

Using Linked Survey Paradata to Improve Sampling Strategies in the Medical Expenditure Panel Survey

Lisa B. Mirel¹ and Sadeq R. Chowdhury²

Using paradata from a prior survey that is linked to a new survey can help a survey organization develop more effective sampling strategies. One example of this type of linkage or subsampling is between the National Health Interview Survey (NHIS) and the Medical Expenditure Panel Survey (MEPS). MEPS is a nationally representative sample of the U.S. civilian, noninstitutionalized population based on a complex multi-stage sample design. Each year a new sample is drawn as a subsample of households from the prior year's NHIS. The main objective of this article is to examine how paradata from a prior survey can be used in developing a sampling scheme in a subsequent survey. A framework for optimal allocation of the sample in substrata formed for this purpose is presented and evaluated for the relative effectiveness of alternative substratification schemes. The framework is applied, using real MEPS data, to illustrate how utilizing paradata from the linked survey offers the possibility of making improvements to the sampling scheme for the subsequent survey. The improvements aim to reduce the data collection costs while maintaining or increasing effective responding sample sizes and response rates for a harder to reach population.

Key words: Sampling; response propensity; paradata; Medical Expenditure Panel Survey; National Health Interview Survey; interviewer observations.

1. Introduction

Costs of conducting surveys are increasing along with a growing reluctance among respondents to participate in surveys. Survey statisticians are exploring innovative ways to improve data collection efforts while minimizing costs through the use of paradata in adaptive/responsive design frameworks. In the 1940's, Hansen and Hurwitz first introduced concepts similar to adaptive/responsive design sampling schemes (Hansen and Hurwitz 1946). Groves and Heeringa (2006) defined responsive design and discussed the use of paradata to develop responsive designs to control survey costs, nonresponse, and improve the precision of the survey estimates. In recent years paradata are increasingly

¹ Office of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention, 3311 Toledo Road, Hyattsville, MD 20782, U.S.A. Email: LMirel@cdc.gov

² Center for Financing, Access, and Cost Trends, Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, U.S.A. Email: Sadeq.Chowdhury@ahrq.hhs.gov

Disclaimer and Acknowledgments: The views expressed in this article are those of the authors and no official endorsement by the Department of Health and Human Services, the Agency for Healthcare Research and Quality (AHRQ), or the Centers for Disease Control and Prevention (CDC). The authors would like to thank Steve Machlin, Sam Zuvekas, Fred Rohde, and Joel Cohen for their helpful feedback and comments on the topics discussed in this article. In addition, the authors would like to thank the editors of this special edition for their thoughtful and valuable comments. This research was conducted while Lisa B. Mirel was employed at AHRQ prior to becoming an employee at CDC.

being used for that purpose (Durrant et al. 2014; Durrant et al. 2015; Kreuter 2013; Wagner 2013; Groves et al. 2009).

Using paradata from a larger survey that is used for subsampling or linked to the sampling frame of a new survey, can also help develop tailored design sampling strategies to reduce data collection effort. One example of how this is being done is in the Medical Expenditure Panel Survey Household Component (MEPS-HC). MEPS-HC is the main component of MEPS and will be referred to as MEPS hereafter. MEPS, administered by the Agency for Healthcare Research and Quality, is a complex, multi-stage, nationally representative sample of the U.S. civilian, noninstitutionalized population. Each year a sample is drawn for MEPS as a subsample of responding households from the prior year's National Health Interview Survey (NHIS). The linkage or connection of these surveys offers a unique opportunity to use paradata from NHIS to inform sampling strategies in MEPS. One paradata variable, whether the NHIS interview was complete or partially complete, is associated with response propensity in MEPS and is currently being used for forming sampling subdomains or substrata in MEPS. Previous research has explored disproportionate sampling as a way to lower data collection costs (Barron et al. 2015). MEPS has implemented a similar strategy where the sample is drawn at different rates in different substrata based on response propensity as a way to reduce the data collection effort while increasing the unweighted and, potentially, the weighted response rates. The main objective of this article is to examine how paradata from a prior survey can be used in developing a sampling scheme in a subsequent survey. We illustrate how innovative methods can reduce the data collection costs without affecting the precision of the survey estimates.

The results from this research are applicable to other surveys, particularly those that use information from a larger survey to plan for a subsequent survey. For example, the 2010 National Survey of College Graduates (NSCG) selected a portion of its sample from the 2009 American Community Survey (ACS) respondents who indicated they had a bachelor's degree or higher in any field of study (National Science Foundation 2016). The ACS collects substantial amounts of paradata and, as noted in the National Academies Press book, "through its paradata, the ACS can also inform the subsequent survey process in ways that would improve the efficiency and quality of the data" (National Research Council 2008, 58). As an additional example, the American Time Use Survey (ATUS) uses paradata collected in the Current Population Survey (CPS) to aid in developing sampling strategies. The ATUS is sponsored by the Bureau of Labor Statistics and is conducted by the U.S. Census Bureau to measure how respondents spend their time. In the ATUS, individuals are randomly selected from a subset of households that completed their eighth and final month of interviews for the CPS (Bureau of Labor Statistics 2016). Similarly, these methods could aid in follow up surveys, such as those being proposed by the National Health and Nutrition Examination Survey (NHANES) Longitudinal Study (Centers for Disease Control and Prevention 2016). Paradata were used in an experiment to follow up sample units from the Survey of Consumer Sentiment to "predict the contact and co-operation propensities and at-home patterns of sample units in a new wave" (Luiten and Schouten 2013, 171). While there are drawbacks in subsampling (e.g., two phases of nonresponse and potential increases in design effects from unequal weighting), having the additional information from the prior survey could help target sampling strategies that would ease burdens on both the interviewers and respondents, and it has the potential to reduce costs of data collection.

This article focuses specifically on incorporating NHIS paradata variables into the MEPS sample design and presents an evaluation of the effectiveness of using two paradata variables complete/partial interview status and the interviewers’ assessment of the likelihood of response in a linked survey. To do this, we present a method for optimal allocation of the MEPS sample, using NHIS paradata. We also present an approach to evaluate the relative performance of alternative stratification and allocation schemes in terms of data collection costs, impact on design effect, and the potential for increasing the response rate. Increasing the response rate at the first round of data collection in MEPS has the potential to help the overall response rates with multiple rounds of data collection. Our evaluation is based on real cost and response propensity data collected in earlier rounds of MEPS. The framework presented for sample allocation and evaluation of alternative strategies can be applicable to other surveys in similar situations.

2. Background

MEPS is a nationally representative sample of the U.S. civilian, noninstitutionalized population. It is an annual survey of about 14,000 households and has been conducted continuously since 1996. MEPS is a panel survey, and the annual sample consists of two overlapping panels (Figure 1). A new panel of sample is selected each year. It is followed up for two consecutive years with five rounds of data collection. MEPS is an in-person survey, and the results from the survey provide national estimates on health care use, expenditures, insurance coverage, sources of payment, access to care, and health care quality (Ezzati-Rice et al. 2008).

As mentioned above, the MEPS sample is drawn as a subsample of households that participated in the prior year’s NHIS conducted by the National Center for Health Statistics, Centers for Disease Control and Prevention. The NHIS is a multi-purpose health survey that serves as the principal source of information on the health status and health behaviors of the civilian, noninstitutionalized U.S. population. NHIS is based on a complex, multi-stage sample design with oversampling of Hispanics, blacks and Asians (National Center for Health Statistics 2014). The NHIS complex sample design carries over to the MEPS through the set of NHIS responding households that comprise the frame for MEPS sample selection.

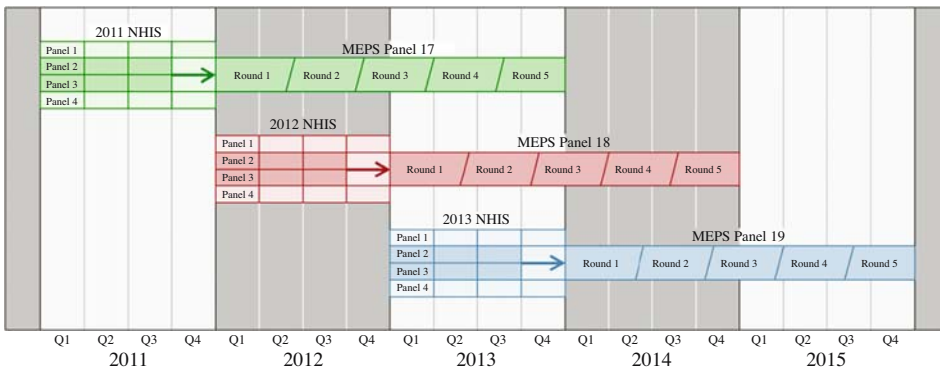


Fig. 1. NHIS-MEPS integrated overlapping panel sample design. Unauthenticated Download Date | 1/22/18 1:24 PM

A disadvantage to this integrated design is that the response rates in MEPS are conditional on response rates in NHIS. More specifically, the overall response rate in MEPS is a compound response rate of NHIS and conditional MEPS response rates. For example, the MEPS Panel 19 response rate in Round 1 is 72% conditional on an NHIS response rate of 76%. Hence, the compound response rate for MEPS Round 1 is 55% ($76\% \times 72\%$) which is much lower than the NHIS response rate of 76%. A detailed discussion of the calculation of MEPS response rates can be found in the MEPS public use file documentation ([Agency for Healthcare Research and Quality 2016](#)). The response rate is calculated independently for each round of data collection by considering the eligibility of a responding unit at that round. Multiplying the NHIS response rate with the product of the conditional response rates for each of the previous and current MEPS rounds produces the overall MEPS response rate up to that round.

Despite the disadvantages of the conditional response rates, there are many advantages to the integrated design. For example, the MEPS sampling frame from the NHIS contains a wealth of information collected in NHIS, including demographic and socioeconomic characteristics of responding members. The integration of the two surveys also means that MEPS does not need to screen households. This increases the efficiency of the design by eliminating the need to independently list and screen households and to locate policy-relevant subgroups of the population. Similarly, the very rich frame of auxiliary variables is used for nonresponse adjustments. The linkage of the two surveys also offers the opportunity to link MEPS data with NHIS data for longitudinal analysis. Another advantage is that the paradata collected in NHIS are used to help inform sampling strategies in the MEPS.

Our article examines the use of paradata from NHIS to inform sampling strategies in MEPS. In our study, the focus is on the non-certainty households and how sample allocation strategies within that domain can be improved, using paradata from the NHIS. The reason for this focus is as follows. Minority households are selected with certainty. Based on race and ethnicity information collected in NHIS, minorities are oversampled in MEPS to improve sample sizes for policy-relevant analyses. Minority households are defined as at least one or more people in the household who identify as Hispanic, Asian, or non-Hispanic black and are sampled with certainty for MEPS. Households that are not defined as minority households are classified as non-Hispanic white/other households. These households are the largest sampling domain in MEPS and are sampled at a non-certainty rate that balances the precision requirement of the estimates for this domain and the pre-assigned targeted sample size for allocation. Over the past several years, the overall sampling rate for the non-Hispanic white/other households has been about 61% of the households on the frame (see [Table 1](#)). Consequently, the MEPS sampling strategy can only be improved for non-certainty households that is, non-Hispanic white/other households.

Starting with Panel 16 of MEPS (2011), NHIS paradata have been used to further stratify the non-Hispanic white/other households and to help develop a tailored sampling strategy. A good predictor of response propensity is a paradata variable from NHIS that indicates if the NHIS interview was complete or partially complete. A complete interview means that the household composition, family, sample adult, and sample child (if a child was present) modules were all completed. A partial interview means that at least a

Table 1. Example of Sampling rates used in various sampling domains and subdomains starting with 2011 (Panel 16) of MEPS.

Domain	Sampling rate (%)
Hispanic	100
Asian	100
Non-Hispanic black	100
Non-Hispanic white/other	61*
NHIS Complete	63.2
NHIS Partial	49.2

*This number is a weighted average of the complete and partial sampling rates

sufficient portion of the family module was completed. Table 1 shows the sampling rates used in recent years for different domains and subdomains in MEPS.

Another paradata variable that may correlate with response propensity is the NHIS interviewer's assessment of cooperation. At the end of each NHIS interview the interviewer records an assessment of how likely she/he thinks the respondent would be to respond to a future linked survey. The interviewer can choose:

1. definitely agree to linked survey,
2. probably agree to a linked survey,
3. probably refuse a linked survey, or
4. definitely refuse a linked survey.

This NHIS variable is not currently used in the MEPS sampling but we discuss it here as a possible future enhancement to the MEPS sampling scheme.

Past research suggests that interviewer assessments can be useful for sampling and for assessing respondent burden. While interviewer assessments may not always be perfectly accurate because they are based on judgments by the interviewers which may add measurement error (West 2013; West and Kreuter 2013), they can still provide insight for sampling and estimating this burden. A recent study revealed that interviewer ratings about participation can "correlate with the cooperation rate" (Eckman et al. 2013, 1). One case study, described in Groves and Heeringa (2006) has illustrated the utility of these types of interviewer ratings: "Sample cases that interviewers expected to have low propensities achieved a second-phase response rate of 38.5%; the high propensity stratum, 73.7%" (442). Similarly in a study conducted using the Consumer Expenditure Survey, the researchers incorporated post survey questions about an interviewer's perception about a respondent's willingness to participate in the survey into their conceptual model predicting response burden in their longitudinal survey (Fricker et al. 2014). Utilizing a similar variable collected in NHIS, we examine possible improvements to sampling the non-certainty households in MEPS.

3. Methodology

We use paradata from NHIS and actual outcomes from previous MEPS fieldwork to create two alternative stratifications to form subdomains or substrata for sampling the non-certainty households. We then allocate the sample at different rates depending on the

relative cost and response propensity of a substratum. We allocate the same overall sample size for each of the stratification schemes to compare their cost-effectiveness. Since sampling at different rates in different substrata increases the variance, we try to optimize the allocation in a manner that balances cost, variance, and response rates. Given the integrated design of NHIS and MEPS, there is inherent variation of the base sampling weights for MEPS. All discussion of variability in the article reflects “additional” variation that the MEPS sampling scheme adds on top of the variation of the NHIS base weight.

We present an approach to optimally allocate the sample to minimize the data collection effort while maintaining the efficiency of the estimates. The proposed optimal allocation approach is used to allocate the sample to different substrata within a domain. We then evaluate and compare the cost effectiveness of different stratification options. Cost effectiveness is defined in the following sections.

3.1. Sample Allocation for a Cost-Effective Design

We allocate the sample to substrata by balancing data collection effort, response rate, and the variance of the estimates. The allocation of the sample is done in a two-step process. First, the sample is optimally allocated to minimize the data collection effort and then the sample size is adjusted to control the increase in variance due to the variation in sampling rates.

We use a cost function that incorporates a fixed cost and a variable cost of data collection in each substratum. The average number of contacts is used as a rough indicator for variable cost of data collection, ignoring any variation in unit cost of a contact by region or primary sampling unit. The number of contacts is affected by many factors, including, but not limited to, locating the study participants, willingness of respondents to participate in the survey, and break offs during the survey. Throughout this article, contacts include actual contacts, contact attempts and calls, but we will generally use the term contacts.

The cost function for a domain or a broad stratum can be considered as follows:

$$C = C_o + \sum C_h n_h \quad (1)$$

where C_o is the fixed cost and all other costs that are invariant to subsampling in substratum h , C_h is the average cost for completing each sampled unit in substratum h and n_h is the sample size in substratum h .

The average cost C_h in substratum h can be defined by factoring in the average number of contacts and response rate as follows:

$$C_h = Q_h / R_h = \text{Overall average number of contacts for achieving a response,} \quad (2)$$

with

Q_h = average number of contacts for each selected household including both respondents and nonrespondents,

$R_h = \frac{n_{hr}}{n_h}$ = response rate, where n_{hr} is the number of respondents in substratum h .

Any other perceived or real cost component can be incorporated in deriving C_h or C . For example, any variation in the unit cost of a contact by geography or other factors can also be accounted for by computing an weighted average cost C_h .

In the absence of any attempt to reduce the number of contacts, no sampling substratum is formed and there is no need for any sample allocation. However, for a comparison at the stratum level with a stratified sampling scheme, the sample selected without stratification can be considered in expectation as proportionally allocated (i.e., the same sampling rate) in different substrata. Therefore, if NHIS paradata were not used for the subsampling of non-certainty households, then the sample in an overall draw is expected to be allocated proportionally in substrata as follows:

$$n_h = n * \frac{N_h}{\sum_h N_h} \tag{3}$$

where n is the overall sample size in the domain or the broad stratum, n_h is the expected allocated sample size in substratum h , and N_h is the frame size in substratum h .

To minimize the cost (in our example, number of contacts) for a fixed sample size n , an appropriate substratification can be formed and the sample can be allocated optimally (Neyman 1934) as follows:

$$n_h = n * \frac{N_h S_h / \sqrt{C_h}}{\sum N_h S_h / \sqrt{C_h}} = n * \frac{N_h S_h / \sqrt{Q_h / R_h}}{\sum N_h S_h / \sqrt{Q_h / R_h}} \tag{4}$$

where S_h is the standard deviation of a target variable in substratum h .

Since the interest here is to control the variance increase due to variation in weights for differential allocation or sampling rates, the variation of a target variable in different substrata within a broad stratum will be assumed the same that is, $S_h = S$. In that case, the above expression for optimal allocation will be reduced to:

$$n_h = n * \frac{N_h / \sqrt{C_h}}{\sum N_h / \sqrt{C_h}} = n * \frac{N_h / \sqrt{Q_h / R_h}}{\sum N_h / \sqrt{Q_h / R_h}} \tag{5}$$

Focusing on the objective of reducing cost, the allocation is set to sample more heavily within a substratum that has larger populations and lower costs (Lohr 2009). The above allocation will minimize costs for a fixed sample size n in a domain. However, as the sampling rate varies by substrata the variance in the domain will increase due to variation in weights. To control the variance, the stratum sample size should be adjusted by considering the higher design effect and increase in response rate.

As we deviate from the proportional allocation to the optimum allocation to minimize costs, the variation in base sampling weights (w) will increase the overall design effect ($deff_w$) due to variation in weights as follows (Kish 1965):

$$deff_w = (1 + CV_w^2) \tag{6}$$

where $CV_w = \frac{\sqrt{V(w)}}{\bar{w}}$ is the coefficient of variation of sampling weights across substrata with the variance of weight defined as,

$$V(w) = \sqrt{\frac{\sum_h n_h (w_h - \bar{w})^2}{n}} \tag{7}$$

In our example, the variation of sampling weights is only for selection in MEPS. As noted previously, this is “additional” variation to the NHIS base weight for those participants who are selected into MEPS. For the proportional allocation, since the subsampling rate is the same in all substrata, the $CV_w = 0$ and hence $deff_w = 1$; the effective sample size will remain the same as n , where n is the realized sample size.

On the other hand, under the optimum allocation, the effective sample size will be reduced to $\frac{n}{deff_w}$.

We will consider this loss in the effective sample size when we consider the cost benefit analysis of the proportional and the optimum allocation.

Considering the increased design effect and increase in response rate, the stratum sample size n will be adjusted as follows:

$$n^* = n \frac{R deff_w}{R^*} \quad (8)$$

where n^* is the adjusted sample size, R is the stratum-level response rate with equal sampling rate in the stratum, R^* is the stratum-level response rate under the above allocation and $deff_w$ is the design effect for variation in sampling rate by substrata. The adjusted sample size n^* can now be used in (5) and reallocated to keep the variance fixed.

For appropriate stratification and optimal allocation, the ratio of increase in stratum-level unweighted response rate $\left(\frac{R}{R^*}\right)$ is usually higher than the increase in design effect ($deff_w$) that is, $\frac{R}{R^*} \leq deff_w$ implying $n^* \leq n$. An optimal allocation with appropriate stratification can reduce costs and also increase response rate while keeping the stratum-level variance the same or lower.

In the above adjustment of the stratum sample size, the design effect due to additional variation in the weights due to nonresponse adjustments is not considered. This is partly because the nonresponse adjustment is usually calculated by forming cells across strata or sampling domains with wide variation in base weights. A marginal increase in the variation of a stratum base weight has relatively small impact on the extra variation of the nonresponse adjusted weight within a cell or an estimation domain (Chowdhury and Baskin 2014). However, if necessary, a compensation for additional increase in the variation of the nonresponse adjusted weight can be done in two ways:

- (a) by inflating $deff_w$ slightly in adjusting the stratum sample size in (8) above; and/or
- (b) by reinvesting some of the cost savings into additional attempts during data collection to increase the response rate in the harder-to-reach substrata with lower sampling rates.

To illustrate the gain under the above procedure, let us consider the following example of a stratum of 10,000 households with 7,000 households in Substratum 1 and 3,000 in Substratum 2. Suppose that on average to obtain a response it requires seven contacts per household in Substratum 1 and ten contacts per household in Substratum 2. If we sample 1,000 households under the optimal allocation and adjust to keep the effective responding sample size fixed, the cost savings and the increase in the overall response rates can be seen in Figure 2 for different levels of response rate differences between Substratum 1 and Substratum 2. For a response rate of 75% in substratum 1 and 65% in Substratum 2, the cost savings compared to proportional allocation is about 2%, and the increase in the

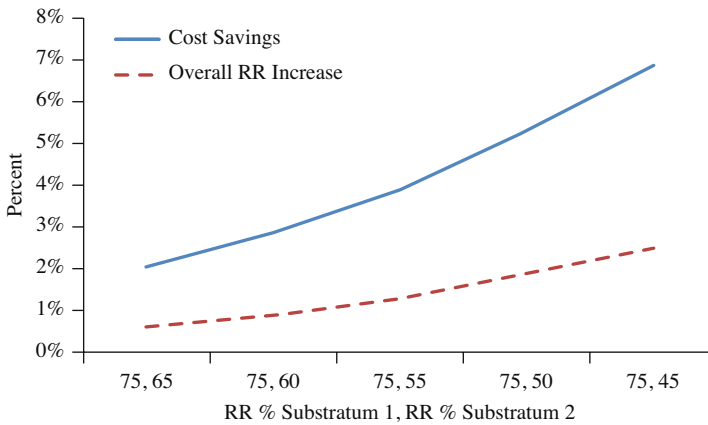


Fig. 2. Cost savings and increase in overall Response Rate (RR) by differences in response rates between substrata.

overall response rate is about 0.8%. For a response rate of 75% in Substratum 1 and 45% in Substratum 2, the increase in the overall response rate is about 2.5% and the cost saving is about 7%.

3.2. Evaluation of Alternative Stratification Schemes

We evaluate the overall savings in terms of the costs and the expected inflation in variance due to the increase in the CV of the weights. We also note the potential for an increase in response rate. To do this evaluation, we compare the cost-effectiveness of alternative schemes using the combined data from MEPS Panels 17 and 18 for the non-certainty sampling domain. We use this as a frame and select a hypothetical sample of size 4,750 households, roughly the usual sample size selected for non-certainty households in MEPS. The mean number of contacts and response rates observed in Panels 17 and 18 will be used for evaluation.

We evaluate two substratification alternatives which differ in how they separate the cases in the non-certainty domain into low and high response propensity. Scheme 1 uses the NHIS paradata variable which captures if the NHIS interview was complete or partially complete. Scheme 2 combines the interview status (complete/partial interview) and the perceived likelihood of response to a future linked survey (definitely respond, probably respond, probably refuse, likely refuse) as assessed by the NHIS interviewer. Since the stratification using the likelihood of response by itself does not offer significant gains compared to the interview completion status, this variable alone is not presented as a stratification scheme. Moreover, some of the categories of the likelihood of response are collapsed because the sample sizes are limited. The blended variable used in Scheme 2 includes three substrata that combine the two paradata variables as shown in Table 4.

The cost-effectiveness of each stratified scheme with optimal allocation is compared with the default scheme. In the default scheme no effort is made to reduce data collection costs, no substrata are formed and there is no varying of sampling rates across substrata. The default scheme without substratification is equivalent to proportional allocation or

equal sampling rate in all substrata as shown in Equation (3). We assumed proportional allocation or equal sampling rate in substrata for the default scheme just for comparison at the substratum level. It has no implication for the overall findings or conclusion of this article as we are not claiming that proportional allocation contributes to any variance improvement or cost reduction.

4. Results

4.1. Stratification Using Complete/Partial Interview Status

Table 2 shows the distribution of the non-certainty sampling domain for MEPS Panels 17 and 18 combined for Scheme 1 that is, by NHIS interview status (i.e., complete or partial) along with the corresponding response rates and mean number of contacts. Those with completed NHIS interviews have a higher response propensity and lower number of contacts on average. The unweighted MEPS response rate was 76.4% for those with a complete interview status based on NHIS compared to 58.5% for those with a partial interview status based on NHIS. The average number of contacts per response in MEPS is much lower (10.38) among NHIS completes compared to the average number of contacts (17.15) for NHIS partials.

Table 3 presents a comparison of sample allocation and the cost-benefit factors between the default scheme and the optimum allocation under Scheme 1. A sample size of 4,750 households was allocated under both schemes. Under the default scheme, the sampling rate is expected to be 61% from both the NHIS interview status of complete and partial substrata while under the stratified scheme with optimal allocation, the sampling rates are 63.2% from the complete substratum and 49.2% from the partial substratum. This difference in sampling rates is due to the higher cost in terms of the number of contacts in the partial substratum, which drives the sample allocation to be lower in the partial substratum and higher in the complete substratum.

As a result, the cost (number of contacts) is expected to decrease by 0.75% under the stratified sampling with optimal allocation. Similarly, since the response rate is lower in the partial substratum, the overall response rate is expected to be 74.2% (3,525 respondents) under the stratified sampling compared to 73.7% (3,500) under the non-stratified sampling. There is also a potential for increasing the weighted response rate due to higher concentration of effort to a smaller sample selected from the hard to reach households in the partial substratum. However, due to the increase in variation of weights

Table 2. Sample size, response rate and number of contacts in MEPS Panels 17 and 18 combined for substrata under Scheme 1.

NHIS Interview status (Scheme 1)	Number of households		Response rate (%)	Average contacts per household	Average contacts per complete
	Sampled	Responded			
Complete	6,599	5,042	76.4	7.93	10.38
Partial	1,183	692	58.5	10.03	17.15
Total	7,782	5,734	73.7	8.25	11.22

Table 3. Sample allocation and effectiveness analysis for Scheme 1 that is, NHIS complete/partial interview status.

NHIS Interview status (Scheme 1)	Default scheme: no stratification/ no variation in sampling rate				Stratification with optimum allocation/ variation in sampling rate			
	Sampled	Sampling rate (%)	Responded	Cost (contacts)	Sampled	Sampling rate (%)	Responded	Cost (contacts)
Complete	4,028	61.0	3,078	31,957	4,169	63.2	3,185	33,072
Partial	722	61.0	422	7,245	581	49.2	340	5,834
Total	4,750	61.0	3,500	39,201	4,750	61.0	3,525	38,906
Response rate			73.7%				74.2%	
CV of weights		0				9.0		
Effective sample size*			3,500				3,497	
Cost difference relative to default							+0.50%	-295
Savings (%)							-3	-0.75%
Difference in effective sample sizes								

*Effective sample size = (total number who responded)/(1 + CV^2) (e.g., 3,525/(1 + (0.09^2)))

(9% CV) under the stratified sampling, the effective responding sample size will come down slightly from 3,500 to 3,497. On the other hand, since there is no additional variation in weights under the non-stratified sampling, the effective sample size will remain the same at 3,500. Under both designs the effective responding sample size is almost the same (3,500 and 3,497). No further adjustment is made to the overall sample size under the stratified sampling. Therefore, while the effective sample size remains almost the same under both schemes, the total number of contacts under the stratified sampling comes down, and the response rate goes up slightly. The number of contacts is used as a proxy for cost, decreasing the number of contacts means there will be a decrease in the costs.

4.2. Stratification Using Complete/Partial Status and Likelihood of Response Status

In this section we examine how stratification Scheme 1 can be made even more beneficial by utilizing an additional NHIS paradata variable, the likelihood of response in a subsequent linked survey as assessed by the interviewer in the NHIS, which we refer to as Scheme 2.

Table 4 shows response rates and average number of contacts for the two paradata variables and their cross classification. The last column of the table shows how the groups were collapsed to form three substrata to be used in Scheme 2. The groups were combined based on similarity of cost (number of contacts) and response rate.

Table 5 shows the response rates and cost (number of contacts) for Scheme 2 after the groups are collapsed into three substrata. The response rate ranges from 48.3% in Substratum 3 to 77.2% in Substratum 1 and the average number of contacts per complete ranges from 10.18 in Substratum 1 to 21.85 in Substratum 3. In comparison, the response rates for Scheme 1 are 58.5% for the Partial and 76.4% for the Complete.

Similar to Table 3, Table 6 presents a comparison of the default scheme and the optimal allocation under Scheme 2. The same sample size of 4,750 households was allocated under both scenarios. The sampling rates under the stratified design with optimal allocation are 63.7% in Substratum 1, 52.0% in Substratum 2 and 43.5% in Substratum 3. The mean number of contacts is negatively associated with the sampling rate; therefore, the cost

Table 4. Response rate and cost (number of contacts) by complete/partial interview status and likelihood of response in MEPS Panels 17 and 18 combined.

NHIS Interview status	Likelihood of response	Response rate (%)	Average contacts per household	Average contacts per complete	Scheme 2 substrata
Complete	Definitely agree	79.8	7.66	9.60	1
	Probably agree	73.9	8.13	11.01	1
	Probably refuse	61.7	9.14	14.81	2
Partial	Definitely refuse	55.0	9.40	17.09	3
	Definitely agree	70.3	8.23	11.71	1
	Probably agree	64.3	10.03	15.60	2
	Probably refuse	48.7	10.88	22.32	3
	Definitely refuse	42.9	9.53	22.24	3
Total		73.68			

Table 5. Sample size, response rate, and number of contacts in MEPS Panels 17 and 18 combined for substrata under Scheme 2.

Scheme 2 substrata	Number of households		Response rate (%)	Average contacts per household	Average contacts per complete
	Sampled	Responded			
1	6,358	4,910	77.2	7.86	10.18
2	909	575	63.3	9.67	15.29
3	515	249	48.3	10.57	21.85
Total	7,782	5,734	73.7	8.25	11.20

(number of contacts) is expected to decrease by 1% under the stratified design. Similarly, the overall response rate is expected to be 74.5% (3,537 respondents) under the stratified design. On the one hand, due to the 11% CV of weights arising from the optimal allocation, the effective responding sample size will only come down slightly to 3,495. On the other hand, since there is no additional variation in weights – as the sampling rate is the same in both sampling substrata under the default scheme – the effective sample size remains the same at 3,500. Since the effective responding sample size under the stratified schemes is very close to that of the default scheme, no further adjustment to the overall sample size is made under the stratified scheme.

Table 7 summarizes the findings for Schemes 1 and 2 compared to the default scheme with no stratification or variation in sampling rate. While the impacts of both stratification methods are similar in terms of a negligible decrease in effective sample sizes, the blended stratification appears slightly better in terms of response rate (higher) and overall costs (lower) for obtaining a response.

5. Discussion

In this article, we discuss an approach to improve sampling strategies in MEPS. The approach is based on optimally allocating the sample to substrata formed using paradata from the linked NHIS. We present a method for optimal allocation of the sample to different substrata to minimize the data collection costs for a fixed variance and present an evaluation approach to select the best alternative stratification.

During the last few years, the NHIS interview status (complete/partial) has been used to form the sampling substrata (Scheme 1) to both reduce costs and potentially increase the response rate. Here we explore using an additional paradata variable, likelihood of response in a subsequent survey as assessed by the NHIS interviewers and blend it with the complete/partial interview status variable to form more effective substrata (Scheme 2). The sampling substrata formed by grouping the households with similar response propensities and number of contacts helps to develop a sampling strategy that reduces the data collection effort for a harder to reach population. An evaluation comparing Scheme 1 and Scheme 2 shows that there are slight savings under both methods in terms of reducing costs (number of contacts) and increasing the response rate but Scheme 2 performs slightly better than Scheme 1.

The results illustrate how a tailored sampling scheme with optimum allocation in substrata formed using paradata can help reduce the cost and potentially increase the

Table 6. Sample allocation and effectiveness analysis for Scheme 2, that is, blending of NHIS complete/partial interview status and likelihood of response.

Scheme 2 Substrata	Default scheme: no stratification/ no variation in sampling rate				Stratification with optimum allocation/ variation in sampling rate			
	Sampled	Sampling rate (%)	Responded	Cost (contacts)	Sampled	Sampling rate (%)	Responded	Cost (contacts)
1	3,881	61.0	2,997	30,513	4,053	63.7	3,130	31,868
2	555	61.0	351	5,367	473	52.0	299	4,574
3	314	61.0	152	3,321	224	43.5	108	2,368
Total	4,750	61.0	3,500	39,201	4,750	61.0	3,537	38,809
Response rate			73.7%				74.5%	
CV of weights		0				11.0		
Effective sample size*			3,500				3,495	
Cost difference relative to default								- 392
Savings								+ 0.79%
Difference in effective sample sizes								- 5

*Effective sample size = (total number who responded)/(1 + CV²) (e.g., 3,537/(1 + (0.11²)))

Table 7. Comparison of two alternative optimum allocation sampling procedures compared to the default scheme.

Sampling substrata	Increase in response rate (%)	Reduction in number of contacts' (%)	Difference in effective sample size
Scheme 1: Complete/Partial	0.50	-0.75	-3
Scheme 2: Blended complete/partial and likelihood of response	0.79	-1.00	-5

response rate without reducing the efficiency of the estimates. Our results indicate that, in a typical MEPS panel, the existing and new stratifications (Scheme 1 and Scheme 2) with optimum allocation of sample can reduce the cost by about 0.75% and 1%, respectively. Similarly, the response rate can increase by about 0.50% under the current scheme and 0.79% under the proposed new cross classification scheme. Although the cost savings by using the existing or the new sampling scheme are marginal, they are achieved with virtually no loss in terms of effective sample size. The method could continue to be improved by applying differential sampling rates using additional available paradata or by collecting and utilizing more relevant paradata to offer further gains in cost reduction. The framework presented for allocating the sample and evaluating cost effectiveness of alternative stratification will be useful in other similar surveys.

It should be noted that with the optimal sampling scheme, it is possible that there could be additional variability in the nonresponse adjusted weight because of variation in substrata sampling rates within a stratum. Our analysis did not address this additional variability, in part, because previous research with MEPS data has shown that the differential sampling within a stratum or a domain has little effect on the overall variability of the weights because the nonresponse adjustment to the weights are made by combining cases from all domains with wide variation in base weights (Chowdhury and Baskin 2014). However, if necessary, any likely increase in the variability of the nonresponse adjusted weight can be mitigated by reinvesting some of the cost savings either by increasing the stratum sample size as discussed in the Methodology section or by making additional contact attempts during data collection to increase the response rate in a harder to reach substrata with lower sampling rates.

Although Scheme 2 has not actually been implemented in the field, MEPS plans to utilize additional paradata variables to tailor the sampling in future surveys. The integration between NHIS and MEPS offers a unique opportunity to continually improve sampling strategies in MEPS.

6. References

Agency for Healthcare Research and Quality. 2016. "MEPS HC-171 2014 Full Year Consolidated Data File." Available at: https://meps.ahrq.gov/data_stats/download_data/pufs/h171/h171doc.pdf. (accessed March 2017).

- Barron, M., M. Davern, R. Montgomery, X. Tao, K.M. Wolter, W. Zeng, C. Dorell, and C. Black. 2015. "Using Auxiliary Sample Frame Information for Optimum Sampling of Rare Populations." *Journal of Official Statistics* 31: 545–557. Doi: <http://dx.doi.org/10.1515/JOS-2015-0034>.
- Bureau of Labor Statistics. (2016, June). *ATUS User's Guide (PDF)*. Available at: <http://www.bls.gov/tus/atususersguide.pdf>. (accessed March 2017).
- Centers for Disease Control and Prevention. (2016, May 23). *Federal Register: The Daily Journal of the United States Government*. Available at: <https://www.federalregister.gov/articles/2016/05/23/2016-12008/proposed-data-collection-submitted-for-public-comment-and-recommendations>. (accessed March 2017).
- Chowdhury, S.R. and R.M. Baskin. 2014. "PPS Subsampling from NHIS to MEPS – Effect on Precision of MEPS Estimates." In Proceedings of the Section on Survey Research Methods: American Statistical Association, August 2014, 2339–2351. Alexandria, VA: American Statistical Association (CD-ROM).
- Durrant, G.B., O. Maslovskaya, and P.W.F. Smith. 2014. "Sequence Analysis as a Tool for Investigating Call Record Data." Working paper, University of Southampton. Available at: <https://eprints.soton.ac.uk/369102/> (accessed March 2017).
- Durrant, G.B., O. Maslovskaya, and P.W. Smith. 2015. "Modelling Final Outcome and Length of Call to Improve Efficiency in Call Scheduling." *Journal of Survey Statistics and Methodology* 3: 397–424. Doi: <https://doi.org/10.1093/jssam/smv008>.
- Eckman, S., J. Sinibaldi, and A. Montmann-Hertz. 2013. "Can Interviewers Effectively Rate the Likelihood of Cases to Cooperate?" *Public Opinion Quarterly* 77: 561–573. Doi: <http://dx.doi.org/10.1093/poq/nft012>.
- Ezzati-Rice, T.M., F. Rohde, and J. Greenblatt. 2008. *Sample Design of the Medical Expenditure Panel Survey Household Component, 1998–2007*, Methodology Report No. 22. March 2008. Rockville, MD: Agency for Healthcare Research and Quality. Available at: https://meps.ahrq.gov/data_files/publications/mr22/mr22.pdf (accessed March 2017).
- Fricker, S., T. Yan, and S. Tsai. 2014. "Response Burden: What Predicts it and Who is Burdened Out?" In Proceedings of AAPOR Section: American Statistical Association, August 2014. 4568–4577. Alexandria, VA: American Statistical Association. (CD-ROM).
- Groves, R.M. and S.G. Heeringa. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society, Series A* 169: 439–457. Doi: <http://dx.doi.org/10.1111/j.1467-985X.2006.00423.x>.
- Groves, R.M., M.D. Mosher, J. Lepkowski, and N.G. Kirgis. 2009. *Planning and Development of the Continuous National Survey of Family Growth*. National Center for Health Statistics. Vital Health Stat, 1(48). Available at: https://www.cdc.gov/nchs/data/series/sr_01/sr01_048.PDF (accessed March 2007).
- Hansen, M.H. and W.N. Hurwitz. 1946. "The Problem of Non-Response in Sample Surveys." *Journal of the American Statistical Association* 41: 517–529. Doi: <http://dx.doi.org/10.1080/01621459.1946.10501894>.
- Kreuter, F. 2013. *Improving Surveys with Paradata: Analytic Uses of Process Information*, edited by Frauke Kreuter. Hoboken, NJ: John Wiley & Sons, Inc.

- Kish, L. 1965. *Survey Sampling*. New York: Wiley.
- Lohr, S.L. 2009. *Sampling: Design and Analysis*. Boston: Richard Stratton.
- Luiten, A. and B. Schouten. 2013. "Tailored Fieldwork Design to Increase Representative Household Survey Response: an Experiment in the Survey of Consumer Satisfaction." *Journal of the Royal Statistical Society A* 176: 169–189. Doi: <http://dx.doi.org/10.1111/j.1467-985X.2012.01080.x>.
- National Center for Health Statistics, *National Health Interview Survey, 2014. Public-use data file and documentation*. Available at: http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm (accessed March 2017).
- National Research Council, N. 2008. *Using the American Community Survey for the National Science Foundation's Science and Engineering Workforce Statistics Programs*. Washington, DC: National Academies Press. Doi: <https://doi.org/10.17226/12244>.
- National Science Foundation. 2016. *National Survey of College Graduates*. Available at: <http://www.nsf.gov/statistics/srvygrads/#tabs-1> (accessed March 2017).
- 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: 558–606. Doi: <http://dx.doi.org/10.2307/2342192>.
- Wagner, J. 2013. "Using Paradata-Driven Models to Improve Contact Rates in Telephone and Face-to-Face Surveys." In *Improving Surveys with Paradata: Analytic Use of Process Information*, edited by F. Kreuter, 145–170. New Jersey: John Wiley and Sons.
- West, B.T. 2013. "An Examination of the Quality and Utility of Interviewer Observations in the National Survey of Family Growth." *Journal of the Royal Statistical Society A* 176: 211–225. Doi: <http://dx.doi.org/10.2307/23355184>.
- West, B.T. and F. Kreuter. 2013. "Factors Affecting the Accuracy of Interviewer Observations Evidence from the National Survey of Family Growth." *Public Opinion Quarterly* 77: 522–548. Doi: <https://doi.org/10.1093/poq/nft016>.

Received January 2016

Revised March 2017

Accepted April 2017