Evaluation of a Sample Design Based on Predicted Occupational Frame Data

Erin McNulty and Alice Yu
Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Washington, DC 20212

Abstract
The Occupational Requirements Survey (ORS) is an establishment survey conducted by the Bureau of Labor Statistics (BLS) for the Social Security Administration (SSA). The survey collects information on the vocational preparation and the cognitive and physical requirements of occupations in the U.S. economy, as well as the environmental conditions in which those occupations are performed. For the first three years of the survey, the establishment sample was allocated proportional to industry employment, and occupations were subsampled within each establishment in proportion to occupational employment. In an effort to publish data on a wider variety of occupations, the ORS sample design was modified using a predicted frame of occupations. Establishments with less-common occupations were selected with greater frequency, and occupations for each establishment were pre-selected with an emphasis on these less-common occupations. This paper describes the sample design and evaluates its effectiveness in its first year of collection.

Key Words: complex sample design, predicted frame, establishment survey, stratification, allocation, hard-to-reach population

1. Introduction

For the first three years of the Occupational Requirements Survey (ORS), the establishment sample was allocated proportional to industry employment, and occupations were subsampled within each establishment in proportion to occupational employment. This resulted in estimates for occupations that represent over 90% of the nation's employment. However, fewer than half of occupations had any estimates because employment is not spread evenly among occupations. In an effort to publish data on a wider variety of occupations, the ORS sample design was modified using a predicted frame of occupations. Establishments with less-common occupations were selected with greater frequency, and occupations for each establishment were pre-selected with an emphasis on the less-common occupations.

The purpose of this paper is to describe the ORS sample design and evaluate its effectiveness in its first year of collection. The next section gives a brief background of the survey, followed by a description of the new survey design. Then, the results of collected are presented and discussed.

2. Background Information on ORS

In the summer of 2012, the Social Security Administration (SSA) and the Bureau of Labor Statistics (BLS) signed an interagency agreement to begin the process of testing the collection of data on occupations. As a result, the ORS was established as a test survey in late 2012. Initial planning for ORS involved several feasibility tests throughout fiscal years (FY) 2013 and 2014. These tests examined the feasibility of gathering the basic information desired and the availability of data, the efficiency of alternative collection procedures, and the probable degree of cooperation from respondents. In FY 2015, the Pre-Production Test was conducted to mimic what will occur during ORS production. The first wave of a three

year production cycle started in FY 2016 and ended in FY 2018. For more details on the development of the ORS, see the papers by Ferguson, McNulty, and Ponikowski (2014); Ferguson and McNulty (2015); and Rhein, Ponikowski, and McNulty (2013). For more details on ORS production processes and outputs to this point, see the BLS Handbook of Methods (2019a) and the ORS website (BLS 2019c).

The goal of the ORS is to collect and publish occupational information that will replace outdated data currently used by SSA (n.d.; U.S. Department of Labor 1991). All outputs generated from ORS data will be made public for use by non-profits, employment agencies, state or federal agencies, the disability community, and other stakeholders.

The ORS data are collected by field economists. The field economists collect approximately 70 data elements related to the occupational requirements of a job. The following four groups of data are collected:

- Physical demands of work such as keyboarding and lifting
- Environmental conditions such as extreme heat and cold
- Vocational preparation including education, prior work experience, and training
- Mental and cognitive demands of work including decision making and communication

3. Motivation for New ORS Survey Design

The ORS aims to produce measures for specific occupations. However, occupations are not evenly spread among the more than 140 million workers in the United States, and certain occupations cover relatively few workers. These low-employment occupations are varied, ranging from transportation occupations such as Ship Engineers, to manufacturing occupations such as Metal Pourers and Casters, to scientific occupations such as Astronomers. Such occupations are therefore not, as a group, concentrated in typical establishment sampling strata, which are generally based on widely available characteristics such as geographic location, industry, and establishment size. Some occupations can even have relatively low employment within establishments where they exist, adding additional difficulty to sampling them.

Tourangeau (2014) defines hard-to-reach populations as those groups that are hard to sample, hard to identify, hard to contact, hard to induce to respond, and/or hard to interview. Because low-employment occupations are hard to sample, the ORS will not often encounter them using typical methods, potentially leaving them poorly represented by the survey. Although low representation of rare sub-populations in a sample can bias population-wide measures, it is a much bigger concern when measures are needed for the rare subpopulations themselves. The need for occupation-specific estimates of job characteristics means that sampling low-employment, or rare, occupations must be a focus of the ORS sample design in order to produce reliable, comprehensive estimates.

Although hard-to-reach populations are commonly associated with demographic groups such as migrants, the homeless, or ethnic minorities, the “the needle-in-a-haystack phenomenon” (Willis, Smith, Shariff-Marco, & English 2014) applies to occupational groups, as well, because available sample frames do not reveal where rare occupations can be found. Sample frames of business establishments without occupational detail are generally available, but frames with data on the specific occupations present at specific

establishments are not. A predicted occupation frame, while unavoidably imperfect as it is based on a model, provides the ORS a nearly universal indicator of the occupations that might plausibly be found in any given establishment and therefore where rare occupations might be located.

Methods for sampling hard-to-reach populations are often focused on a single rare population. Often, though, as in the case of the ORS, measures will be estimated for various subgroups. In fact, the ORS aims to produce estimates for hundreds of occupations, around one-fifth of which are considered rare. This could make some sampling approaches inefficient or impractical, in particular because the ORS is attempting to conduct probability-based sampling (Kalton 2009).

A relatively straightforward and time-tested probabilistic approach for targeting rare populations is disproportionate stratified sampling. Since rare occupations are so disparate, they are spread across the economy. Establishments with these occupations cannot be separated from the rest of the population by typical establishment characteristics such as location, industry, or employment size. However, using a predicted occupation frame as a guide to where rare occupations might be, each "typical" sample stratum can be separated into two sub-strata: (1) establishments predicted to have at least one rare occupation and (2) establishments predicted to not have any rare occupations. Then, establishments can be oversampled from the first sub-stratum at the first (establishment) stage of selection. In the second (occupation) stage, rare occupations can be sampled with near certainty to maximize their presence in the sample while still allowing collection of all occupations in the economy (Kalton 2009; Kish 1965).

4. Overview of New ORS Survey Design

4.1 ORS Sample

The new ORS sample design will be applied to the five years of ORS sample that began collection in autumn of 2018. This paper will sometimes refer to the new sample as the second wave of the ORS to distinguish it from the first wave of the ORS, which was the separate, three-year ORS production sample that was completed in 2018.

The major goals of the new ORS sample design are to:

(1) Improve the distribution of the number of observations sampled across all occupations.
(2) Produce more published estimates across a greater number of occupations while maintaining current resource levels.

4.2 ORS Sample Frame

The predicted occupation frame used for the ORS is created by researchers in the BLS's Occupational Employment Statistics (OES) program (BLS 2019b). The OES uses the staffing patterns identified in its 1.2 million-establishment sample to model occupation distributions at the detailed geography, industry, and establishment employment level. Their model is then applied to all private industry establishments in the BLS's Quarterly Census of Employment and Wages (QCEW), which is the main establishment sampling frame for the BLS's surveys. This results in a list of potential occupational observations for each establishment.

Because of data constraints, the occupation frame is not available for state and local government establishments. Government establishments are also selected from the QCEW frame, but occupational observations are selected in the field from occupation lists provided by ORS respondents, following the procedures from the first wave of the ORS.

4.3 ORS Establishment Sample Selection

The private industry establishment sample comprises 42,500 establishments over five years, accounting for 85% of the total five-year ORS sample size of 50,000. The remaining 15% is allocated to State and Local Government establishments. Sample allocation is carried out each year using updated frame information, and 8,500 private industry and 1,500 government establishments are selected annually.

The private industry sample is a two-stage stratified sample of private industry establishments and occupations within selected establishments. Strata are formed by the cross-classification of the predicted presence or absence of a rare occupation in the establishment, the Census Region (Northeast, Southeast, Midwest, West), and the aggregate industry (Education, Goods Producing, Health Care, Financial Activities, Service Providing), leading to forty strata. A higher proportion of the total private industry sample size is allocated to the twenty rare-occupation strata than to the twenty non-rare-occupation strata. Establishment allocation to the cells within the rare/non-rare strata is proportional to total employment within the stratum. At the first stage of sample selection, private industry establishments are selected with probability proportional to the employment size of the establishment.

Because of the government frame limitations, the sample design and selection procedures for the State and Local Government sample are largely unchanged from the first wave of the ORS. The government sample is a two-stage stratified sample of establishments and occupations within selected establishments. The government establishment sample is allocated by industry proportional to the total employment within each of ten sample strata. Establishments are selected with probability proportional to the employment size of the establishment.

4.4 ORS Job Selection Process

ORS occupations are classified according to the 2018 Standard Occupational Classification (SOC) system. Prior to sampling, the ORS program defines occupations at the six-digit SOC level as rare or non-rare based on their national employment.[1] The 200 least common occupations are designated rare. The rare-occupation designation is used to prioritize occupations found on the predicted occupation frame, which applies only to private industry establishments. In order to focus sampling on occupations that have a viable chance of collection, certain occupations are ineligible to be designated as rare (though they continue to be in-scope for ORS collection):

1. Agricultural occupations, because the agriculture industry is out of scope for the ORS
2. Occupations primarily found in government, because government establishments are not on predicted occupation frame

[1] Based on May 2017 employment reported by the Occupational Employment Statistics program (https://www.bls.gov/oes/home.htm).

3. Residual-category occupations, such as "Drafters, All Other," because pre-sampling tests showed them to be difficult to collect (more information on these occupations is provided later in the paper)
4. Occupations in SOC codes that were not present in the predicted frame, because OES uses the 2010 version of SOC while ORS uses the 2018 version of SOC

ORS job selection is begun at the time of establishment sampling for the majority of ORS establishments, including most private industry establishments. These pre-selected jobs (PSJ) are the source of ORS data most of the time. For an establishment using a pre-selected job list, any rare occupations are selected with certainty and are placed first on the list (with a few exceptions for establishments with a large number of rare occupations, in order to allow all occupations the chance for inclusion in the sample and collection in the field). Remaining occupations are selected using an equal probability sampling method and are added to the end of the job list. The data collector proceeds through the pre-selected job list in order, stopping collection when the target is met or the list is exhausted (potentially coming up short of the target).

There are three categories of ORS establishments where occupations are selected at the time of data collection rather than via a pre-selected job list. For these establishments, data collectors follow the probability selection of occupations (PSO) method that was used during the first wave of ORS collection:

1. Private industry establishments that overlap with the National Compensation Survey (NCS). ORS information is collected for the NCS occupations, rather than the pre-selected jobs, in order to reduce respondent burden.
2. Private industry (non-overlapping) establishments in which none of the pre-selected occupations match to jobs within the establishment. ORS occupations are selected from the establishment's existing occupations using the PSO method, as a fallback.
3. State government and local government establishments. ORS occupations are selected using the PSO method because predicted occupation information is not available for them from the OES.

Each establishment, regardless of the occupational selection method, is assigned a target number of occupations for collection, based on its employment. When pre-selected jobs are provided, the occupation sample for an establishment provides two times the target number of occupations (up to the number of distinct occupations on the frame for the establishment) to allow for the fact that frame is based on a model and was not verified within in the establishment before sample selection. See Table 1 for the targets.

Table 1. Target Number of Occupations by Establishment Size

Establishment Employment	Target Number of Occupations*	Pre-selected Jobs Provided*, If Applicable
1	1	2
2	2	4
3	3	6
4 – 49	4	8
50 – 249	6	12
250 or more	8	16

*Maximum. The target and number of jobs provided may be limited by the number of distinct jobs at the establishment.

For establishments using PSO, the target number of occupations is selected in the field. The PSO method gives more common occupations a greater chance of selection and allows occupations to overlap (sometimes reducing the unique occupation yield).

Although pre-selected jobs are provided at the six-digit SOC level, all occupations are coded at the eight-digit O*NET-SOC level, regardless of occupation selection method (U.S. Department of Labor 2019). For most occupations, there is no difference between the two levels, but some six-digit SOC codes are broken out into more detailed occupations at the eight-digit O*NET-SOC level. For example, Registered Nurses (six-digit SOC code 29-1141) can be classified into one of five eight-digit SOC codes:

- Registered Nurses (29-1141.00)
- Acute Care Nurses (29-1141.01)
- Advanced Practice Psychiatric Nurses (29-1141.02)
- Critical Care Nurses (29-1141.03)
- Clinical Nurse Specialists (29-1141.04)

5. Results of Collection

5.1 Scope of Results

Results are based on ORS data collected as of August 6, 2019, which was the end of the data collection period. At that time, 100% of the first one-year sample under the new sample design had been collected, but data review was ongoing. The final set of reviewed and validated data was not available in time for this paper. As a result, all measures are based on unweighted counts.

Occupational results for “PSJ and PSO Establishments” include data from all usable establishments, regardless of occupational selection method (covering private industry and government establishments). Occupational results for “PSJ Establishments Only” include data only from usable establishments where a pre-selected job list was used to select occupations (covering most, though not all, private industry establishments).

In about half of establishments, a pre-selected job list from the predicted occupational frame was used to determine the ORS occupations (52%). Fallback PSO, done when collection of pre-selected jobs was attempted but not possible, was uncommon (2%). PSO was used for 17% of establishments, including establishments using fallback PSO. PSO

was specified by the sample design for all private industry NCS overlaps (3%) and state and local government establishments (12%). Respondent refusal at the establishment level accounted for 29% of establishments, and the remaining establishments were unusable as a result of business closure, out-of-scope location, out-of-scope industry, or lack of in-scope occupations (3%). (See Figure 1.)

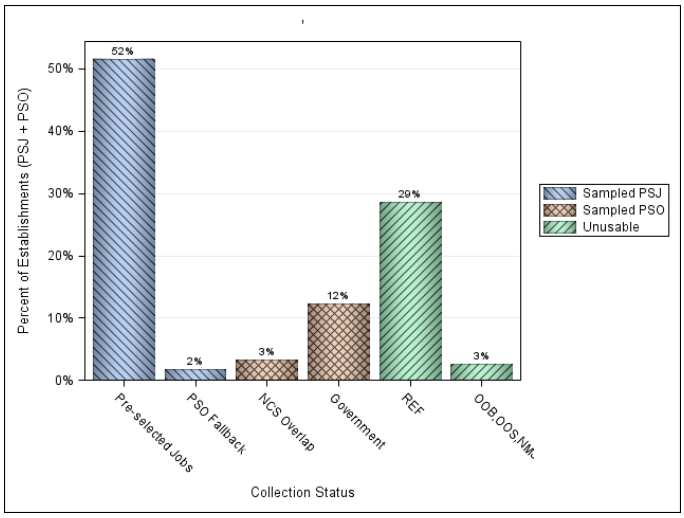

Figure 1. Percent of completed establishments, by collection status

A responding establishment was defined as an establishment with at least one responding occupation, which means that at least one occupation had at least one piece of usable ORS data. This definition led to a 69% establishment response rate from all sampled units and a 71% response rate from all viable units. The remainder of the paper deals with data from only the responding establishments.

5.2 Establishment Level Results

Data from establishments were usually collected using the occupation sampling method that had been assigned during sampling. Seventy-five percent of the responding establishments provided data on occupations from a pre-selected job list (PSJ establishments), compared to 77% that were sampled as PSJ establishments. In the remaining establishments, probability selection of occupations was performed (PSO establishments). Most PSO establishments were assigned as PSO establishments (as government establishments or NCS overlaps), but some establishments used PSO as a fallback after none of the pre-selected jobs were found at the establishment. Fallback PSO was most common in establishments with fewer than 50 employees, where 8% of PSJ establishments used fallback PSO, compared to 1% of establishments in larger establishments.

Each establishment was assigned a target or ideal number of occupations to collect. While it was expected that the target would sometimes be missed because of incorrect frame information (for PSJ establishments) or overlapped occupations (for PSO establishments), the sample design assumed that the number of collected occupations would come reasonably close to the number of target occupations. A PSJ establishment where the number of matched occupations equaled the target number of unique occupations was

considered to have met the target, regardless of whether usable occupation data were collected for every occupation (a matched occupation could be out of scope or a refusal). The target was met in 43% of PSJ establishments. Meeting the target was least common in PSJ establishments with fewer than 50 employees, where 32% met the target. In larger PSJ establishments, 47% of establishments met the target. In PSO establishments, the target was met if there were no collapsed occupations. Meeting the target was more common in PSO establishments than PSJ establishments overall; 52% of all PSO establishments met the target. (See Figure 2.)

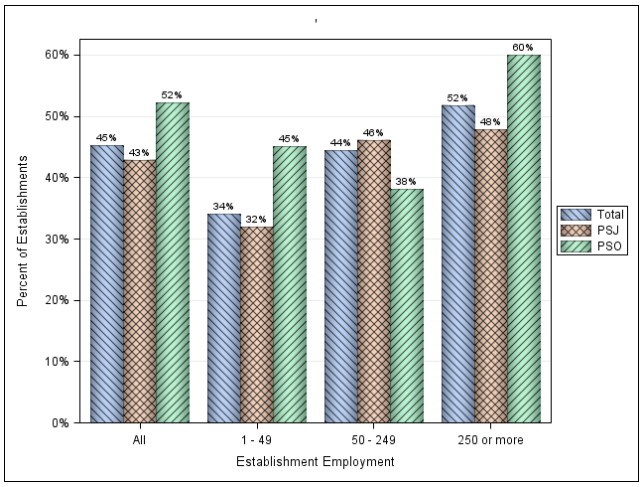

Figure 2. Percent of usable establishments where the target number of occupations was collected, by establishment employment size class and occupational selection method

Across all PSJ and PSO establishments, an average of 6.23 unique occupations per establishment were targeted for collection, and 76% of this target (4.72) were matched or collected. In establishments with fewer than 50 employees, about two-thirds of the total target number of occupations was collected. The proportion in small establishments has decreased compared to the first wave of the ORS survey. At the same time, the proportion has increased in establishments with 50-249 employees. The average number of occupations collected per establishment, or occupation yield, depended on the establishment employment, ranging from 2.08 for establishments with fewer than 50 employees to 4.41 for establishments with 250 or more employees. (Such a range was expected because the target number of occupations correlates with establishment size.) The occupation yield for the current ORS sample has increased slightly compared to the ORS first-wave sample overall, while both the smallest (with less than 50 employees) and largest (with 250 or more employees) establishments have experienced a decrease compared to the first-wave sample. On average, 3.59 unique occupations per establishment were collected with usable data, which is 58% of the targeted number of occupations. (See Figure 3 and Figure 4.)

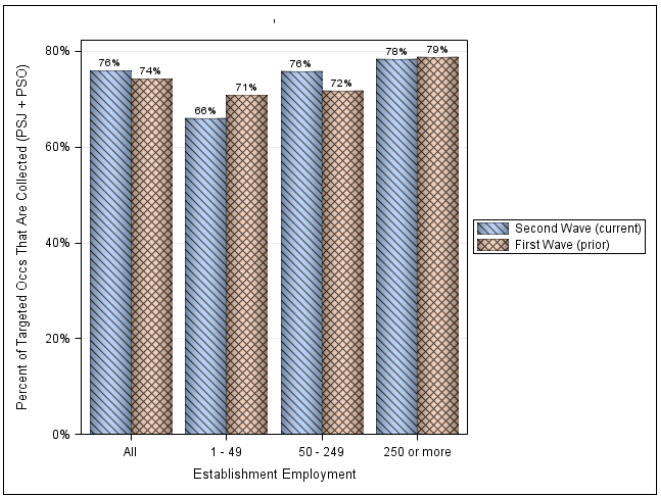

Figure 3. Percent of target number of occupations that was collected, by establishment employment size class

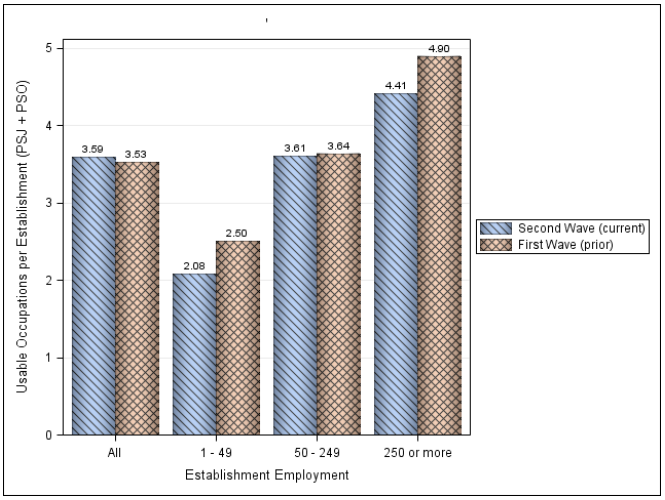

Figure 4. Average number of usable occupations per usable establishment, by establishment employment size class

Rare occupations were sampled in all establishments predicted to have at least one rare occupation (sampled-rare establishments). However, the presence of rare occupations was not certain at the time of sample selection because the frame was based on a model. A rare occupation was collected from 59% of sampled-rare establishments. Matching rare occupations was more common in large establishments with 250 or more employees, where at least one rare occupation was found in 70% of sampled-rare establishments. An average of 3.1 rare occupations were provided for each of these large sampled-rare establishments. Finding rare occupations was less common in smaller establishments, where an average of 1.2 rare occupations were provided per sampled-rare establishment; a rare occupation was found in 43% of establishments with fewer than 50 employees. (See Figure 5.)

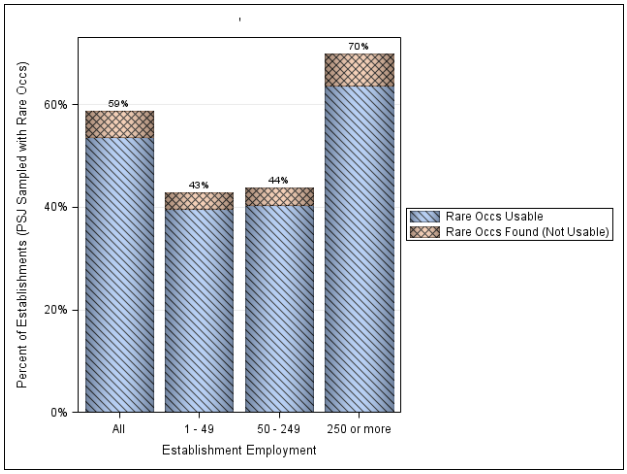

Figure 5. Percent of establishments where rare occupations were found, out of establishments sampled with rare occupations, by establishment employment size class

When a list of an establishment's occupations was obtained before the interview, the data collector could do preliminary work to match sampled jobs to actual establishment jobs. This could potentially leave more time for the collection of ORS element data. A small establishment's occupation list was obtained before the ORS interview 26% of the time. In larger establishments, the list was obtained before the interview 52% of the time.

5.3 Occupation Level Results

5.3.1 Pre-selected Jobs (PSJ)

Collection was attempted for occupations on an establishment's pre-selected job list until the target was met, the list was exhausted, or the respondent ended the interview. As a result, some jobs on the list might not be used. On average, 25% of all occupations on the lists were not used, and 11% of rare occupations were not used. Results in this section are based on occupations that were used – that is, they were either matched to an establishment position or not found at the establishment.

About half of occupational observations for which collection was attempted were matched to an establishment position (52%); the other half were not found at the establishment. The match rate among rare occupations was similar (53%). The overall match rate varied by occupation group, ranging from 48% in Management, Business, and Financial occupations to 60% in Service occupations. The rare occupation match rate ranged from 22% in Sales and Related occupations (where there are only two rare occupations) to 57% in Service occupations. (See Figure 6.)

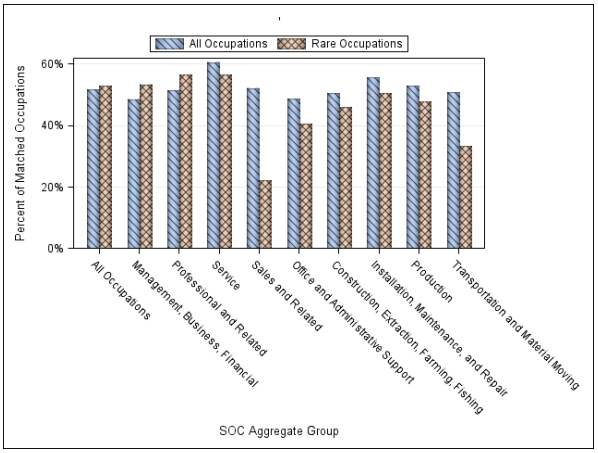

Figure 6. Percent of occupational observations that were matched in establishments with pre-selected jobs, comparing all occupations to rare occupations, by aggregate SOC group

While occupational observations were matched about half of the time, the plurality of the 848 non-military detailed occupations (occupational titles) had match rates between 25% and 34%. About one-fifth of detailed occupations had match rates of 50% or more. A larger percentage (37%) of the rare detailed occupations had match rates of 50% or more. (See Figure 7.)

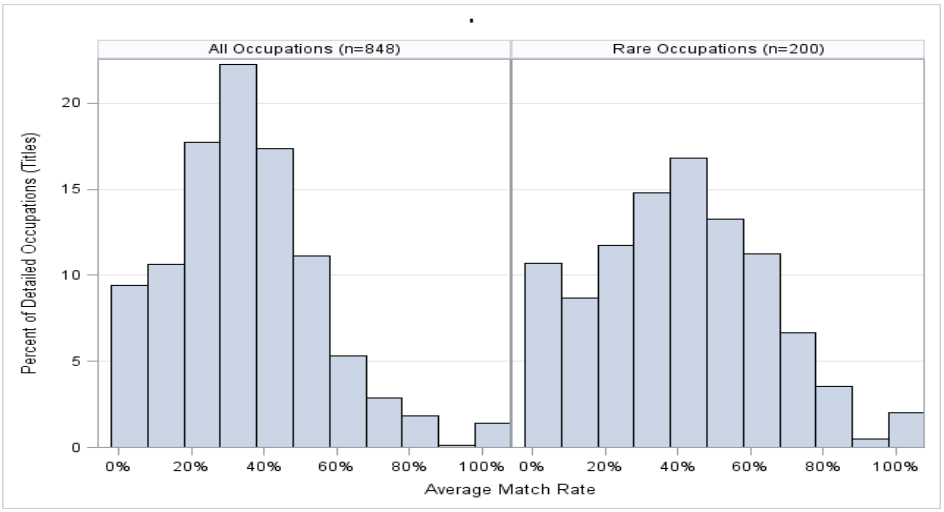

Figure 7. Percent of detailed occupations (titles) by average match rate for the occupation, for all occupations and rare occupations.

Some occupations were matched less frequently than average. These include the "all-other"-labeled occupations, which are residual groups for workers not included in specifically defined occupations. For example, the "Drafters, All Other" occupation includes all drafters that are not "Architectural and Civil Drafters," "Electrical and Electronics Drafters," or "Mechanical Drafters." In the 74 residual-group occupations, 27% of observations where collection was attempted were matched, compared to 52% of jobs matched across all occupations. (See Figure 8.)

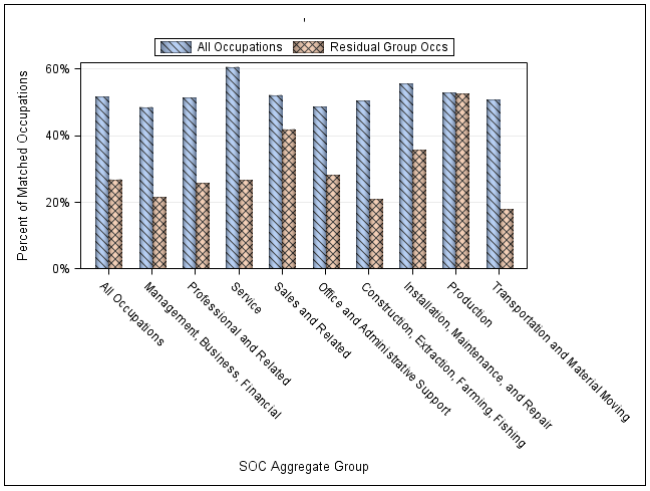

Figure 8. Percent of occupational observations that were matched in establishments with pre-selected jobs, comparing all occupations to occupations in the "All Other" residual categories, by aggregate SOC group

Three-quarters of the 74 residual-group detailed occupations (titles) had match rates under 30%, compared to 37% of all detailed occupations. (See Figure 9.)

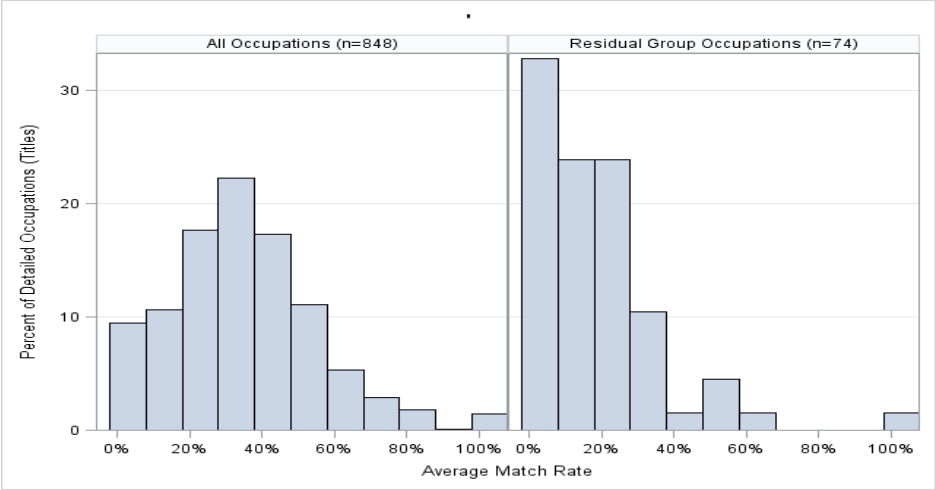

Figure 9. Percent of detailed occupations (titles) by average match rate for the occupation, for all occupations and residual group occupations.

After a data collector matched an occupation to a position in the establishment, the collection of ORS data provided additional information that was useful for classifying the occupation by SOC. It was sometimes determined that the collected occupation belonged to a different SOC than was sampled and subsequently matched. At this point, it was not feasible to start over with a new occupation in the sampled SOC (if one even existed in the establishment), so data collectors were able to update the SOC code of a collected occupation if necessary. Updates to SOC codes after occupation matching were uncommon. About 86% of matched observations were coded within the same 6-digit SOC for which they were sampled. The rate for rare occupations was 89%. The rates ranged from 81% for Office and Administrative Support occupations to 91% for Service

occupations. Most occupations (95%) remained in the same two-digit (major group) SOC designation as they were sampled. (See Figure 10.)

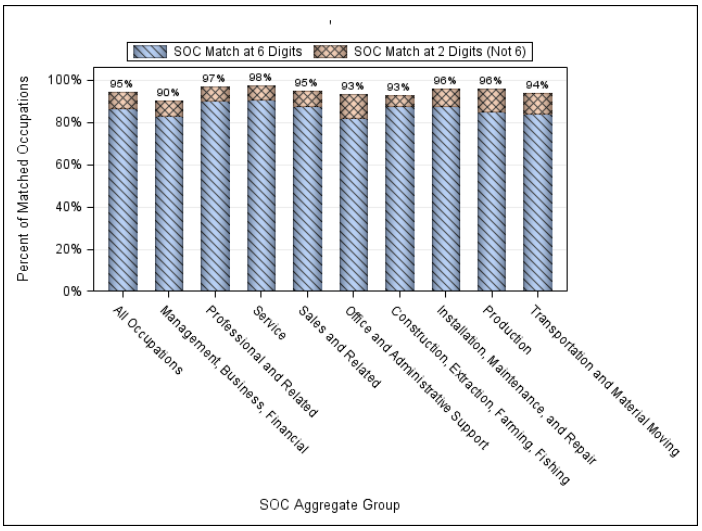

Figure 10. Percent of occupational observations that were matched to the sampled SOC code, by level of match, by aggregate SOC group

The majority of the 848 detailed occupations (titles) had six-digit SOC match rates above 85%. This was also true among rare detailed occupations. (See Figure 11.)

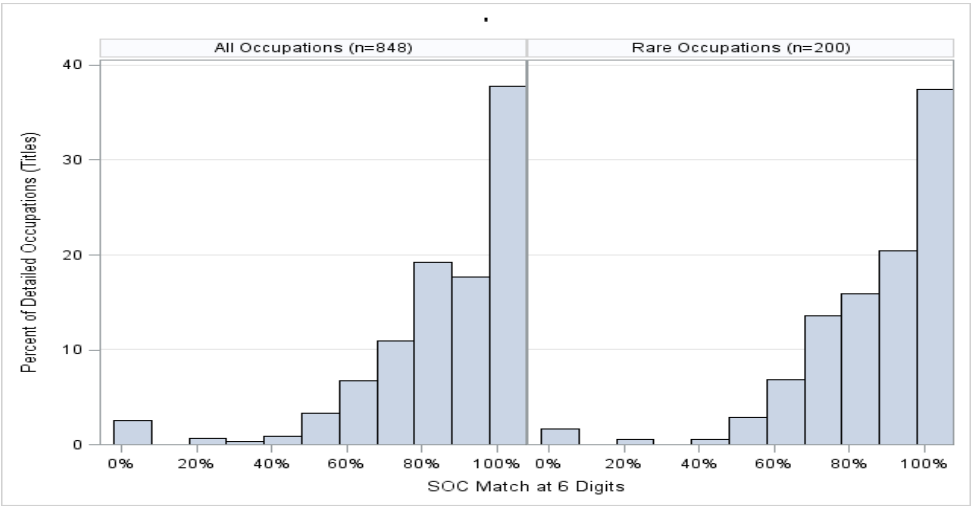

Figure 11. Percent of detailed occupations (titles) that were matched to the sampled SOC code, for all occupations and rare occupations.

A collected occupation was considered usable when data for ORS elements were provided by the respondent. The majority of occupational observations in PSJ establishments provided usable ORS data (78%). The occupational response rate was higher for rare PSJ occupations (83%). (See Figure 12.)

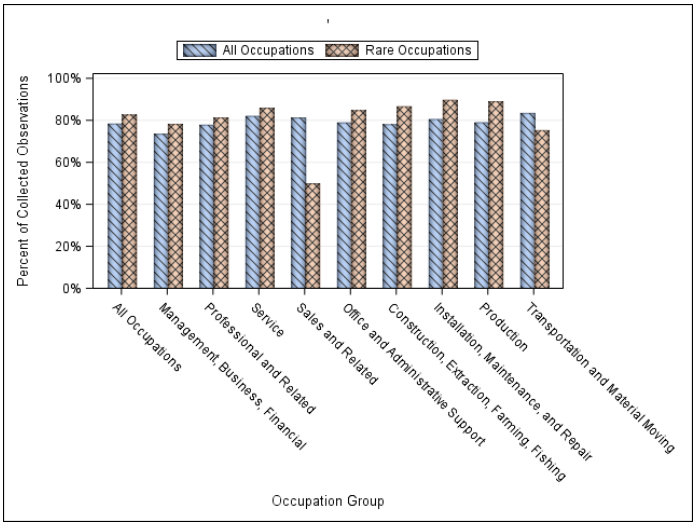

Figure 12. Percent of usable occupational observations out of observations that were collected, PSJ establishments, by aggregate SOC group

The majority of the 848 detailed occupations (titles) had response rates above 75%. The majority of rare detailed occupations had response rates above 80%. (See Figure 13.)

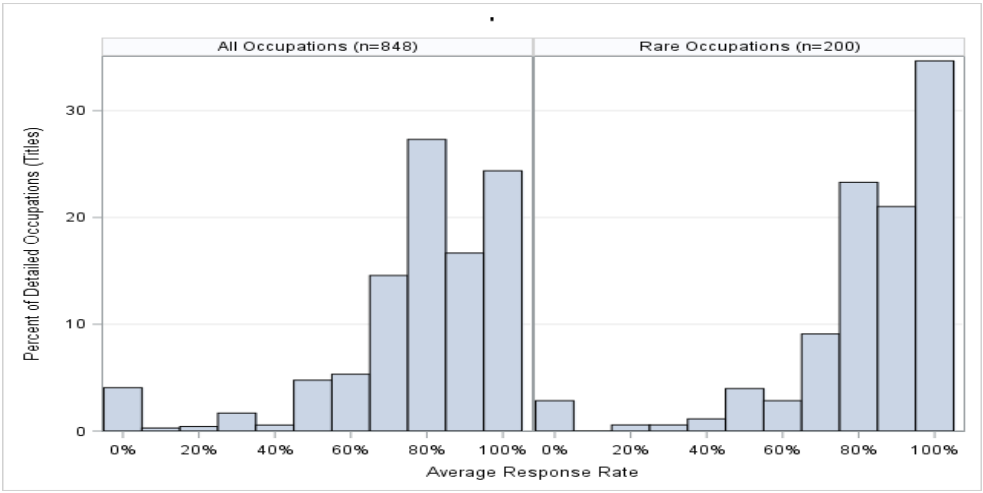

Figure 13. Percent of detailed occupations (titles) that were usable, PSJ establishments, for all occupations and rare occupations.

5.3.2 Total Occupations (PSJ + PSO)

Response rates were higher among occupations in PSJ establishments (78%) than in PSO establishments (71%); the difference was especially marked among rare occupations (83% PSJ compared to 58% PSO). When occupational data from PSO establishments were combined with PSJ data, the occupational response rate was 76%, and the response rate for rare occupations was 81%. The response rate ranged from 72% in Management, Business, and Financial occupations to 82% in Transportation and Material Moving occupations. (See Figure 14.)

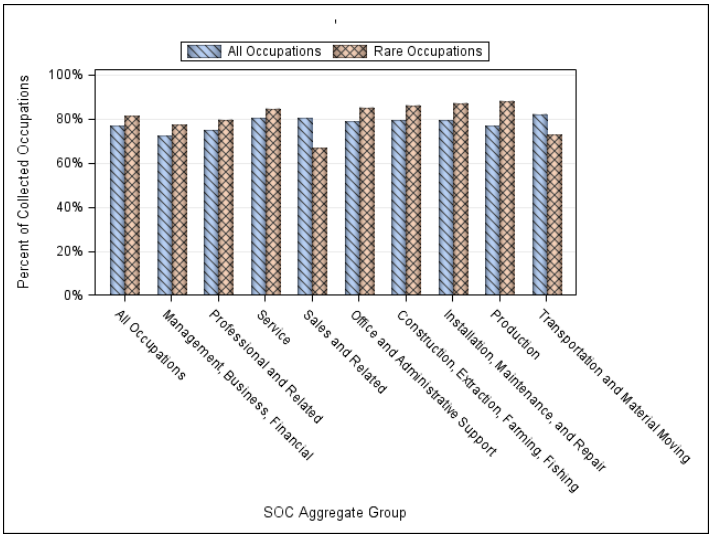

Figure 14. Percent of usable occupational observations out of observations that were collected, PSJ and PSO establishments, by aggregate SOC group

The majority of the 848 non-military detailed occupations (occupational titles) had a response rate of 75% or more. The majority of rare detailed occupations had response rates above 80%. (See Figure 15.)

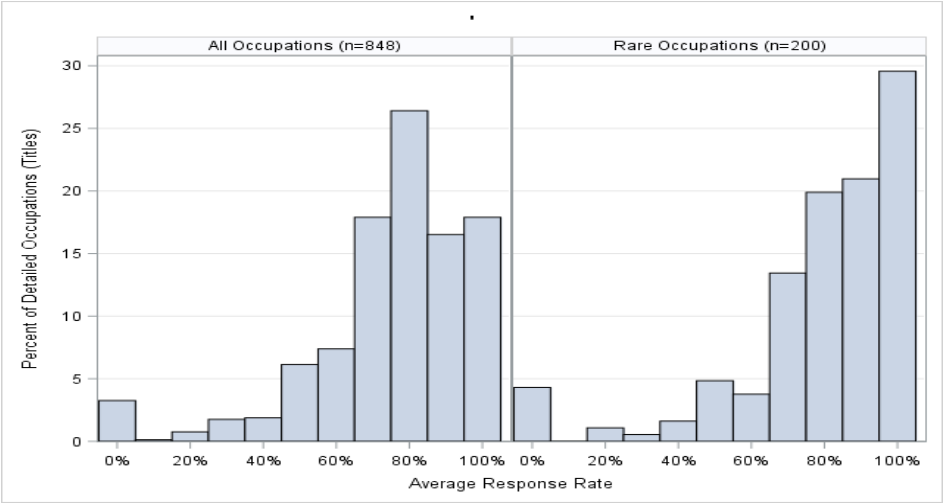

Figure 15. Percent of detailed occupations (titles) that were usable, PSJ and PSO establishments, for all occupations and rare occupations.

Detailed occupations (occupational titles) will be eligible for ORS estimates if at least 30 usable observations are collected (other publication criteria are also applied, so 30 observations does not guarantee estimates). Of the 848 non-military detailed occupations, 206 had 30 or more usable observations; 26 of the 206 potentially publishable occupations were rare occupations. About one quarter of the potentially publishable occupations (57) were not published during the first wave of ORS (which included three years of sample). None of the 26 rare occupations were published during the first wave of the ORS.

Overall, one-quarter of the detailed occupations are potentially publishable after the first year of the sample design; 6% of potentially publishable occupations were not published after the three years of the first wave. After the first wave, none of the rare occupations were publishable, but 13% of rare occupations are potentially publishable after the first year of the second wave.

On the other end of the spectrum, there were 48 occupations (6% of occupations) for which observations were sampled in PSJ establishments, but no usable observations were collected in either PSJ or PSO establishments. Twenty-one of these were rare occupations (11% of rare occupations). An additional 46 occupations (5%), including one rare occupation, had no sampled observations in PSJ establishments and no usable observations in PSO establishments.

6. Discussion of Results

As stated early in the paper, the major goals of the new ORS sample design are to:

(1) Improve the distribution of the number of observations sampled across all occupations.
(2) Produce more published estimates across a greater number of occupations while maintaining current resource levels.

To achieve the goals, it is helpful to maximize the overall number of occupational observations collected. Since resource levels, in particular sample sizes, remain the same in the second wave compared to the first wave, large gains in the overall number of usable observations were not expected. It is therefore positive that the second wave is performing slightly (though not substantially) better than the first wave on this measure.

However, it appears that total occupation counts per establishment lagged in small establishments. Establishments with fewer than 50 employees were less likely to meet the target number of occupations, and predicted occupations were less likely to be matched in these establishments. Compared to the first wave, the proportion of target occupations that were usable decreased. Small establishments were also more likely to require fallback PSO, which dilutes the advantage of targeting specific occupations. These details indicate that the predicted occupation frame may be less efficient for small establishments. By definition their employment is low, so small establishments simply did not have as many frame occupations as larger establishments. Only 29% of establishments with fewer than 50 employees had sufficient frame occupations to be provided twice the target number of occupations, while 85% of larger establishments were provided twice the target. As a result, small establishments on average were provided 1.47 occupations per targeted occupation, compared to 1.92 in larger establishments. Any negative effects of the lower performance among small establishments may be mitigated somewhat in the second wave, though: Small establishments comprised a smaller proportion of the ORS sample in the second wave (23%) than the first wave (39%), a consequence of small establishments not having as many rare occupations generally (establishments with rare occupations were sampled at a higher rate).

Occupational response rates also contribute to how much data are collected. The overall response rate of 76% was relatively good; the response rate was a bit higher in PSJ establishments, which made up three-quarters of the collected sample, and lower in PSO

establishments. At the same time, meeting the target number of occupations was more common in PSO establishments, even though these establishments were generally larger and had more occupations to collect (they were mostly government units and overlap units). The timing of collection activities within the ORS interview might have contributed to these PSJ-to-PSO differences. In PSO, the occupation-matching step was not done, and all occupations were immediately identified from the job list provided by the respondent. Therefore, it might have been easier to hit the target number of occupations given a limited amount of time. Time might have run out, however, at the occupation data collection stage because the establishments were often large enough to have a target of eight occupations, resulting in more occupation refusals. It is telling that in PSJ establishments, rare occupations, which are listed first, had a higher response rate than non-rare occupations, hinting that response propensity decreases as the interview progresses. (Note that weighting procedures will take into account situations where the target is not met; it is only the loss of unique occupational observations that is a concern.)

To improve both the quantity and distribution of occupational observations, it is helpful to maximize the amount of pre-selected occupations that are matched, because this makes it more likely that the target will be met and provides control over the occupations that are collected. Providing twice the target number of occupations seemed generally sufficient, as the overall match rate was 52%. The larger job list on average accounted appropriately for inaccuracies in the frame without burdening the respondent with too many job-matching attempts. The match rate, however, varied by occupation and by establishment size, so adjustments might be needed in sampling rates at the occupation level if there are shortfalls. Since most occupations have match rates below 50%, it appears that the best match rates occur in common (heavily sampled) occupations; familiarity with an occupation might contribute to frame accuracy, match rates, or both.

Collecting ORS data elements for rare occupations is one of the goals of the sample design and seems to be succeeding generally. The occupation frame on average predicts the presence of rare occupations as well as it predicts the more common occupations. Upon collection, overall occupation code (SOC) match rates were not different by rare/non-rare status, so the less common occupations were not more difficult to identify accurately. Also, rare occupations actually had a higher response rate than average. It seems likely that the occupations listed first resulted in usable data more often because they were collected before respondent fatigue set in or the respondent ran out of time.

The prospect of publishing estimates for rare occupations has improved. During the first wave of the ORS, estimates were not published for any of the 200 rare occupations, but 26 are now eligible for published estimates. This is only the first year of a five year sample design, so it is expected that additional occupations, rare and non-rare, will become publishable in the future. In addition, since there are about 70 ORS elements, each with various estimates, the breadth of estimates for occupations published after the first year can increase over the next four years.

7. Conclusions and Future Work

The ORS data collected so far show that the new sample design is yielding a similar amount of occupation data as the first wave ORS sample design while allowing more control over

how the occupations are distributed. Rare occupations as a group are performing at least as well as all occupations for several measures of interest, such as the occupational match rate and occupational response rate. At the same time, the number of unique usable occupations per establishments has increased slightly, showing that the emphasis on less common occupations is not hindering the collection of more common occupations. In the aggregate, the ORS continues to collect data on enough occupations to get reasonably close to collecting the ideal (target) number of occupations.

Over the remaining years of the second wave of the ORS, the ORS program will continue to monitor the collection of occupations. Occupational distributions and occupational response will be studied to see if patterns shift. The ORS program will examine occupational data to determine how to allocate future samples, with the goal of maximizing the number of occupations with published data. Similarly, data from small establishments will be tracked to determine if procedures should be tailored to establishment size. There will also be a more thorough examination of the amount of ORS element data that is usable within an occupation, in particular ORS element durations.

References

Ferguson, G., McNulty, E., and Ponikowski, C. (2014). Occupational Requirements Survey Sample Design Evaluation. JSM proceedings, Government Statistics Section, American Statistical Association.

Ferguson, G. and McNulty, E. (2015). Occupational Requirements Survey Sample Design. JSM proceedings, Government Statistics Section, American Statistical Association.

Kalton, G. (2009). Methods for oversampling rare subpopulations in social surveys, *Survey Methodology*, Vol. 35, No. 2, pp. 125-144.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons, Inc.

Rhein, B., Ponikowski, C., and McNulty, E. (2013). Sample Design Considerations for the Occupational Requirements Survey. FCSM Papers and Proceedings, Federal Committee on Statistical Methodology Research Conference.

Social Security Administration (n.d.). Occupational Information System Project, https://www.ssa.gov/disabilityresearch/occupational_info_systems.html.

Tourangeau, R. (2014). Defining Hard-to-Survey Populations. *Hard-to-Survey Populations*, R. Tourangeau, B. Edwards, T.P. Johnson, K.M. Wolter, and N. Bates (eds). Cambridge: Cambridge University Press, (In Press).

U.S. Bureau of Labor Statistics (2019a). *BLS Handbook of Methods*, Occupational Requirements Survey, https://www.bls.gov/opub/hom/ors/home.htm.

U.S. Bureau of Labor Statistics (2019b). Occupational Employment Statistics Program, http://www.bls.gov/oes/.

U.S. Bureau of Labor Statistics (2019c). Occupational Requirements Survey, http://www.bls.gov/ors/.

U.S. Department of Labor, Employment and Training Administration (1991). *Dictionary of Occupational Titles*, Fourth Edition, Revised 1991.

U.S. Department of Labor (2019). O*NET Online, http://www.onetonline.org/.

Willis, G., Smith, T., Shariff-Marco, S., and English, N. (2014). Overview of the Special Issue on Surveying the Hard-to-Reach. *Journal of Official Statistics*, Vol. 30, No. 2, pp. 171–176.