS unassigned/unidentified in QCEW)

➢ Establishments with NAICS 624120 (serv. for elderly and persons with disabilities) and average monthly employment < 2 excluded

For sample design purposes, the QBS universe will then be segmented into the following sub-populations:

➢ ARS-eligible; ARS 2023 (eligible to be selected for this QBS sample)
➢ ARS-eligible; ARS non-2023 (not eligible to be selected for this QBS sample)
➢ ARS-ineligible (eligible to be selected for this QBS sample)

Each year, the ARS is administered to roughly one-third of the ARS-eligible establishments on a rotating schedule. The ARS primarily uses the 7th-8th digits of the EIN to subset the QCEW universe into thirds for data collection purposes. Based on previous evaluation, the 7th-8th digits of EIN were determined to be a suitably random method of selection.

Samples will be selected for the ARS-eligible sub-population by drawing from only the ARS-eligible/ARS 2023 group. Samples will also be selected for the ARS-ineligible sub-population by drawing from the entire sub-population.

<u>Sample Sufficiency</u>

Sufficiency is determined for each of the four survey analysis breakouts listed in the Research Goals section. The four sets of sufficiency counts are then meshed together to create a single unified sample allocation design.

Sufficiency is determined based on estimating proportions to a certain degree of precision, where precision is based on researcher needs weighed versus survey burden and cost. The formula for the sample sufficiency of an estimation cell is based on the deconstruction of the formula for the variance of a proportion (using simple random sampling within the cell):

$$r_h = \frac{s_h^2 + p_h(1 - p_h)}{s_h^2 + \left(\frac{p_h(1 - p_h)}{N_h} \right)}$$

$$n_h = \frac{r_h}{est\_rr}$$

Where :

h is the stratum or analysis cell

r_h is the number of actual responders needed to estimate to the desired precision in the stratum or cell

est_rr is the estimated response rate (right now, a constant is used across all strata and cells)

n_h is the response rate adjusted sample size

N_h is the stratum h population

p_h is a guess at the eventual proportion estimate in stratum h (always set to 0.5)

s_h is the standard error chosen by the researcher to set the precision level for stratum h

Sample Design

The sample will follow a stratified design, with sample drawn from each state * modified NAICS sector * narrow industry size stratum with at least one establishment. Strata sample sizes will be allocated in an attempt to achieve sample sufficiency for all analysis cells across all four research goals. Strata estimates will be calculated within each stratum, provided the stratum has usable survey responses. Composite estimates will be constructed for analysis cells that comprise multiple strata. For example, survey estimates for Virginia establishments of broad industry size 20-99 (i.e. VA/20-99) will be built as a weighted composite of thirty-two strata estimates – one strata estimate for each of the sixteen modified NAICS sectors for VA/20-49 and one strata estimate for each of the sixteen modified NAICS sectors for VA/50-99.

Sample Selection

Most strata can be categorized into a single universe sub-population. Within these strata, samples will be selected using simple random sampling.

Some strata will consist of establishments from multiple universe sub-populations. Each of these strata will be segmented into ARS-eligible and ARS-ineligible components. The strata sample size will be proportionately allocated to these components. For any such stratum, a simple random sample will be drawn from within the ARS-ineligible component, using the corresponding proportionately allocated sample size. For the stratum's ARS-eligible component, a simple random sample will be drawn from only the 2023 ARS sub-component, using the corresponding proportionately allocated sample size. For each of these strata, the 2023 ARS sub-component is considered to be a random first stage of selection of ARS-eligible establishments. Although establishments from the non-2023 ARS sub-component are not allowed to be surveyed, they will be accounted for in weighting.

For establishments that are in-scope for the 2022 BRS but out-of-scope for the ARS (establishments with one to three employees; establishments that are in low-change NAICS industries), BLS is also researching the possibility of relying predominantly on UI accounts that have an email address available for email blasts. This would reduce the cost of printing and mailing. This research is ongoing.

2b. Estimation Procedure

The primary measure of interest will be an estimated proportion possessing an attribute being assessed by a survey question. Proportions will be estimated for each of the analysis breakout cells described in section 2a. The proportion estimate formula is:

$$\hat{p} = \sum_{i=1}^{H} W_h \hat{p}_h$$

Where:

$\hat{p}$ is the est. weighted sample prop. (for a particular survey question) for a breakout cell

$\hat{p}_h$ is the est. sample prop. (for a particular survey question) for stratum h within the breakout cell

W_h is the stratum weight within the breakout cell

H is the number of strata in the breakout cell

Each stratum weight is the population proportion of each stratum relative to the composite population of interest:

$$W_h = N_h / N$$

Where:

N_h is the population of stratum h within the analysis breakout cell

N is the population of the entire analysis breakout cell

The formula for the estimated sample proportion for some stratum (h), generalized for non-response adjustment, is:

$$\hat{p}_h = w_h^{resp} \hat{p}_h^{resp} + w_h^{nonresp} \hat{p}_h^{nonresp}$$

Where:

$\hat{p}_h$ is the non-response-adjusted estimated proportion possessing the attribute of interest in stratum h

$\hat{p}_h^{resp}$ is the estimated proportion of responders possessing the attribute of interest in stratum h

$\hat{p}_h^{nonresp}$ is the estimated proportion of non-responders possessing the attribute of interest in stratum h

w_h^{resp} is the proportion of the stratum h sample that responded to the survey question

$w_h^{nonresp}$ is the proportion of the stratum h sample that did not respond to the survey question

The proportion of non-responders possessing the attribute of interest is generally unknowable. Therefore, the assumption is made that, within each stratum, responders and non-responders possess the attribute of interest in the same proportion, and therefore the formula reduces as follows:

$$\hat{p}_h = \hat{p}_h^{resp}$$

At a later date, survey sponsors may decide to request composite estimates for new analysis breakouts that were not determined ahead of time. The definitions above can be generalized to composite estimators over various strata groupings by redefining the universe to the one of interest. However, the sample was not designed with the intention of being sufficient for these new analyses.

2c. Reliability

Variance estimation will involve (i) the application of the general formula for the variance of a composite proportion estimator drawn from a stratified random sample and (ii) the application of the basic formula for the variance of a proportion drawn from a simple random sample. Specifically, the variance of the proportion estimator for some particular analysis cell is:

$$Var(\hat{p}) = \sum_{h=1}^{H} W_h^2 Var(\hat{p}_h), \text{ where } W_h = \frac{N_h}{N}.$$

Under the assumption that, within each particular stratum, non-responders possess the attribute of interest in the same proportion as responders, the formula for the within stratum variance of a proportion calculated from a simple random sample is:

$$Var(\hat{p}_h) = \left(1 - \frac{r_h}{N_h} \right) \left(\frac{\hat{p}_h (1 - \hat{p}_h)}{r_h - 1} \right)$$

In the formula above, note that r_h is the number of establishments in stratum h that *responded* to the survey question of interest. It is *not* the stratum h sample size.

The formulas above can be tailored to the desired composite estimator by applying it across only the set of strata that are relevant to that particular composite.

2d. Data Collection Cycles

Since the QBS will be collected directly after respondents submit the ARS, the data collection cycle for the QBS will follow the established pattern of collection used for the Annual Refiling Survey (ARS) as outlined in materials for OMB Control No. 1220-0032.

Overall data collection is expected to be conducted over an eight to sixteen week period starting in late July 2022. This will give time for response rate review and analysis to determine the optimal data collection outreach methods to maximize response rates and reduce burden. BLS will rely on review and analysis tools developed for the ARS and QBS to assist in decision making.

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.

BLS expects a response rate of 25%. This expectation is based on responses to prior test estimates of the QBS pilot tests and results observed from the BRS in 2020 and 2021.

3a. Maximize Response Rates

To maximize response rates, all units selected in the sample will be transitioned to the QBS questions once they complete the ARS. All multi-units and any singles that are out of scope for the ARS, but are selected for the QBS, will be contacted using established ARS contact methods, i.e. email and ARS solicitation letters. Non response follow-up will be conducted per the current ARS procedures as outlined in materials for OMB Control No. 1220-0032.

BLS will rely on established, tested data collection processes to ensure maximum response rates to the QBS. BLS will make use of existing processes from the ARS, BRS, and QBS pilot tests, along with consulting other established BLS survey programs to maximize efficiency and reduce burden.

All of the data collection will take place online via the BLS Internet Data Collection Facility (IDCF). This method of fully online data collection has been successfully employed and tested with the ARS, BRS and QBS pilots and has been effective in reducing cost to the government, reducing respondent burden and maximizing response rates[1][2][3].

3b. Non-Response Adjustment

BLS expects a response rate of 25 percent. However, this response rate will not be uniform across all strata. Additionally, some strata are smaller than others. Consequently, there will be strata that end up with no usable survey responses. When building composite estimates that consist of "empty" strata, the empty strata will be imputed.

Because BLS expects a response rate of 25 percent, BLS will be performing a nonresponse bias analysis and determining appropriate nonresponse bias adjustments.

For strata imputation, survey strata and question combinations with no usable item response will have their establishment proportions and variances imputed according to an ordered hierarchy of related composite estimates that will be determined at a later date.

---

[1] Stang and Thomas "Web Collection in the Quarterly Census of Employment and Wages Program", ICES-V, 2016. http://ww2.amstat.org/meetings/ices/2016/proceedings/072_ices15Final00299.pdf

[2] Stang and Thomas "Email Solicitation for a Business Establishment Survey – Results from the 2015 Annual Refiling Survey", JSM 2016. http://9004e5e16f4a25df17a0-290e28d0a6d5d71f78b4f59d5f323756.r86.cf1.rackcdn.com/ASA-JSM/pdf/389517.pdf

[3] Stang and Thomas "Developing and Testing the Business Research Survey," JSM 2018. https://ww2.amstat.org/meetings/jsm/2018/onlineprogram/AbstractDetails.cfm?abstractid=328621

Other non-response bias adjustment methods may be assessed based on survey results. For example, another method to explore is to impute missing questionnaire responses using the response of the nearest responding neighbor. Final methods will be documented along with all of the other statistical design methods on the public webpage used for dissemination of the results.

3c. Confidentiality

Before 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 estimate fails a predetermined primary confidentiality threshold, such as the p% rule, then the cell can be protected. Whether this protection is suppression, rounding or other method is somewhat dependent on the objective of the survey. Secondary confidentiality protection is also implemented to protect respondent information at this level.

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 QBS is built on years of testing for the QBS and the ARS. This testing has informed the platform used, contact strategies, data processing methods, and expected data review time frames.

The content proposed for inclusion on the 2022 BRS has been reviewed by the BLS cognitive experts and currently undergoing by the Office of Survey Methods Research staff at BLS. The final content of the questionnaire may change slightly in response to the testing.

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. Ed Robison, Division Chief of the Statistical Methods Staff, Office of Employment and Unemployment Statistics, is responsible for the statistical aspects of this survey.