Appendix G: Paper on Model-Aided Sampling for the O*NET Data Collection Program[1]

1 Berzofsky, M. E., Welch, B., Williams, R. L., & Biemer, P. P. (2006). Using a model-assisted sampling paradigm instead of a traditional sampling paradigm in a nationally representative establishment survey. Proceedings of the American Statistical Association, Section on Survey Research Methods, 2763–2770.

Using a Model-Assisted Sampling Paradigm instead of a Traditional Sampling Paradigm in a Nationally Representative Establishment Survey

Marcus Berzofsky,[1] Brandon Welch,[1] Rick Williams,[1] and Paul Biemer[1]
[1]RTI International

Key Words: traditional sampling, model-based sampling, quota sampling, business establishment surveys, O*NET

1. Introduction

1.1 Two Sampling Paradigms

Historically, the sampling of finite populations has been conducted with one of two methods: a probability-based approach or a pure model-based approach (Moser, 1952; Moser & Stuart, 1953). For large, federally funded surveys, the probability-based approach, as defined by Neyman (1934), has been deemed the superior of the two methods by the statistical community (Kish, 1965). However, in situations where the population of interest is difficult to find or the sample size is very small, a third approach may be best. This approach, called *Model-Assisted Sampling* (MAS), combines traditional probability sampling with quota sampling and may be viewed as a type of model-based sampling. It can be highly effective in providing results that allow inference to the general population while controlling costs (Sudman, 1966). This paper describes the application of MAS to the Occupation Information Network (O*NET) Data Collection Program and evaluates how it compares with probability-based sampling. We also consider the utility of MAS in future iterations of the O*NET program.

Before defining MAS, it is important to review the key elements of the traditional sampling paradigm and contrast them with the model-based sampling paradigm. In particular, we consider the sample selection mechanism and all requirements associated with it, the data collection requirements, the types of inference that can be made, and the basis for these inferences.

If the population of interest is well defined, then the usual approach is to design the sample so that the selected units are in some sense representative of the whole population (Smith, 1983). Both traditional sampling and model-based sampling strive for this but accomplish it in very different manners. The traditional sampling paradigm requires that a precise specification of the sampling frame be made and that its coverage of the population of interest be acceptable (King, 1985). In traditional sampling, the sample can support inference only to the population implied by the sampling frame (Deming, 1960). Therefore, to minimize coverage bias, the sampling frame should have a high coverage level of the population of interest. Furthermore, under traditional sampling, the sampling units must be selected from the frame under a random process with known probabilities of selection (King, 1985). Random selection is the central tenet of the traditional paradigm and the process by which representativeness and population inference is justified.

Under model-based sampling, a model is used to define the distribution of the target population (Stephenson, 1979) with respect to the variables of interest. The model is usually defined by quotas for subgroups or cells based on the cross-classification of known demographic information relevant to the outcome of interest. Examples of quota cells include geographic region by age and, in the case of business establishments, by the industry in which the business operates. Moser and Stuart (1953) point out that the quotas can be either “independent,” which means that the quotas are based on the marginal distribution, or “interrelated,” which means that the quota requirements are made for each cross-classified subgroup. In either case no frame is explicitly required; however, knowledge of the population of interest is required for proper specification of the sampling distribution (Deville, 1991; Moser, 1952). Either a frame or another external source of information can be used for this purpose. Because a predefined model is being used to determine the sampling distribution of respondents, there are no coverage requirements for the sample. If the model assumptions hold, there is no bias in the estimates produced (Deville, 1991). Moreover, the model-based sampling paradigm does not require known selection probabilities or even random sampling. Once the quotas are defined, essentially any sampling method can be used to identify and select sample members for each quota cell (Moser, 1952).

Thus, the requirements for data collection differ greatly between the two paradigms. Under the traditional paradigm, rigid controls of field procedures are specified so that the sampling instructions are properly executed and any interviewer effects on response are minimized. In carrying out the sampling instructions, interviewers must complete data collection on the entire sample, regardless of the achieved response rate, and conduct callbacks sufficient to reduce the proportion of nonrespondents and minimize the impact of nonresponse on the survey results (King, 1985). Conversely, the model-based sampling paradigm allows data collection to stop in a particular quota cell once the quota is met. In addition, interviewers are allowed great flexibility in how they collect the data. Callbacks and other attempts to recontact nonrespondents are not required, so long as the quota requirements are achieved (Moser, 1952).

Because of the differences in sample selection and data collection methods, the two paradigms also differ from one another in methods for analysis. The traditional paradigm uses randomization to allow the creation of probability-based weights to represent the entire frame population; it argues that, even if the achieved sample is not proportionally representative, the use of survey weights minimizes any potential bias. Furthermore, standard errors are used to express the level of precision of the survey estimates. Under the model-based sampling paradigm, inference is based on a superpopulation model, which King (1983) and Deville (1991) argue can be made if the a priori sampling distribution is achieved during data collection. Deville even defines a variance estimator for quota samples, and previously Moser and Stuart (1953) defined a "standard error" for quota sample designs using resampling methods. Furthermore, although the model-based sampling paradigm does not use probability-based weights, it often incorporates poststratification for making descriptive inferences to a specific population (Smith, 1983).

Although these two paradigms appear to be diametrically different and incompatible, the model-based sampling paradigm is often used to complement more traditional methods as the last sampling technique used in a multistage stratified survey (Deville, 1991). Here we empirically examine the accuracy of a MAS design that combines elements of both paradigms for estimates obtained in the O*NET program.

1.2 Application to the O*NET Data Collection Program

The O*NET project is a survey of workers contacted through a nationally representative business establishment survey that produces estimates for more than 800 occupations in the United States, across four occupational domains—skills, work context, work activities, and knowledge. Hence, O*NET is simultaneously conducting over 3,200 surveys. The O*NET program differs from most large-scale surveys in that it is targeting a large number of subpopulations, which yields a large number of completed questionnaires in aggregate, but at the occupation-by-domain level the sample sizes are relatively small. Furthermore, with limited empirical information, predicting eligibility and response rates for each of these subpopulations is difficult. Thus, it is problematic to accurately determine the number of sampling units to release in order to obtain the desired number of responses.

The current data collection began in 2001 and has compiled information by more than 110,000 survey respondents. To date, estimates have been derived under the traditional paradigm for more than 700 of the 810 occupations at the national level, with an average of 144 questionnaires collected per occupation (median = 117). For each occupation, information across each of the four occupational domains—skills, work context, work activities, and knowledge—is collected, with each respondent completing a questionnaire for one domain. The goal of the current data collection for a particular occupation is to complete a minimum of 15 questionnaires per domain for a total of 60 completed questionnaires. Of the occupations that have completed data collection, an average of 36 (median = 29) questionnaires per occupation by occupational domain have been collected. Within each domain, the O*NET program collects information on the importance of an occupational attribute (e.g., reading comprehension) on a 5-point scale, the level of need for that attribute on a 7-point scale, and estimates of proportions for "mark-all-that-apply" questions.

The sample design is a traditional multistage design that first selects establishments and then selects employees in the occupations of interest for the selected establishments. Selected employees may complete the survey by mailed paper instrument or by Web instrument. The design takes advantage of the correlation in the industries for which occupations are employed by collecting data on several occupations at a time. Currently, this design follows the guidelines of the traditional paradigm. Although the current design is effective in identifying persons of interest in aggregate, the sample size for a particular occupation by domain can be highly variable, depending on the ease with which that occupation is found in the population. This variability causes an inequality in the number of questionnaires collected across occupations.

One of the constraints on the O*NET program is the number of public burden hours approved by the Office of Management and Budget. As data collection progressed, it was observed that for some occupations a higher than desired sample size was obtained. For example, occupations, such as Secretaries, which are found in many industries, were more easily found than many others and would return a larger than desired number of questionnaires. In order to make the best use of the available burden hours, it was necessary to control the number of completed questionnaires. We found that a small number of occupations completed a large number of questionnaires and disproportionately used burden hours. Unlike other large-scale surveys, the O*NET program's large number of targeted subpopulations makes it particularly sensitive to excessive burden and cost involving any one subpopulation. In such situations, after the initial sampling units are drawn, the traditional paradigm does not provide much flexibility for sample modifications to help limit overproduction of respondents. It is therefore of interest to incorporate methods that can help control the sample sizes across occupations while ensuring that the questionnaires collected still represent the occupation of interest.

MAS, as defined for this study, incorporates a sample selection mechanism from a traditional sampling paradigm, uses data collection techniques from both paradigms, and uses analysis techniques from a model-based sampling paradigm. Our approach proposes continuation of the random, multistage design to select employees in the occupations of interest, in order to ensure that no selection bias occurs. However, before sample selection, a sampling distribution, in the form of quotas, is defined for each occupation, based on the distribution of the occupation by region, establishment size, and industry groupings for which the occupation is employed. Furthermore, during data collection a strict protocol is used to identify and contact establishments, as dictated by a traditional sampling paradigm, including multiple contact attempts to minimize nonresponse bias. Unlike the traditional paradigm, however, once enough questionnaires are projected to be completed in a quota cell for an occupation, further sampling contacts in that cell for that occupation cease. Once all quota cells are met, data collection is stopped for the entire occupation, whether or not data collection on all selected business establishments has been completed. At this point, weighted survey estimates using poststratification weights to known population totals are created for inference to the population. Here we hypothesize that estimates for occupations created under MAS will not significantly differ from the estimates created under the current traditional paradigm.

1.3 Other Studies of MAS

In the 1950s, statisticians treated the two sampling paradigms dichotomously and argued the merits of each. Leading proponents of the model-based sampling paradigm were based in England and led by Moser and Stuart (1953), and Stephan and McCarthy (1979). Proponents of the traditional sampling paradigm argued that model-based sampling led to biased results (Kish, 1965). Moser (1952) countered that, although model-based sampling may be biased with regard to certain characteristics, it may be quite satisfactory for others. The quality of estimates produced through model-based sampling depends on the model used to derive the sampling quotas. If the model holds, model-based sampling will likely give good estimates of the population quantity, but if it does not then the estimates may be badly biased (Lohr, 1999). In fact, Moser and Stuart found in their experiments comparing the traditional paradigm and the model-based paradigm few major differences in the results. However, Moser and Stuart admit that there is no theoretical evidence to suggest that model-based sampling will always produce estimates as unbiased as those from traditional sampling.

In order to bridge the theoretical gap, statisticians began developing hybrid approaches. Sudman (1966) developed "probability sampling with quotas." Under this design, the probability of respondents' being available to be interviewed defines the quota for each cell. Interviewers comply, as well, with tighter controls on how survey participants are selected; however, rules are relaxed regarding number of callbacks an interviewer must make to a selected sampling unit. In empirical testing, Sudman found that estimates under this design resembled estimates determined by traditional sampling methods. Stephenson also (1979) empirically compared "probability sampling with quotas" to traditional sampling, finding, as Sudman suggested, that it behaves much like traditional sampling, with no detectable bias for most questionnaire items. He cautioned, however, that it carries greater risk of bias due to exclusion of people who are hard to find or interview.

More recently, statisticians have argued that nonprobability samples can be analyzed through model-based inference. Smith (1983) demonstrated how a model-based approach to inference allows one to analyze nonrandom sampling in a formal way while making explicit the underlying assumptions. Smith argues that randomization is advantageous in model-based designs, not necessarily because it is essential, but because the scientific community will find the design more acceptable. Moreover, Smith advocates the use of poststratification in model-based designs when the goal is to make inference to a specific population. King (1985) used a Bayesian model based on prior information to determine the allocation of a model-based design. King determined that the classes used to define quotas had to be highly correlated to the outcome of interest in order to ensure nearly unbiased results. He concluded that the researcher must ascertain agreement between model-based sampling results and traditional sampling results before he or she implements a model-based design.

Hybrid designs have also been implemented to ensure a representative sample when response rates are expected to be very low. Sanzo, Garcia-Calabuig, Audicana, and Dehesa (1993) used a combination of random sampling and model-based sampling to estimate the prevalence of *Coxiell burnetii* infection within a region in northern Spain. Under this design, the investigators used stratified random sampling to select health care centers. However, because of concerns about an expected low response rate during the second stage of selection, the investigators derived age and gender quotas that would make the results representative of the population. Once the investigators filled a particular quota cell, they stopped collecting data in that cell. After the completion of all cells, the investigators stopped data collection.

Another recent hybrid design is multiple inverse sampling (MIS) proposed by Chang, Liu, and Han (1998). This design partitions the population into two or more subpopulations with known sizes. MIS is effective when one of these subpopulations is rare and it would be undesirable to obtain no or very few responses from the rare subpopulation. MIS selects sampling units one at a time, without replacement, until the predetermined sample sizes are obtained for all subpopulations. Through simulations, Chang et al. found that MIS is reasonably efficient when compared to simple random sampling.

2. Methods

2.1 Data

Data collected for the O*NET program were used to compare, from 79 occupations, estimates derived under each of the two sampling paradigms. Of all 810 occupations, these 79 were a representative cross section based on the educational requirements of each occupation and its relative rarity in the population. For each occupation, estimates were created for 36 items. These items spanned all four domain questionnaires and all question types (e.g., 5-point and 7-point types, and estimates of proportions). Therefore, our analysis consisted of 2,844 occupations by item-level estimates.

2.2 Quota Definitions

The first step in the MAS design is to define the model by which each occupation will be defined. This model should be based on known attributes of the occupation and incorporate characteristics that help explain all aspects of the occupation. For the O*NET project, three classifications were used to define the model: industry division, Census region, and number of employees, as shown in Table 1. MAS uses "marginal quotas with unequal rates" to represent the occupation and define each class (Deville, 1991). Under this design, the marginal totals for each subgroup must be met, but no constraints are made on the joint distribution between classes.

Table 1. MAS Quota Classifications

Industry division	
▪ Agricultural, Forestry, and Fishing	
▪ Wholesale Trade	
▪ Mining	
▪ Retail Trade	
▪ Construction	
▪ Finance, Insurance, and Real Estate (FIRE)	
▪ Manufacturing	
▪ Services	
▪ Transportation, Communications, Electric, Gas, and Sanitary Services	
▪ Government (Federal, State, and Local)	
Census region	
▪ Northeast	▪ Midwest
▪ South	▪ West
Number of employees	
▪ Unknown, 1–24	▪ 250 or more
▪ 25–249	

The industry division quotas are defined first according to the proportional distribution of employment in an occupation as found in the Occupational Employment Statistics (OES) Survey conducted by the U.S. Bureau of Labor Statistics. For each occupation, the quota for particular industries may be altered to allow for "overrepresentation" in that cell (Deville, 1991). Furthermore, small industry cells for an occupation are collapsed into a single cell. These adjustments are done to allow for a more cost-efficient data collection process and to reduce respondent burden. Once the industry quotas are determined, the region and establishment size quotas are defined according to the industries' distribution in the Dun and Bradstreet (D&B) frame. Because of the right-skewed distribution of size of establishments (i.e., number of employees), further "overrepresentation" is made in the "250 or more" employees cell to ensure that it is represented. Within each class, the quotas sum to 60, the desired sample size for each occupation.

2.3 Simulation, Stopping Rules, and Collapsing Rules

In order to create MAS estimates, a simulation using existing data was conducted to determine which questionnaires would have been collected had a MAS design been used. The O*NET program is primarily a mail survey (questionnaires are mailed to potential respondents at their place of employment). Because of this design, a lag exists between selection and response. Therefore, the stopping of a quota cell must be based on the projected number of respondents from those selected. Thus, the date a potential respondent was selected was used as the basis for inclusion in the MAS estimate, instead of the date a questionnaire was returned. In other words, the simulation was performed by ordering questionnaires according to the date they were mailed. Respondents were included chronologically, and cumulative tally counts were generated by occupational domain, region, business size (number of employees), and industry division.

Under the simulation, stopping rules were created to determine when a quota cell should be stopped. Moreover, minimum quotas for each cell were set, in case the targeted quota could not be achieved. Because it was not known whether the choice in stopping rule, minimum quota level, and the manner by which the collapsed industry cell was created would affect the MAS estimates, a sensitivity analysis was incorporated into the study evaluation. For each rule, two criteria were defined. The combination of these criteria gives a total of eight stopping rules. Table 2 outlines the criteria used to define the eight different rules by which the simulation was conducted.

Under MAS, establishments and employees are selected under the same procedures currently being used in the traditional paradigm. The first point at which MAS differs from the current design is after a questionnaire is mailed to an employee. Thus, the purpose of the simulation was to determine which questionnaires would have been collected had a MAS design been in place. The stopping rules were used to determine when to stop the simulation

Table 2. Rules Used in Sensitivity Analysis

Minimum Quota Rules	Stopping Quota Cell Rules	Collapsing Quota Rules (Industry Class Only)
1. 5 completed questionnaires in the cell.	1. Stop cell if projected no. of completed questionnaires exceeds the quota plus 5.	1. Collapse cell if quota is less than 10.
2. 5 completed questionnaires allocation based on OES distribution is less than 25; 10 completed questionnaires otherwise.	2. Stop cell if projected no. of completed questionnaires exceeds the quota plus 10.	2. Collapse cell if quota is less than 15.

for a particular quota cell. Because MAS has a marginal design, if a stop rule was met for a cell, then all remaining completed questionnaires from that cell would not be included, even if they were needed to fill cells in the other two classes. The simulation was complete if 20 questionnaires were collected in each domain and the minimum cell counts were met for all quota cells.

Once the MAS respondents were determined, point estimates were created for all the items being analyzed. In order to help minimize potential bias, a poststratification weight based on OES information was applied. This process was conducted for each of the eight stopping or collapsing rules.

2.4 Analysis

For each stopping or collapsing rule, we used two statistical methods to compare the simulated MAS estimates to the published traditional estimates. For MAS-to-traditional comparisons, analyses were performed on three different item types: means of 5-point and 7-point scales, and estimates of proportions. Additional analyses were also performed by the occupation's education-level category to verify that MAS was not biased for particular occupation types. Two education-level categories were created: less than bachelor's degree, including vocational degree, and bachelor's degree or above required.

Substantive confidence bands were the primary tools used to compare simulated MAS estimates with traditional estimates. Based on O*NET research findings, the variation around 5-point item estimates is approximately 0.5 to 1.0 scale points, whereas variation around 7-point item estimates is approximately 1.0 to 1.5 scale points (Mumford, Peterson, & Childs, 1997). In other words, the population estimate is within one point or 1.5 points of the traditional estimate for 5-point and 7-point scale items, respectively. We concluded that using substantive limits for 5-point and 7-point items to compare the MAS estimates with the traditional estimates was more meaningful than using statistical confidence intervals.

Thus, we define substantive confidence limits in the following manner: For 5-point and 7-point scale items, define μ_M as the mean item by occupation value under the MAS process, and $\hat{\mu}_M$ as its corresponding estimate. Similarly define μ_T as the item-by-occupation mean under the traditional approach, with $\hat{\mu}_T$ as its corresponding estimate. Define

$$\hat{\mu}_T \pm 1 \text{ and } \hat{\mu}_T \pm 1.5$$

as substantive confidence limits for 5-point and 7-point scale items, respectively. If $\hat{\mu}_M$ fell outside the substantive limit, then the MAS estimate was substantively different from the traditional estimate.

On the basis of a review of the literature, for estimates of proportions no substantive limit was known; therefore, we used statistical confidence bands to determine a statistically significant difference between MAS and traditional estimates. In order to standardize this difference for all estimates, we used the mean sample size, $\bar{n}$, for each item when we calculated the half width of a 95% confidence interval, as if all estimates were based on a sample size of $\bar{n}$. Thus, the confidence limit for estimates of this type was calculated by the following formula:

$$\hat{p}_T \pm z_{0.025}\sqrt{\frac{\hat{p}_T(1-\hat{p}_T)}{\bar{n}}},$$

where $\hat{p}_T$ is the estimated proportion under the traditional sampling design.

In addition to confidence limits, effect sizes were computed for each occupation and item. For 5-point and 7-point scale items, the effect size was defined as

$$d = \frac{|\hat{\mu}_M - \hat{\mu}_T|}{\hat{\sigma}_T}.$$

For estimates of proportions, we used the chi-square equivalent to calculate the effect size as described by Cohen (1988). The effect size standardizes the difference between the two means, using the standard deviation estimated under the traditional design. We compared the effect sizes to a standard normal distribution and determined the percentage of items falling outside its inter-quartile range (IQR) of a standard normal distribution. A small percentage of estimates falling outside the IQR would indicate that the traditional estimates and the MAS estimates were similar.

3. Results

3.1 Sensitivity Analysis

Results from comparing each of the eight quota stopping or collapsing rules yielded no significant differences. For 5-point items, the percentage of items that fell outside the 1-point substantive band did not differ between methods by more than 0.5%. Similarly, the percentage of estimates that fell outside the IQR was never more than 0.4% different. In addition, the results for the 7-point items and the estimates of proportions never deviated by more than 0.5% for any two sets of rules. Therefore, it was determined that the choice in stopping rule, minimum quota rule, and collapsing rule did not bias the results produced under MAS. Thus, the most flexible rule was selected, which set a minimum quota of 5, allowed quota cells to exceed the targeted quota by 10 questionnaires, and provided that industry cells be collapsed into one cell if their quota was less than 15.

3.2 Substantive Limits, Statistical Confidence Bands, and Effect Sizes

Overall there were not significant differences between estimates generated by each method. For 5-point items, 99.84% of items fell within the 1-point substantive band. For 7-point items, 99.58% of estimates fell within the 1.5-point substantive band. Figure 1 illustrates how almost all occupation-by-item data points fall within substantive bands for 5-point and 7-point items. Similar results for 5-point and 7-point items were found in the analysis of effect sizes. In this analysis 97.93% of 5-point items and 97.44% of 7-point items fell within the IQR when compared to the traditional estimates. These results suggest no statistical difference between the two methods for 5-point and 7-point items. For estimates of proportions, 88.7% of estimates fell within the statistical confidence intervals, and 89.22% of estimates fell within the IQR when compared to the traditional estimates.

3.3 Impact on Burden

Under the traditional paradigm, the 79 occupations in the analysis produced 15,871 completed questionnaires. However, under MAS these occupations produced only 6,583 completed questionnaires. Table 3 illustrates the amount of employee burden saved because of MAS. This table indicates that the number of burden hours expended by respondents would decrease by more than 50%. Thus, MAS would reduce the burden hours and associated cost for future occupations studied in the O*NET program.

4. Discussion

Similar to the goal of the other hybrid designs discussed in the introduction, the intent of MAS (as implemented in this paper) was to retain as many of the probabilistic features underlying the traditional sampling paradigm as possible while incorporating quota cells to minimize any bias induced by the cutoff sampling rules. MAS departs from the traditional paradigm in two key areas. First, once the randomly selected sample was released to the field, interviewers proceeded to fill quota cells defined by the MAS model. As quotas were achieved for some cells, interviewing shifted to other cells until the specified criteria

Figure 1. Substantive Confidence Bands for 5-Point and 7-Point Items

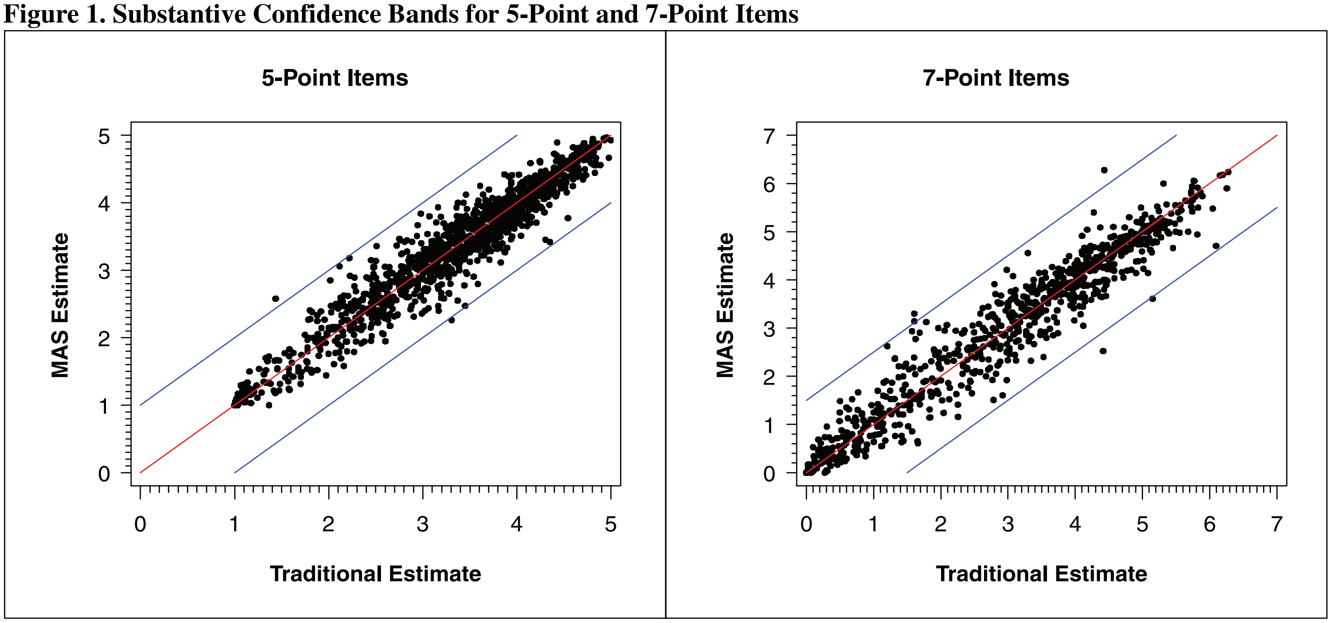

Table 3. Impact to Employee Burden Due to MAS for 79 Analyzed Occupations

A. Estimated burden hours per responding employee	0.5
B. Number of completed questionnaires under traditional paradigm	15,871
C. Burden hours under traditional Paradigm (A * B)	7,935.5
D. Number of completed questionnaires under MAS Paradigm	6,583
E. Burden hours under MAS paradigm (A* D)	3,291.5
F. Burden saved under MAS (C – E)	4,644
E. Change in burden (E/C – 1) * 100	–58.5%

were met for all cells. At that point, interviewing was terminated on all outstanding samples that had not yet been contacted. Second, the survey estimates were not weighted for the selection probabilities. But, as Smith (1983) recommended, poststratification weights were applied. The other areas of the sample design, such as the way establishments and employees were selected, and the way interviewers were to contact establishments, followed a traditional paradigm design.

Like the earlier studies, our analysis suggests that MAS produces estimates comparable to the traditional design currently employed. MAS did not substantively alter the estimates across all occupations and questionnaire items. Under each measurement scale type, the MAS estimates were consistently in agreement with the traditional estimates. Moreover, our sensitivity analysis indicates that our choice of criteria regarding quota cell fulfillment does not bias the estimates, as evidenced by their agreement with traditional estimates. Furthermore, as in most establishment surveys (see, e.g., Knaub, n.d.), the O*NET data exhibit a tendency to be skewed toward smaller establishments (i.e., many more small establishments—those with fewer employees—respond to the survey than larger establishments). MAS is designed to control the number of survey respondents by establishment size and minimize the bias that may be created by this inherent skewness in the size distribution of responding establishments.

As Sudman (1966) and Stephenson (1979) state, there is no theoretical argument for suggesting that hybrid approaches, such as MAS, will always fare as well as the traditional estimates. There are only empirical arguments based on empirical experiments or simulations like the one we conducted. We believe that our simulation performed well because we were able to accurately define a model for each occupation. In addition, we agree with King (1998) that if we had been unable to specify a correct model, our MAS results would not have been as close as they were to the traditional estimates. This qualification suggests that MAS may not be an effective design for an initial data collection study where there is little prior information about the target population. MAS may be effective in update studies that are collecting data on a target population a second time and can use the information collected in the first study to assist in the model definitions.

Also, in studies where the population of interest is difficult to identify in the general population, the use of model-based designs such as MAS can help ensure that survey estimates are representative and include members from all areas that are necessary to fully describe the population of interest. The O*NET Data Collection Program uses MAS to ensure that each occupation has respondents from all industries and all sizes of establishment that appropriately represent the occupation. Furthermore, MAS can help ensure that these respondents come from the entire country and not just one region.

5. Conclusions

Our simulation suggests that our MAS approach does not significantly bias the estimates as compared to a traditional design. Moreover, using MAS, we found no evidence of a bias in the estimates of the standard errors. In other words, both the estimates and confidence intervals for these estimates are not significantly different under MAS than under the traditional paradigm. MAS substantially reduced establishments' burden of providing many more responses than are required for some occupations. MAS does not appear to negatively impact the O*NET program's ability to reliably produce data for users, and it obtains those data more cost-efficiently than traditional designs.

We emphasize that one cannot assume these findings apply to all large-scale surveys. General surveys without the issues found on the O*NET survey, such as sampling a large number of subpopulations, will not benefit from MAS more than from the traditional paradigm. Furthermore, before the implementation of the MAS strategy, research and testing must be conducted to determine whether the strategy is appropriate.

Because of these findings, the O*NET program has incorporated some features of MAS for its second iteration of data collection. Specifically, before data collection a model is defined for each occupation, based on experience gained during the initial data collection period. These models are used to help guide the sample selection process so that the set of respondents for each occupation is representative. MAS cells are stopped when it is clear that the quota will be met; however, traditional probability-based weighted estimates are still produced, and respondent weights are adjusted to account for any stopped cells. This hybrid method incorporates the theoretical strengths of the traditional method, while including steps to ensure a representative respondent sample.

References

Chang, K.-C., Liu, J.-F., & Han, C.-P. (1998). Multiple inverse sampling in post-stratification. *Journal of Statistical Planning and Inference, 69,* 209–227.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Deming, E. W. (1960). *Sample design in business research*. New York: Wiley.

Deville, J.-C. (1991). A theory of quota surveys. *Survey Methodology, 17,* 163–181.

King, B. (1985). Surveys combining probability and quota methods of sampling. *Journal of the American Statistical Association, 80*(392), 890–896.

King, B. (1998). Quota, representative, and other methods of purposive sampling. In *The encyclopedia of biostatistics* (Vol. 5, pp. 3656–3659). New York: Wiley.

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Knaub, J. R. (n.d.). *Model-based sampling, inference and imputation.* Retrieved September 1, 2006, from http://tonto.eia.doe.gov/FTPROOT/electricity/mbsii.pdf#search=%22model-based%20sampling%2C%20inference%2C%20imputation%22

Lohr, S. L. (1999). *Sampling: Design and analysis*. New York: Duxbury Press.

Moser, C. A. (1952). Quota sampling. *Journal of the Royal Statistical Society, Series A, 115,* 411–423.

Moser, C. A., & Stuart, A. (1953). An experimental study of quota sampling. *Journal of the Royal Statistical Society, Series A, 116,* 349–405.

Mumford, M. D., Peterson, N. G., & Childs, R. A. (1997). *Basic and cross-functional skills: Evidence for the reliability and validity of the measures* (O*NET Final Technical Report Contract No. 94-552).

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society, 97,* 558–622.

Sanzo, J. M., Garcia-Calabuig, M. A., Audicana, A., & Dehesa, V. (1993). Q fever: Prevalence and antibodies to *Coxiella burnetii* in the Basque country. *International Journal of Epidemiology, 22,* 1183–1188.

Smith, T. M. F. (1983). On the validity of inference from non-random samples. *Journal of the Royal Statistical Society, Pt. 4, 146,* 394–403.

Stephan, F. F., & McCarthy, P. J. (1979). *Sampling opinions*. New York: Wiley.

Stephenson, B. C. (1979). Probability sampling with quotas: An experiment. *Public Opinion Quarterly, 43,* 477–495.

Sudman, S. (1966). Probability sampling with quotas. *Journal of the American Statistical Association, 61,* 749–771.