METHODS REPORT

Using a Model-Aided Sampling Paradigm Instead of a Traditional Sampling Paradigm in a Nationally Representative Establishment Survey

Marcus Berzofsky, Brandon Welch, Rick Williams, and Paul Biemer

February 2008

RTI Press

RTI INTERNATIONAL

About the Authors

Marcus Berzofsky, MS, is a statistician at RTI International in Research Triangle Park, NC.

Brandon Welchm, MS, was formerly a statistician at RTI; he is now at Rho.

Rick Williams, PhD, is a principal scientist and statistician at RTI International in Research Triangle Park, NC.

Paul Biemer, PhD, is a Distinguished Fellow in Statistics at RTI International in Research Triangle Park, NC.

RTI Press publication MR-0004-0802

This PDF document was made available from **www.rti.org** as a public service of RTI International. More information about RTI Press can be found at **http://www.rti.org/rtipress**.

RTI International is an independent, nonprofit research organization dedicated to improving the human condition by turning knowledge into practice. The RTI Press mission is to disseminate information about RTI research, analytic tools, and technical expertise to a national and international audience.

Suggested Citation

Berzofsky, M., Welch, B., Williams, R., and Biemer, P. (2008). Using a Model-Aided Sampling Paradigm Instead of a Traditional Sampling Paradigm in a Nationally Representative Establishment Survey. RTI Press publication No. MR-0004-0802. Research Triangle Park, NC: RTI International. Retrieved [date] from http://www.rti.org/rtipress.

This publication is part of the RTI Press Methods Report series.

RTI International
3040 Cornwallis Road
PO Box 12194
Research Triangle Park, NC
27709-2194 USA

Tel: +1.919.541.6000
Fax: +1.919.541.5985
E-mail: rtipress@rti.org
Web site: www.rti.org

doi:10.3768/rtipress.2008.mr.0004.0802

www.rti.org/rtipress

Using a Model-Aided Sampling Paradigm Instead of a Traditional Sampling Paradigm in a Nationally Representative Establishment Survey

Marcus Berzofsky, Brandon Welch, Rick Williams, and Paul Biemer

Abstract

We compare traditional survey inference, which is based on probability sample selection and weighting, with a model-based approach based on sampling quotas and model-based weighting. Compared with the traditional approach, the model-based approach more efficiently controls subgroup sample sizes when a large number of rare subgroups are studied. Using data from a national survey of US businesses, we simulated a model-based paradigm and compared estimates with those under the traditional paradigm. In this study, the findings suggest that the model-based approach offers advantages over the traditional sampling approach; however, a hybrid approach capturing the advantages of both paradigms proved best.

Contents

RTI INTERNATIONAL

Introduction

This methods report presents the results of an empirical experiment comparing the use of two sampling paradigms for the US Department of Labor's Occupational Information Network (O*NET) Data Collection Program. In doing so, we describe the two paradigms and how these two are combined to be applied to the current O*NET sample design, and we also discuss other references in the literature. We then describe the analysis and results of the experiment and how the results of the experiment led to a modification of the O*NET sample design.

Two Sampling Paradigms

Historically, sampling finite populations has been conducted using one of two methods: a probability-based approach or a pure model-based approach (Moser, 1952; Moser and Stuart, 1953). For large, federally funded surveys, the statistical community (Kish, 1965) deemed the probability-based approach, as defined by Neyman (1934), the superior of the two methods. 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, which we suggest calling model-aided sampling (MAS), combines traditional probability sampling with quota sampling and is a type of model-based sampling. MAS is an inferential tool and is separate from model-assisted sampling, described by Särndal, Swensson, and Wretman (2003). The design we propose can be highly effective in providing results that allow inference to the general population while controlling costs. We describe the application of MAS to the 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.

We consider quota sampling to be a subset of model-based sampling, and the use of quotas is the essence of MAS. Valliant, Dorfman, and Royall (2000) describe a broader class of model-based finite population sampling than we discuss here.

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 end but accomplish it in very different manners.

The traditional sampling method 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). Random selection is the central tenet of the traditional paradigm and the process by which representativeness and population inference are justified.

Under model-based sampling, a model is used to define the distribution of the target population with respect to the variables of interest (Stephenson, 1979). The model is usually used to determine 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 must be determined 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, the sample has no coverage requirements. If the model assumptions hold, there is no bias in the estimates produced (Deville, 1991). Moreover, the model-based sampling approach 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 approaches. Under the traditional paradigm, rigid controls of field procedures are specified to properly execute the sampling instructions and minimize any interviewer effects on response. 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 enables data collection in a particular quota cell to be stopped once the quota is met. In addition, interviewers are allowed great flexibility in how they collect the data. Moreover, under a quota design, because it is a non-probability-based design, 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 methods also differ in methods for analysis. The traditional design uses randomization so that probability-based weights can be created 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 (1985) 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 Moser and Stuart (1953) define a "standard error" for quota sample designs using resampling methods. Furthermore, although the model-based sampling design does not use probability-based weights, it often incorporates poststratification for making descriptive inferences to a specific population (Smith, 1983).

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

Application to the O*Net Data Collection Program

Sponsored by the US Department of Labor and conducted by the National Center for O*NET Development and RTI International, the O*NET Data Collection Program provides information about a multitude of occupational attributes. The O*NET program is a survey of workers sampled from a sample of businesses that represent all businesses in the nation. It produces estimates for more than 800 occupations in the United States across four occupational domains—the skills required for the occupation, work context (i.e., the conditions in which one's work is completed), the types of work activities conducted on the job, and areas of knowledge (e.g., sales and marketing, or mathematics) that are important for the job. Hence, the O*NET program is simultaneously conducting more than 3,200 surveys.

The O*NET program differs from most large-scale surveys in that at the occupation-by-domain level, the sample sizes are relatively small; however, it targets a large number of subpopulations, which yields a large number of completed questionnaires in aggregate. With limited empirical information, predicting eligibility and response rates for each of these subpopulations is difficult, as is accurately determining the number of workers in each subpopulation to survey in order to obtain the desired number of responses.

Current O*NET data collection began in 2001 and has compiled information from more than 110,000 survey respondents. To date, estimates have been derived under the traditional paradigm for more than 700 of the 810 US occupations targeted by the O*NET program, with an average of 144 questionnaires collected per occupation (median = 117).

For each occupation, respondents complete a questionnaire for one of the four occupational domains—skills, work context, work activities, and knowledge. The goal of the current data collection is to complete at least 15 questionnaires per domain, for a total of 60 completed questionnaires for each occupation. Of the occupations for which data collection has been completed, an average of 36 (median = 29) questionnaires have been collected for each occupational domain. 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. Currently, the sample design incorporates a wave design to control sample overproduction. In a wave design, the sample is released periodically in small, randomly selected increments. Once the desired sample size is achieved, no more increments are released. In the case of the O*NET program, groups of similar occupations are released in waves in an attempt to control the sample size for a given occupation.

Although the current design is effective in identifying persons of interest in aggregate, locating sufficient respondents for each domain in an occupation can be highly variable, depending on the ease with which that occupation is found in the population. Even with the wave design, controlling the sample size across all occupations simultaneously is difficult. This variability causes an inequality in the number of questionnaires collected across occupations.

One constraint on the O*NET program is the number of public burden hours approved by the US Office of Management and Budget. As data collection progressed, some occupations had a higher than desired sample size. 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. To make the best use of the available burden hours, we needed 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. After the initial sampling units are drawn, the traditional sampling approach 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. Similar to general inverse sampling (GIS) (Salehi and Seber, 2004), our approach proposes continuation of the random, multistage design to select employees in the occupations of interest, but also 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 groups for which employees in the occupation are employed.

Furthermore, during data collection a strict protocol is used, as dictated by a traditional sampling paradigm, to identify and contact establishments, 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 based on poststratifying 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.

Other Studies of Model-Aided Sampling

In the 1950s, statisticians treated the two sampling approaches dichotomously and argued the merits of each. Leading proponents of the model-based sampling approach were based in England and led by Moser and Stuart (1953) and Stephan and McCarthy (1979). Proponents of the traditional sampling method 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, in their experiments comparing the traditional paradigm and the model-based paradigm, Moser and Stuart found few major differences in the results. However, Moser and Stuart admit that the lack of theoretical evidence suggests that model-based sampling will always produce estimates as unbiased as those from traditional sampling.

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 (1979) also 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 all cells were completed, the investigators stopped data collection.

Another recent hybrid design, proposed by Chang, Liu, and Han (1998), is multiple inverse sampling (MIS) for finite populations. This design partitions the population into two or more subpopulations with known sizes. MIS is effective when one of these subpopulations is rare and obtaining no or very few responses from the rare subpopulation would be undesirable. MIS sequentially selects sampling units, 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. The final sample size under MIS is a random number, but Liu and Chang (2000) derived formulas to compute the mean and variance of the final random sample size. Furthermore, Salehi and Seber (2001) showed that Murthy's estimator can be used to construct unbiased estimators of the mean and variance in a sequential sampling design. Salehi, Levy, Jamalzadeh, and Chang (2006) extended this method to allow the estimation of the parameters in multiple logistic regression.

Of specific interest to MAS, Salehi and Seber (2004) developed general inverse sampling (GIS). GIS is an adaptive sampling procedure in which the population is divided into predefined quota cells. Then a preliminary sample is drawn across all quota cells following the traditional paradigm. Sampling is completed if the initial sample contains a prespecified number of units in each population cell. Otherwise, a sequential sample is drawn until either the prespecified number of units within each population cell is met or the total sample size reaches a predetermined amount.

Methods

Data

We used data collected for the O*NET program to compare, from 79 occupations, estimates derived under each of the two sampling approaches. 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 77 items across three question types, which were classified by the response option range: a 5-point scale, a 7-point scale, and estimates of proportions (those constructed by dichotomizing "mark all that apply" questions). Our analysis included 16 5-point items, 10 7-point items, and 51 estimates of proportions. Therefore, our analysis consisted of 6,083 occupation-by-item-level estimates.

Quota Definitions

The first step in the MAS design is to determine 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 (Table 1). Each MAS classification category is further split into MAS cells. The industry division class has 11 cells, the

Table 1. Model-assisted sampling quota classifications

Industry division	
	Agricultural, Forestry, and Fishing
	Mining
	Construction
	Manufacturing
	Transportation, Communications, Electric, Gas, and Sanitary Services
	Wholesale Trade
	Retail Trade
	Finance, Insurance, and Real Estate
	Services
	Government (Federal, State, and Local)
	Unknown
Census region	
	Northeast
	South
	Midwest
	West
Number of employees	
	Unknown, 1–9
	10–49
	50–249
	250 or more

census region class has 4 cells, and the number of employees class has 4 cells. 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.

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 US 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, industry cells for an occupation that are small are collapsed into a single cell. These adjustments are made 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 frame. Because the distribution of establishments is skewed toward smaller establishments (i.e., those with fewer 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.

Simulation, Stopping Rules, and Collapsing Rules

To create MAS estimates, we conducted a simulation using existing data to determine which questionnaires would have been collected had we used a MAS design. 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, stopping a quota cell must be based on the projected number of respondents from those selected.

Thus, the date a potential respondent was selected became 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 of stopping rule, minimum quota level, and the manner by which the collapsed industry cell was created would affect the MAS estimates, we incorporated a sensitivity analysis into the study evaluation. For each rule, two criteria were defined. The eight possible combinations of these rules (2 x 2 x 2) formed the stopping rules tested during the analysis. Table 2 outlines the criteria used to define the eight different rules by which the simulation was conducted.

Table 2. Rules used in the 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 when the allocation based on OES 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.

Under MAS, establishments and employees are selected following the same procedures used in the traditional paradigm. The first point at which MAS differs from the traditional 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 for a particular quota cell. Because MAS has a marginal design, if a stop rule was met for a cell, then no remaining completed questionnaires from that cell would 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. We used 20 questionnaires as the termination point for the simulation instead of 15 to simulate the fact that some of the collected questionnaires fail edit checks and, therefore, are not used during estimation and analysis.

Once we determined the MAS respondents, we created point estimates for all the items being analyzed. To help minimize potential bias, we applied a poststratification weight based on OES information. We conducted this process for each of the eight stopping or collapsing rules.

Analysis

For each stopping or collapsing rule, we used two statistical methods to compare the simulated MAS estimates with the published traditional estimates. For MAS-to-traditional comparisons, analyses were performed on three different item types: means of 5-point scales, means of 7-point scales, and estimates of proportions. We performed additional analyses according to 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. Substantive confidence bands are used to determine whether two estimates differ by more than an amount that would be of substantive significance to the outcome. We used these bands are used in lieu of statistical confidence intervals because standard errors cannot be produced, and a statistical confidence interval may indicate a statistical difference even though substantively the two estimates are the same. For the 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, and Childs, 1997). In other words, the population estimate is within 1 point of the traditional estimate for 5-point scale items or within 1.5 points of the traditional estimate for 7-point scale items.

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.

No substantive limit for estimates of proportions was found in the literature; therefore, we used statistical confidence bands to determine a statistically significant difference between MAS and traditional estimates. 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 percent confidence interval, as if all estimates were based on a sample size of $\bar{n}$. We calculated the confidence limit for estimates of this type using the following formula:

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

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

We also computed effect sizes for each occupation and item. For 5-point and 7-point scale items, the effect size was defined as

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

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 interquartile 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.

Results

Sensitivity Analysis

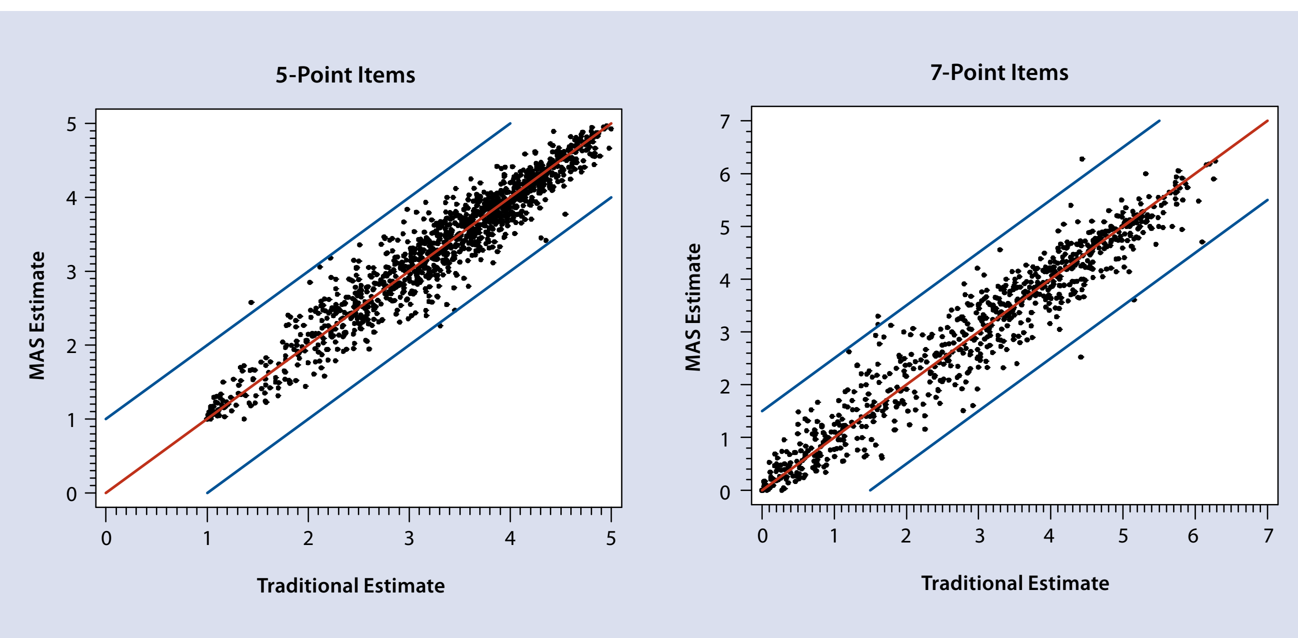
Results from comparing each of the eight quota stopping or collapsing rules yielded no significant differences. For 5-point items, the percentage of items that fell outside the 1-point substantive band did not differ by more than 0.5 percent between methods. Similarly, the percentage of estimates that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

Substantive Limits, Statistical Confidence Bands, and Effect Sizes

Overall there were no significant differences between estimates generated by the two methods. For 5-point items, 99.84 percent of items fell within the 1-point substantive band. For 7 point items, 99.58 percent of estimates fell within the 1.5-point substantive band. Figure 1 contains the 1,264 occupation-by-5-point-item comparisons and the 790 occupation-by-7-point-item comparisons and illustrates how almost all occupation-by-item data points fall within substantive bands for item types.

We found similar results for 5-point and 7-point items in the analysis of effect sizes. In this analysis, 97.93 percent of 5-point items and 97.44 percent of 7-point items fell within the IQR when compared with the traditional estimates. These results suggest that no statistical difference exists between the two methods for 5-point and 7-point items. For estimates of proportions, 88.7 percent of estimates fell within the statistical confidence intervals, and 89.22 percent of estimates fell within the IQR when compared to the traditional estimates.

Figure 1. Substantive confidence bands for 5-point and 7-point items



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 when using MAS. This table indicates that using MAS would decrease the number of burden hours expended by respondents by more than 50 percent. Thus, MAS would reduce the burden hours and associated cost for future occupations studied in the O*NET program.

Table 3. Impact to employee when using model-aided sampling (MAS) to analyze 79 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 x B)	7,935.5
D.	Number of completed questionnaires under MAS paradigm	6,583
E.	Burden hours under MAS paradigm (A x D)	3,291.5
F.	Burden saved under MAS (C – E)	4,644
G.	Change in burden (E/C – 1) x 100	−58.5%

Discussion

Similar to the goal of others in designing hybrid designs discussed in the introduction, our intent in using MAS (as implemented in this study) was to retain as many of the probabilistic features underlying the traditional sampling method as possible while incorporating quota cells to minimize any bias induced by the cutoff sampling rules. MAS departs from the traditional method 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 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. Smith (1983) recommended, however, that we apply poststratification weights. The other areas of the sample design, such as the way we selected establishments and employees and the way interviewers were to contact establishments, followed a traditional 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, 1999), 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) stated, there is no theoretical argument for suggesting that hybrid approaches, such as MAS, will always fare as well as the traditional approach. 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 for which little prior information exists 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 for which 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 to appropriately represent the occupation. Furthermore, MAS can help ensure that these respondents come from the entire country and not just one region.

Conclusions

Our simulation suggests that our MAS approach does not significantly bias the estimates as compared with a traditional design. Moreover, using MAS, we found no evidence of a bias in the estimates of the standard errors. In other words, neither the estimates nor the substantive confidence bands for these estimates are significantly different under MAS than under the traditional approach. MAS substantially reduced establishments' burden of providing many more responses than are required for some occupations. MAS does not appear to negatively affect the O*NET program's ability to produce reliable data for users, and it obtains those data more cost-efficiently than do traditional designs.

We emphasize that one cannot assume that these findings apply to all large-scale surveys. General surveys without the issues found in the O*NET survey, such as sampling a large number of subpopulations, will not benefit more from MAS than from the traditional method. 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, we are incorporating some features of MAS in our second round of data collection for the O*NET program. Specifically, before data collection, we are defining a model for each occupation, based on experience gained during the initial data collection period. In rounds beyond the second round, we will incorporate all prior experience gained to develop an updated MAS model. We will continue to use these models to help guide the sample selection process so that the set of respondents for each occupation is representative. We will stop MAS cells when it is clear that the quota will be met; however, we will still produce traditional probability-based weighted estimates, and we will adjust respondent weights 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., and 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. (1999). Model-based sampling, inference and imputation. Retrieved September 1, 2006, from http://tonto.eia.doe.gov/FTPROOT/electricity/mbsii.pdf

Liu, J.-F., & Chang, K.-C. (2000). A note on multiple inverse sampling. *Journal of Statistical Planning and Inference, 87,* 347-352.

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.* In N. G. Peterson, M. D. Mumford, W. C. Borman, P. R. Jeanneret, E. A. Fleishman, & K. Y. Levin (Eds.), O*NET Final Technical Report, 1. Prepared for Utah Department of Workforce Services and US Department of Labor under 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.

Särndal, C. E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling.* New York: Springer.

Salehi, M. M, Levy, P. S., Jamalzadeh, M. A., & Chang, K.-C. (2006). Adaptation of multiple logistic regression to a multiple inverse sampling design: Application to the Isfahan healthy heart program. *Statistics in Medicine, 25,* 71-85.

Salehi, M. M., & Seber, G. A. F. (2001). A new proof of Murthy's estimator which applies to sequential sampling. *Australian & New Zealand Journal of Statistics, 43*(3), 281-286.

Salehi, M. M., & Seber, G. A. F. (2004). A general inverse sampling scheme and its application to adaptive cluster sampling. *Australian & New Zealand Journal of Statistics, 46*(3), 483-494.

Smith, T. M. F. (1983). On the validity of inference from non-random samples. *Journal of the Royal Statistical Society: Series A., 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.

Valliant, R., Dorfman, A. H. & Royall, R. M. (2000). *Finite population sampling and inference: A prediction approach.* New York: John Wiley.

Acknowledgments and Funding Sources

We wish to thank the National Center for O*NET Development and the Employment and Training Administration, US Department of Labor, for their support of the O*NET Data Collection Program and this research.

RTI International is an independent, nonprofit research organization dedicated to improving the human condition by turning knowledge into practice. RTI offers innovative research and technical solutions to governments and businesses worldwide in the areas of health and pharmaceuticals, education and training, surveys and statistics, advanced technology, democratic governance, economic and social development, energy, and the environment.

The RTI Press complements traditional publication outlets by providing another way for RTI researchers to disseminate the knowledge they generate. This PDF document is offered as a public service of RTI International. More information about RTI Press can be found at www.rti.org/rtipress.