

Department of Commerce
United States Census Bureau
OMB Information Collection Request
Annual Survey of Manufactures (ASM)
OMB Control No. 0607-0449

B. Collection of Information Employing Statistical Methods

1. Universe and Respondent Selection

The 2018 ASM statistics will be based on a sample that was selected in 2014 and supplemented annually with manufacturing births. The frame, used for selecting the 2014 sample, was assembled from the 2012 Economic Census – Manufacturing. The frame consisted of 294,600 active manufacturing establishments.

In order to reduce the reporting burden on small- and medium-sized single-location companies, the frame was partitioned into two groups: establishments eligible to be mailed a questionnaire (101,250 establishments) and establishments not eligible to be mailed a questionnaire (193,350 establishments). The group of establishments that is not eligible to be mailed a questionnaire still contributes to the ASM estimates. The group of establishments that is eligible to be mailed a questionnaire is defined as the mail stratum. It is comprised of larger single-location manufacturing companies and all manufacturing establishments of multi-location companies. Of the 101,250 establishments in the mail stratum, 47,800 establishments were selected for the ASM sample using methodology similar to what was used for previous ASM samples. The initial sample was supplemented with manufacturing establishments that were newly opened in 2013 (births) to yield a sample of 50,200 establishments for the 2014 ASM. Births added to the mail stratum are large, single-location companies and new manufacturing establishments of multi-location companies. Since births are added annually to the mail sample, the 2018 ASM sample size is expected to be approximately 55,000 establishments.

The group of establishments that is not eligible to be mailed a questionnaire is defined as the nonmail stratum. The nonmail stratum contained the remaining 193,350 single-location companies. Although this group still contributes to the ASM estimates, no data are collected from companies in the nonmail stratum. Rather, data are imputed using administrative records of the Internal Revenue Service (IRS), the Social Security Administration (SSA), and the Bureau of Labor Statistics (BLS) or are imputed based on industry averages. This administrative information, which includes payroll, total employment, industry classification, and physical location, is obtained under conditions which safeguard the confidentiality of both tax and census records. Although the nonmail

companies account for nearly two-thirds of the establishments in the universe, they account for only about 6 percent of the manufacturing output. The nonmail stratum is supplemented annually with small manufacturing births that are not included in the mail stratum.

A new sample will be drawn for the 2019 ASM using methodology similar to the 2014 methodology. The size of the 2019 ASM mailed sample is expected to return to the 50,000 range.

Two types of response rates are computed for the ASM: unit response rate (URR) and total quantity response rate (TQRR). The URR is the percentage of reporting units, based on unweighted counts, that were eligible (E) or of unknown eligibility (U) and were respondents (R) in the statistical period. Cases are assumed to be active and in scope in the absence of evidence otherwise. This includes cases that are Undeliverable as Addressed. To be considered a respondent to the ASM, a reporting unit must provide both of the key items: value of shipments and total payroll. The formula for calculating the URR is as follows: $URR = [R/(E+U)] * 100$. The URR for the 2016 ASM was 63%. This rate is lower than in previous years, primarily due to a shortened collection period.

The TQRR is defined as the percentage of the estimated (weighted) item total that is obtained from directly reported data or from sources determined to be equivalent quality to reported data. The 2016 TQRR was 64% for value of shipments, and 76% for total payroll.

2. Procedures for Collecting Information

a. Description of Reporting Questionnaires

The Census Bureau will send an initial contact letter to approximately 55,000 manufacturing establishments. The contact letter will direct respondents to report the MA-10000 online.

b. Matching to BLS Establishment List

Incomplete industry codes can cause potentially serious errors in our coverage of new producers and in our ability to perform accurate editing and imputation. Although the SSA requests industry classification in the original application for the EI number, frequently only a 3- or 4-digit code can be assigned for a large number of the new businesses. This amounts to about 20,000 manufacturing establishments annually. In addition, a large number of newly active establishments are received from the IRS without industry classification. Classification information is requested from the BLS for both types of cases.

c. Sampling Methodology

The 2019 ASM sample will be selected from the Manufacturing Sector of the 2017 Economic Census – Manufacturing using methodology similar to the 2014 methodology. The sample will be supplemented annually to include new establishments in the Manufacturing Sector. This sample will be used through reference year 2023. Below is an overview of the sample design.

The universe will again be partitioned into mail and nonmail strata. Within each of the 360 NAICS industries, small- and medium-sized single-location companies will be identified and defined as the nonmail component. Establishments comprising the remaining portion, including all establishments of multi-location companies, will be defined as the sample frame.

On the sample frame, establishments that meet specified criteria will be selected in the sample with certainty.

The sampling strategy will be to select an independent sample within each of the 360 NAICS industries. This will allow optimization of the probabilities of selection within each industry, which will improve the representativeness and reliability of the survey estimates. Within each industry, each establishment will initially be assigned multiple probabilities. These probabilities will be based on the establishment's relative importance within the industry that it is classified and the set of product classes that it produces; and, the target reliability constraints defined by the survey manager. For example, an establishment that has activity in three product classes will initially be assigned a total of four probabilities (one would be industry-based and the other three would be product class based). For sample selection purposes, the establishment's maximum probability will be used. The use of the maximum probability ensures that target reliability constraints will be satisfied.

d. Estimating Procedures

A primary objective of the ASM is to estimate year-to-year change between the censuses. The variances of estimated changes are always reduced when the sample overlap is high between both periods and the year-to-year correlations are positive. Since the ASM sample is selected and maintained for a period of five years and the year-to-year correlations are high for most ASM data variables, an estimator that takes advantage of both the constancy of sample and the positive correlations is highly desirable. For the ASM this is achieved via the use of the "difference

estimator.” Essentially, an estimate of the “difference” between the current year and the census year is derived from the sample and added to or subtracted from the corresponding census value.

For a given sample size, the difference estimator generates more reliable estimates than most estimators of totals. It also offers the attractive feature that estimates for different subgroups are additive, ensuring that estimates are arithmetically consistent.

The formula for the difference estimator is as follows:

$$Y''_{cy} = Y'_{cy} + (Y_{cen} - Y'_{cen}) + I_{cy}$$

Where Y''_{cy} is the published estimate for the current year. Y'_{cy} is the linear estimate obtained by multiplying each mail sample establishment's current year data by the corresponding establishment weight.

I_{cy} is the estimate obtained from the use of administrative records and industry averages for establishments in the nonmail portion of the universe.

Y_{cen} is the census value from the sampling frame.

Y'_{cen} is the linear estimate of Y_{cen} from the sample selected from the sampling frame.

For selected variables with poor year-to-year correlations, estimates of total are generated as follows:

$$Y''_{cy} = Y'_{cy} + I_{cy}$$

3. **Methods to Maximize Response**

a. Follow-up Procedures

The contact strategies will include an initial mailing, due date reminder as the due date approaches and systematic mail follow-up for nonresponse (Attachment A), supplemented by telephone follow-up for selected firms. We call larger delinquent companies at the time of our processing closeout prior to the tabulation review stage. In addition, the analyst staff contacts individual establishments of these larger companies as part of the tabulation review stage.

b. Estimating for Missing Data

The procedures for handling missing data essentially are the same as the prior years. For single-establishment companies that do not respond, we obtain employment and payroll data from the IRS. We then estimate the other data items for the nonrespondent, using a combination of the prior year establishment operational relationship and industry averages.

For establishments of multiunit companies that do not respond, we obtain operational status information from the Company Organization Survey to identify the establishments that actually are in business and, therefore, candidates for imputation. We then estimate the detailed items for the nonrespondent establishments using a combination of the prior year establishment operational relationships, industry averages, and changes in industry levels developed from data supplied by the Census Bureau's Manufacturers' Shipments, Inventories, and Orders Survey (M3) survey (inventories) and the Bureau of Labor Statistics (BLS) (employment and payroll).

c. Reliability

The estimates developed from the sample are likely to differ from the results of a complete canvassing of all eligible establishments in the population. The particular sample selected for the ASM is one of many probability samples that could have been selected under identical circumstances. Each of the samples would yield a slightly different set of results. These differences are known as sampling errors or standard errors of the estimates. Estimates of the magnitude of these sampling errors are included, in the publications, in the form of relative standard errors (the standard error divided by the corresponding estimate). At the total manufacturing level, the relative the standard errors for the key data items (value of shipments, total payroll) were 0.1 percent for 2016. At the 3-digit subsector levels, the relative standard errors for the key items were all less than 1.7 percent for 2016.

4. Tests of Procedures or Methods

a. Cognitive Testing of New Questions

From October 2017 through May of 2018, researchers conducted three rounds of cognitive testing, totaling 36 interviews, for new questions for the Annual Survey of Manufactures that focused on how companies are investing in robotics (see OMB Generic Clearance submitted 9/27/2017, approved 10/3/2017). The final version of the questions appear as a Special Inquiry in Item 28 of the ASM, and ask companies to report, for each establishment:

- the capital expenditures on industrial robotic equipment in the reporting year,
- the number of industrial robots in operation in the reporting year, and
- the number of industrial robots purchased in the reporting year.

The questions focus on “industrial robotic equipment,” defined as “automatically controlled, reprogrammable, and multipurpose machines used in industrial automated operations.”

Results from pretesting suggest that response burden for these new questions will correlate roughly with the size of the company:

- Smaller companies using robotic equipment likely have a small amount of equipment; thus requested data tend to be easier to identify in accounting records, and respondents are more likely to have personal knowledge of the robotic equipment or easy access to someone who can provide this information.
- Survey respondents for large companies, particularly those with many manufacturing locations, are more likely to be unfamiliar with their company’s manufacturing process, and are less able to identify expenditures for robotic equipment in their records without assistance. Additionally, obtaining the requested data would likely require contact with individual plant managers, adding burden to the response task.
- Some participants from larger companies reported that providing accurate information for these questions would be extremely difficult, if not impossible.

These and other results suggest that data quality may be adversely impacted, particularly among the largest multi-location manufacturing companies. Because of this potential difficulty, respondents who are unable or unwilling to provide these data are asked to provide reasons in an associated open question.

In addition, because we are familiar with only a few of the industries using robotics, these questions are being included for all units in our sample in order to discern the pervasiveness of the technology, possibly enabling future targeting.

Due to concerns remaining after completion of cognitive testing, plans are in place to conduct respondent debriefings during data collection, as responses are received, in order to evaluate the accuracy and burden across a variety of manufacturing industries and company sizes.

Particular, although not exclusive, attention will be on companies opting to provide reasons for item nonresponse. The goal is to schedule 60-70 interviews, lasting up to 1 hour each. In order to reach this goal, we expect to make approximately 210 calls in order to identify and obtain cooperation from appropriate respondents. Recruiting calls with respondents that agree to meet will last 5 minutes with refusals or nonresponses lasting 2 minutes. As such, the estimated maximum recruiting burden is expected to be 10.5 hours (70 screening calls x 5 minutes = 350 minutes; 140 refusals/nonresponses x 2 minutes = 280 minutes; 350 minutes + 280 minutes = 630 minutes = 10.5 hours). Thus the total public reporting burden for this research is approximately 80.5 hours (a maximum of 70 interviews x 1 hour/interview = 70 hours; 70 hours for interviews + 10.5 hours for recruiting = 80.5 hours).

Findings from respondent debriefings will be used to further refine and update these questions, and resulting revisions are expected to require subsequent cognitive testing.

b. Instrument Evaluation Based on Analysis of Paradata

Paradata are defined as data collected as a byproduct of the survey data collection process. For an internet based instrument, examples of paradata include browser information, time in survey, and patterns of movement through instrument (e.g., backing up, changing answers, clicking buttons, using help links).

Paradata from the 2016 ASM Centurion web instrument were analyzed in order to identify problematic screens and better inform design decisions. The ASM paradata was comingled with paradata from the 2016 Company Organization Survey (COS) for the entire analysis, as these two surveys share the Centurion web instrument and are conducted during the same time frame every year. Key research questions and findings from the 2016 ASM/COS paradata analysis are presented in the table below:

Research question	Key results
Help link usage: A high number of clicks on “Help” on a screen may indicate a design issue.	The two most frequently clicked Help links were to obtain: 1. Census Bureau contact information, most commonly accessed from the Main Dashboard screen and the Message screen. 2. Additional information about the survey, most commonly accessed from the screen displaying the Overview/Reporting Steps.
Error messages: Screens where many	The screens triggering error messages for the largest numbers of respondents were:

respondents triggered an error may have a design issue.	<ol style="list-style-type: none"> 1. The screen containing establishment location information and requesting operational status. 2. The screen containing the group of questions about employment and payroll.
Number of clicks on specific buttons indicates respondents' -- Navigation through the instrument Use of particular features	<p>The buttons most frequently clicked by respondents, overall, were:</p> <ol style="list-style-type: none"> 1. "Go to Overview" 2. "Go to Step 1 – Report" 3. "Go to Step 2 – Review" 4. "Fix" clicked from the Review screen

Several of the research questions identified for paradata analysis were based on usability testing that was conducted from July, 2016, through October, 2016, for the development of the 2016 ASM electronic instrument. For example, the 2016 usability testing suggested that the top banner navigation buttons, "Overview," "Step 1 – Report," and "Step 2 – Review," – were under-utilized by respondents, because they did not bear the expected design for a clickable button. This design was corrected in the production instrument, and their prominent use is indicated in results from the paradata analysis shown above. While the 2016 usability testing and lessons learned from fielding the 2016 ASM were used to improve the 2017 Economic Census instrument, this design is being carried over for use in the 2018 ASM.

The Economic Programs Directorate plans to develop a routine annual evaluation of the Centurion instrument using paradata, in order to support continuous improvement to facilitate the response process and reduce respondent burden. Future paradata analysis that leads to changes in the instrument may require additional usability testing.

c. Evaluations Supporting Continuous Quality Improvements

The Census Bureau's Economic Programs Directorate plans to implement an ongoing program of research to evaluate the performance and effectiveness of existing ASM questions, data collection instruments, and communication materials. These evaluations are intended to support timely continuous quality improvements in data collection methodology. The primary methodologies for conducting these evaluations is expected to include real-time and/or post-collection respondent debriefings, along with paradata analysis, as appropriate. To ensure timely implementation of these evaluation studies during data collection, up to 100 hours of total

public reporting burden, supporting 80-90 debriefing interviews, including recruiting, are required annually.

The Economic Programs Directorate will continue to submit requests under the Census Bureau's generic clearance for pretesting for survey questions covering new content, not previously collected on the ASM, and new or substantial changes in instrumentation and/or technology for data collection.

5. Contacts for Statistical Aspects and Data Collection

Mr. Julius Smith, Jr., Assistant Division Chief for Manufacturing, Mining, and Construction Sectors of the Economy-Wide Statistics Division, serves as consultant on the collection, analysis, and the dissemination of the ASM data. He can be reached on (301) 763-7662.

Ms. Amy Newman-Smith, Methodology Director for Manufacturing, Investment, and Construction Programs of the Economic Statistical Methods Division, serves as consultant on the statistical aspects of the ASM. She can be reached on (301) 763-6595.

Attachments:

- A. Draft Letters
- B. Draft Definitions and Instructions for the Annual Survey of Manufactures, MA-10000
- C. Former MA-10000(S) Questionnaire
- D. Draft Questionnaire Paths
- E. Electronic Instrument Selected Screen Shots
- F. List of Contacts
- G. BEA Letter of Support
- H. Robotics Use
- I. Eliminated Questions