Permanent Random Number Technique to Minimize Response Burden in Repeated Surveys

Darryl Creel, Don Jang, and Daniel Kasprzyk, Mathematica Policy Research, Inc.
Thomas Williams, Department of Defense

Key Words: Permanent Random Number, Negative Coordination, Repeated Survey

Introduction

Survey data collection organizations are all concerned about producing high quality surveys. One way to measure the quality of a survey is the survey's response rate or, conversely, the survey's nonresponse rate. There is a general increasing trend in nonresponse rates for a number of surveys (Atrostic 2001), but this trend is not universal and depends on several factors related to the specific survey, such as country, mode of data collection, sampling procedure, saliency of topic, agency performing the survey, and societal changes (De Heer 1999). The possible trend in increasing nonresponse rates is an area of concern, because survey nonresponse could potentially bias population estimates developed from the survey. Consequently, minimizing nonresponse should be a high priority for survey researchers. While the decision of a sampling unit to participate in a survey is based on a number of factors, many feel the amount of time to complete a survey and the number of times a sampling unit is asked to participate in sample surveys strongly influence that decision. Consequently, survey practitioners strongly support the reduction of "survey respondent burden" to achieve this priority.

In regularly conducted repeated surveys, some sampling units may be selected more than once in a relatively short time span, thereby increasing the sampling units survey response burden and possibly decreasing the unit's willingness to respond.

For the Health Care Survey of Department of Defense Beneficiaries, we investigated this conjecture by reviewing odds ratios for beneficiaries who were selected more than once in a relatively short period of time, e.g., the odds of a sampling unit responding in the current year divided by the odds of a sampling unit responding in some previous year. Because of the importance of minimizing respondent burden, defined in this paper as the selection of a sampling unit on more than one occasion in a relatively short time period, we proposed a permanent random number technique, i.e., a unique random number is assigned to each sampling unit and remains with that unit as long as the sampling unit is on the sampling frame.

We provide below a brief description of the survey on which the permanent random number technique is applied. Then, we discuss the importance of maintaining high response rates and provide evidence that controlling sample selection to reduce respondent burden may contribute to an improved response rate. Finally, we discuss the permanent random number technique as a scheme to reduce response burden.

Health Care Survey of Department of Defense Beneficiaries

The Adult Health Care Survey of Department of Defense Beneficiaries (HCSDB) is the primary tool with which the TRICARE Management Activity (TMA) of the Assistant Secretary of Defense (Health Affairs) monitors the opinions and experiences of military health system (MHS) beneficiaries. Specifically, the HCSDB is designed to monitor satisfaction with, access to, and use of care in the MHS. From 1995 to 2001, the HCSDB was an annual mail survey. Since 2001, the HCSDB has been conducted quarterly and has used a permanent random number technique to reduce respondent burden.

The survey's target population consists of adults[1], i.e., active duty military, spouses of military members, retired military, and spouses of retired military, who are eligible for military health benefits. The sampling methodology used is Stratified Simple Random Sampling Without Replacement. MPR designs the sample, selects the sample, implements the weighting procedures, and implements the estimation procedures. Defense Manpower Data Center maintains the sampling frame that is created from the Defense Enrollment Eligibility Reporting System. The National Research Corporation conducts the data collection.

Response Rates

Our initial concern about response rates stemmed from the fact that we felt selecting the same sampling unit for the HCSDB within a short period of time would make the sampling unit less likely to respond after the first survey. This concern guides this application. Below, we discuss the potential for bias in estimates in the presence of nonresponse, the HCSDB response rates, and, finally, how we might minimize respondent burden in a repeated survey.

Why are response rates important? If you have a high nonresponse rate, i.e., a low response rate, then a potential for bias in the population estimates exists. For example, assume we select a simple random sample from a population that consists of two unknown groups: respondents and nonrespondents. Let $\mu = \frac{N_R}{N}\mu_R + \frac{N_{NR}}{N}\mu_{NR}$ be the population mean, where μ_R is the population mean for respondents, μ_{NR} is the population mean for the nonrespondents, N_R is the number of respondents in the population, N_{NR} is the number of nonrespondents in the population, and $N(= N_R + N_{NR})$ is the number of units in the population. Also, let $\bar{x}_R$ be an unbiased estimator of μ_R, i.e., $E[\bar{x}_R] = \mu_R$. Given this information, we have

$$Bias[\bar{x}_R] = E[\bar{x}_R] - \mu$$

$$Bias[\bar{x}_R] = \mu_R - \frac{N_R}{N}\mu_R - \frac{N_{NR}}{N}\mu_{NR}$$

$$Bias[\bar{x}_R] = \frac{N-N_R}{N}\mu_R - \frac{N_{NR}}{N}\mu_{NR}$$

$$Bias[\bar{x}_R] = \frac{N_{NR}}{N}\mu_R - \frac{N_{NR}}{N}\mu_{NR}$$

$$Bias[\bar{x}_R] = \frac{N_{NR}}{N}(\mu_R - \mu_{NR})$$

For the bias of $\bar{x}_R$ to be small, we must have a small ratio of the number of respondents to the number of units in the population, $\frac{N_{NR}}{N}$, or a small difference between the population mean of respondents and population mean of nonrespondents, $\mu_R - \mu_{NR}$, or both. Unless we have external information about the population mean of nonrespondents, we will not have an estimate of the population mean for the nonrespondents. Consequently, we will not know the magnitude of the difference between the population mean of respondents and the population mean of nonrespondents. Therefore, we focus on the ratio of the number of nonrespondents to the number of units in the population.

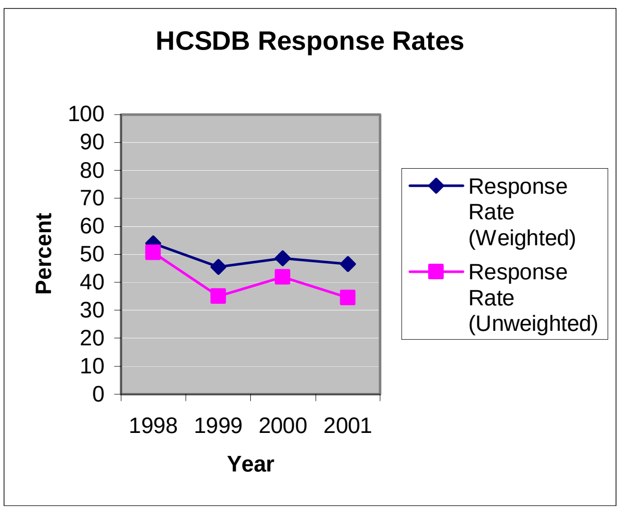
In the annual HCSDB, a number of factors caused sampling units to be selected more than once in a short period of time. These factors include deep stratification,

[1] Within the HCSDB there is also a child survey that is separate from the adult survey.

oversampling for Department of Defense research objectives, and oversampling to account for differential response. Because of these factors and without coordinating the samples, the quarterly HCSDB creates a larger probability for sampling units to be selected more than once in a short period of time.

While investigating the response rates for the HCSDB, we reviewed three questions: What are the response rates for the HCSDB? Were there sampling units selected in consecutive surveys? If a sampling unit was selected in multiple years, did they respond differently among the surveys?

The following graph shows the response rates for HCSDB.



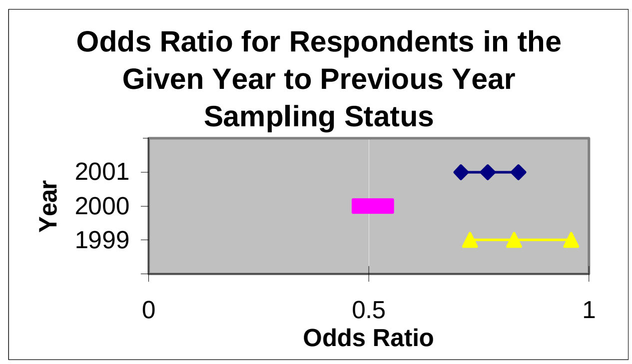
Although the HCSDB is a mail survey and response rates are not particularly high in such surveys, we had hoped the response rates would be higher than actually achieved. Because of the response rates, we were especially concerned with implementing any procedure that would further reduce the response rates.

The following table provides information about the number of sampling units selected in consecutive years of the survey. Sample size is the sample size for the given year. Count is the number of sample units in the given year who were sampled in the previous year. Percent is the count divided by the sample size.

Consecutive Selections

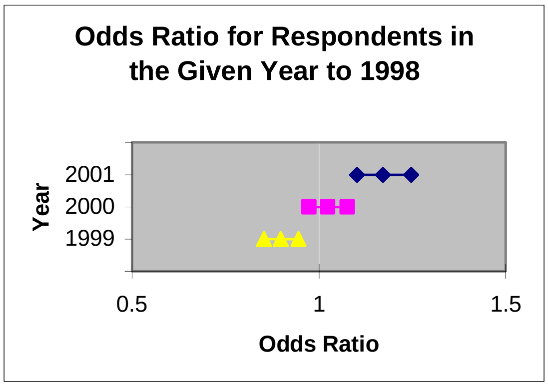
Year	Sample Size	Count	Percent
1999	206,007	7,011	3.4
2000	205,994	10,694	5.2
2001	180,000	8,672	4.8

The absolute number and percent of sampling units sampled in consecutive years is considerable. So, repeated sampling in consecutive years did occur, resulting in an increase in respondent burden. But does being selected in two consecutive years affect a person's willingness to respond? To answer this question, we examined 95 percent confidence intervals (CI) for the odds ratio.



The graph above shows the year of survey on the vertical axis, the odds ratio on the horizontal axis, and 95 percent CI for the odds ratios. The odds ratio is the estimated odds of people responding in the current year who were in previous year's survey to the estimated odds of people responding in the current year who were not in the previous year's survey.

An odds ratio of one, corresponding to independence, serves as a baseline for comparison. All two-year comparisons show a 95 percent confidence interval below one. This indicates that people in the current year were less likely to respond if they were in the previous year's survey than those who were not in the previous year's survey. We see that sample units selected in two consecutive years are less likely to respond in the second year. What happens if a sampling unit is selected in multiple years and those years may or may not be consecutive? The following graph provides the answer.



This graph shows the year of survey on the vertical axis, the odds ratio on the horizontal axis, and 95 percent CI for the odds ratios. The odds ratio is the estimated odds of people responding in the survey year who were in the 1998 survey to the estimated odds of people responding in the survey year who were not in the 1998 survey.

An odds ratio of one, corresponding to independence, serves as a baseline for comparisons. The estimated odds of people responding in 1999 who were in the 1998 sample to the estimated odds of people responding in 1999 who were not in the 1998 sample has a 95 percent confidence interval below one. This indicates that people in the 1999 survey were less likely to respond if they were in the 1998 survey than those who were not. This odds ratio was also presented in the previous graph. The estimated odds of people responding to the 2000 survey who were in the 1998 sample to the estimated odds of people responding to the 2000 survey who were not in the 1998 sample has a 95 percent confidence interval that contains one. This indicates that people in the 1999 survey were as likely to respond if they were in the 1998 survey than those who were not. The estimated odds of people responding in the 2001 survey who were in the 1998 sample to the estimated odds of people responding in the 2001 survey who were not in the 1998 sample has a 95 percent confidence interval above one. This indicates that people in the 2001 survey were more likely to respond if they were in the 1998 survey than those who were not. Consequently, a sampling unit selected in consecutive years responds at a lower rate than sampling units who were not selected in consecutive years. Therefore, we would like to control the sampling to ensure we do not select a sample unit in consecutive years.

Permanent Random Number Technique

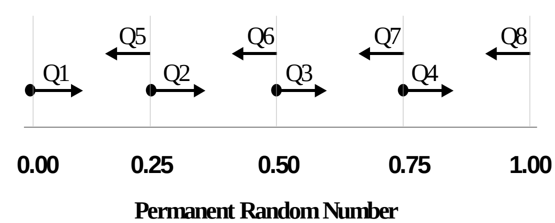
Why are we using the permanent random number technique? The primary reason is our longstanding concern with response rates and our desire to not increase respondent burden, a possible contributor to higher nonresponse rates. The goal is to have two consecutive years, i.e., eight consecutive quarters, without selecting a sampling unit more than one time.

Information from the surveys conducted is not used to update the sampling frame. Although not numerous, sampling units move into and out of the frame and move from one stratum to another stratum over time. Since independent samples are drawn every quarter, there is the possibility that some sampling units will be selected several times in a relatively short period of time while other sampling units will not be selected at all. In order to avoid this situation, we use a general permanent random number technique called negative coordination to minimize the possibility of overlap between the quarterly samples. In particular, we use sequential simple random sampling without replacement (sequential *srswor*). Describing sequential *srswor*, Ohlsson (1995) writes, "The following approach leads to an efficient algorithm for selecting a *srswor*. Each unit in the list frame is assigned a random number drawn independently from the uniform distribution on the interval [0, 1]. Let X_i denote the random number assigned to the unit *i*. The frame units are sorted in ascending order of the X_i. The sample is composed of the first *n* units in the ordered list. Ohlsson (1992) presents a formal proof that this technique produces an *srswor*. Fan et al. (1962) describes this technique, which they label 'sequential.' Following their lead, I refer to this type of selection as *sequential srswor*." This technique is similar to cumulated representative samples or cumulated representation (Kish 1999).

We implement the algorithm in the following manner. We assign the permanent random numbers as described above. When births occur, we generate permanent random numbers beginning in the generation sequence where the last set of generated random numbers ended. Then the random numbers are checked to ensure that there are no duplicate permanent random numbers.

The design is stratified, so we will only discuss the algorithm for a particular stratum for the two-year period that is our goal for having minimal overlap. The permanent random number (prn) is sorted in ascending order within the stratum. We partition the interval, (0,1) into four intervals. For the first quarter of the first year, beginning with the smallest prn we select *n* units. For the second quarter of the first year, select *n* units starting with the smallest prn that is greater than or equal to 0.25. For the third quarter of the first year, select *n* units starting with the smallest prn that is greater than or equal to 0.50. Similarly, for the fourth quarter of the first year, we select n units starting with the smallest prn that is greater or equal to 0.75. Sort the prn in descending order within the stratum. For the first quarter of the second year, start with the largest prn and select *n* units. For the second quarter of the second year, start with the largest prn that is less than 0.75 and select *n* units. For the third quarter of the second year, start with the largest prn that is less than 0.50 and select *n* units. Finally, for the fourth quarter of the second year, start with the largest prn that is less than 0.25 and select *n* units. The following graphic combines our first and second year sampling procedures and demonstrates our sequential srswor procedure.

Sequential Simple Random Sampling Without Replacement



Final Remarks

We have completed eight quarters of sampling using this method and have not selected a sampling unit more than once. Our implementation of a permanent random number technique has allowed us to use negative coordination to reduce the possibility of selecting a sampling unit in a short period of time and thus minimize respondent burden in this repeated survey.

Previous data from the annual surveys show that response rates of sampling units are lower for sampling units who are drawn into sample for two consecutive years of surveys. Using the permanent random number technique, we have drawn valid probability samples while reducing the likelihood of a sampling unit being selected into sample for two consecutive years.

References

Atrostic, B. K., et al. "Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights." *Journal of Official Statistics*. Vol 17, No. 2, 2001. 209-226.

De Heer, Wim. "International Response Trends: Results of an International Survey." *Journal of Official Statistics*. Vol. 15, No. 2, 1999. 129-142.

Fan, C. T., M. E. Muller, and I. Rezucha. "Development of Sampling Plans by Using Sequential (Item by Item) Techniques and Digital Computers." *Journal of the American Statistical Association*. Vol. 80, 1962. 387-402.

Kish, Leslie. "Space/Time Variations and Rolling Samples." *Journal of Official Statistics*. Vol. 14, No. 1, 1998. 31-46.

Ohlsson, Esbjorn. "SAMU-The System for Co-Ordination of Samples from the Business Register at Statistics Sweden-A Methodological Description." R&D Report 1992: 18, Stockholm: Statistics Sweden.

Ohlsson, Esbjorn. "Coordination of Samples Using Permanent Random Numbers." *Business Survey Methods*. Eds. Cox, Brenda, et al. New York : John Wiley & Sons, Inc., 1995. 153-169.