

## **H: Small Business Pulse Survey Methodology**

### **Target Population**

The target population is all nonfarm, single-location employer businesses with less than 500 employees and receipts of \$1,000 or more in the 50 states, District of Columbia, and Puerto Rico. The sampling frame was extracted from the Business Register in April 2020. The following industries were designated as out of scope for the Business Pulse Survey:

- Agriculture production (NAICS in ('110000', '111', '112')),
- Railroads (NAICS = '482'),
- U.S. Postal Service (NAICS = '491'),
- Monetary Authorities – Central Bank (NAICS = '521'),
- Funds, Trusts, and Other Financial Vehicles (NAICS = '525'),
- Religious grant operations and Religious organizations (NAICS = '813'),
- Private households (NAICS = '814'),
- Public administration (NAICS = '92'), or
- Unclassified with legal form of organization as tax-exempt or unknown

The set of businesses in the target population that responded to the 2017 Economic Census were identified as the subset of businesses eligible to participate in the Small Business Pulse Survey (SBPS). The 2017 Economic Census (EC) utilized an all-electronic data collection strategy. Respondents to the EC were mailed a letter containing an authentication code and were invited to create an account using the Respondent Portal. To establish their Respondent Portal account, respondents provided a valid email address, their name, phone number, and were required to establish a password.

Of the 8 million employer establishments in scope to the EC, about 2 million were classified as multi-unit and 6 million are classified as single units. Of the 6 million single unit employer businesses, almost 5 million have paid employees between 1 and 500. Approximately 1.7 million single unit businesses received an invitation to respond to the EC and administrative data was used for the remaining cases to minimize respondent burden. Out of the 1.7 million, 1.1 million had employees between 1 and 500.

In total, 4 million establishments received an invitation to the 2017 EC and the associated response rate for single unit businesses was 70% and for multi-unit businesses was 79%. Nearly 1.2 million responses including authenticated email address from the single unit universe were captured as part of the EC data collection process. Roughly 550K units of the units that were

active at the time of the EC would now be considered deaths due to having no payroll in 2018, 2019, or 2020; these were excluded from the SPBS frame.

In April 2020, the Census Bureau extracted from the Business Register single-location businesses with e-mail addresses and with employment between 1 and 500 and payroll in 2018, 2019, or 2020. This resulted in approximately 950K units. We then matched these single location businesses to our current survey databases to determine if we had a more recent email address. Using date time stamps associated with update to the email address field, we found 91K email addresses that were more recent than what was collected in the EC. An additional 10K emails were removed due to validity checks, resulting in a target population of about 940K.

An initial comparison of firm age, as calculated for the Business Dynamics Series (BDS), in the set of businesses in the target population vs. those with e-mail addresses, demonstrates a slightly smaller percentage of younger firms (0-5 years) and a slightly larger percentage of older firms (21+ years) in the e-mail population than in the population overall. These differences might be due in part to different scoping criteria in BDS, which excludes businesses that do not have positive March 12 employment.

Firm Age	BDS Small SUs	E-mail Sample
0-5 years	33.8	25.1
6-10 years	16.0	15.7
11-20 years	22.3	25.4
21+ years	27.9	33.8

We plan to use firm age, where available, to monitor and compare distributions of responding firms to the target population to assess potential nonresponse bias in the estimates.

### Sample Design

To maximize the use of available e-mail addresses, the Census Bureau intends to use the full set of 940K businesses with e-mail addresses in the group. The full set will be divided into nine panels of roughly equal size (104,500 businesses) and representation for the weekly e-mail invitations to respond to the pulse survey. To establish the panels, the entire sampling frame (including businesses with and without e-mail addresses) was stratified by state and by 3-digit NAICS. While we intend to publish by sector/2-digit NAICS, the 3-digit NAICS stratification ensures that businesses in the 3-digit industries are distributed across the nine panels. Within each sampling stratum, the units were sorted by an indicator for presence of an e-mail address, MSA, and 2018 annual payroll. Each business in the stratum was then systematically assigned to one of the nine panels. The sample weight (*wgt*) for each unit is the ratio of the count of the total number units in the stratum (both with and without email addresses) divided by the number of selected sample units (with e-mail addresses).

## Tabulation

Weekly visualizations will be based on weighed percentages for response categories. Non-response adjustments may be applied to the current weekly panel as well as to late-responding units included in the current period's estimates.

Within any tabulation domain, we define the following:

- $R$  is the number of respondents in the current weekly panel
- $NR$  is the non respondents in the currently weekly panel
- $wgt$  is the sample weight for the current weekly panel
- $R'$  is the late respondents from the previous weekly panels included in the current estimate
- $wgt'$  is the sample weight for the previous weekly panels

Initial on-response adjustments are calculated for each state by 3-digit NAICS.

$$NAF_h = \frac{\sum_{R+NR} wgt}{\sum_R wgt + \sum_{R'} wgt'}$$

The non-response weight adjustment factor is applied to the current weekly panel and the late response units. The tabulation weight is calculated as follows:

$$WGT_{tab} = \begin{cases} NAF_h \times wgt, & \text{for responses} \in \text{current week panel} \\ NAF_h \times wgt', & \text{for late responses} \end{cases} \quad \text{previous weeks } \downarrow$$

For a specific question and response category  $a$ , the published response percentage is calculated as:

$$P_a = \frac{\sum_{R, R \in a} WGT_{tab}}{\sum_{R+R'} WGT_{tab}} * 100$$

## Quality Measures

The Census Bureau will be transparent about the survey and its limitations. We will publish quality metrics including response rates, nonresponse bias indicators, and estimates of variability so that users may assess the quality and suitability of the results for their purposes.

### - Calculation of Response Rate

The unit response rate is calculated as follows:

$$URR = \frac{\text{number} \in R + \text{number} \in R'}{\text{number} \in R + \text{number} \in NR} * 100$$

In this calculation, we are assuming that the late respondents from previous week's panel are similar to the non-respondents in this week's panel who will be delayed respondents for the following week. Thus we do not adjust the denominator for the late responders; but add them to the numerator.

Response rates and collected data will be monitored internally on a daily basis at the national, state, and MSA level.

- **Estimates of Sampling Variability**

Variance estimates will be calculated using a delete-a-group jackknife estimator with 10 replicates. A description of the method can be found in the paper [Using the Delete a Group Jackknife Estimator in NASS Surveys](#), by Phil Kott. This method is used in a number of Census Bureau surveys of businesses, including the Annual Business Survey and the Annual Capital Expenditures Survey. We will be adapting code used in the Annual Business Survey to do the calculations.

We will provide standard errors with each survey and notes about data limitations. Tables of estimates will include notes that flag low quality estimates with text similar to the following:

*The standard error of this estimate yields a coefficient of variation greater than 30 percent, which is an indicator of potential quality issues. Caution should be used when interpreting this estimate.*

- **Indicators of Nonresponse Bias**

Census Bureau business surveys generally account for nonresponse through imputation or non-response adjustment. Following OMB standards, we conduct a non-response bias study for each program with an expected unit response rate below 80%. Based on the study results, we attempt to mitigate the nonresponse bias where possible. For example, in ongoing surveys, extra follow-up efforts for smaller units or particular industries may be implemented to improve response when bias is indicated in estimates for these domains.

The Census Bureau intends to monitor the potential non-response bias as part of the on-going collection of the Small Business Pulse Survey. To assess potential nonresponse bias, we will compare distributions of characteristics, such as employment, annual payroll, and firm age, of survey respondents to similar distributions of non-respondents and the full in-scope business

population from the Business Register. The Census Bureau will make these comparisons at all published levels, including the national, state, and published metropolitan area levels as well as by NAICS sector. To address bias, a post-stratification adjustment by employment size class will be applied at the state by 2-digit NAICS level. Note that the size adjustments, developed from Business Register data, will not yet reflect the full impact of the pandemic on businesses, such as business closures. Data users will need to take into consideration the potential effects of the pandemic on particular geographic locations and industries in assessing potential nonresponse bias.

We expect to release quality metrics, including information on potential nonresponse bias, with each data release. The release of Week 1 estimates will be held until all planned quality metric calculations are programmed, tested, and ready for release.

### **Disclosure Avoidance**

Business Pulse Survey releases will comply with Census Bureau disclosure avoidance practices. U.S.-level estimates will be suppressed if based on fewer than three contributors. In compliance with IRS Publication 1075, Tax Information Security Guidelines for Federal, State, and Local Agencies, state-level and MSA-level estimates with fewer than 10 contributors will be suppressed.