An Analysis of Adaptive Sampling Procedures in the National Health Interview Survey

Chandra Erdman and Darcy Morris
Center for Statistical Research and Methodology
U.S. Census Bureau

August 19, 2013

Abstract

Adaptive design is a means to reduce survey costs by taking focus away from maximizing response rate exclusively. An optimal adaptive procedure would minimize cost, while maintaining or improving the quality of the estimates. Through empirical analyses, this work assesses the quality of key estimates from the National Health Interview Survey (NHIS) under various stopping rules. Our results provide insight into potential optimal decision theoretic stopping rules for achieving an appropriate balance between properties of the estimator and survey costs.

This document was prepared for the 24^{th} International Workshop on Household Survey Nonresponse. It represents work in progress – please do not cite or circulate without permission of the authors.

Acknowledgments: This work was completed with the help of a team consisting of staff from the National Center for Health Statistics: Jim Dahlhamer, Renee Gindi, and Chris Moriarity.

1 Background

The National Health Interview Survey collects information on the prevalence and distribution of disease, disability and chronic impairments, and the type and completeness of health services people receive. The survey is sponsored by the Center for Disease Control and Prevention (CDC) and the National Center for Health Statistics (NCHS), and conducted by the U.S. Census Bureau on an ongoing monthly basis. The NHIS sample is designed to produce nationally representative estimates for the civilian, non-institutionalized population living in the United States. Each year, a sample of approximately 64,000 households is selected, with certain minority and elderly groups selected at a higher rate. Specifically, in areas with significant minority population, a number of households are designated as “screeners” and once demographic information is obtained, the household is only interviewed if one or more of its members are either Black, Asian, or Hispanic.

2 Data

In order to provide access to the most recent health statistics from the NHIS, the CDC and the National Center for Health Statistics publish “Early Release” (ER) estimates of fifteen health measures three times each year: (1) lack of health insurance coverage; (2) usual place to go for medical care; (3) cost barrier to medical care; (4) receipt of influenza vaccination (ages 18+); (5) pneumococcal vaccination (ages 65+); (6) obesity (ages 20+); (7) sufficiently leisure-time physical activity (ages 18+); (8) current smokers (ages 18+); (9) 5+ drinks in 1 day (ages 18+); (10) HIV test (ages 18+); (11) excellent or very good health; (12) needs help with personal care (ages 65+); (13) serious psychological distress in the past 30 days (ages 18+); (14) diagnosed diabetes (ages 18+); (15) asthma attack in the past 12 months. Our analyses focus on these key health measures and our data consist of unedited responses to the NHIS from the entire 2012 calendar year, along with base weights to correct for oversampling when computing estimates. The 2012 NHIS sample produced data on 65,626 households, 99,921 adults, and 32,672 children.

In addition to survey responses, we have two data sources from which to estimate response propensities. First, we have complete contact history records from the Contact History Instrument (CHI). The CHI captures basic information such as the time of day, mode (telephone or in-person) and outcome (partial or complete interview, refusal, etc.) of each contact attempt. In addition, each time contact is made with a sample member, the interviewer completes the CHI “reluctance questionnaire” which lists twenty-two categories of respondent reluctance or concerns, and behaviors that may be exhibited during contact. Interviewers may check any number of items in the questionnaire which lists behaviors such as “slams door,” and “hangs up,” statements such as “too busy,” or “not interested,” and a “no concerns” item. Our second data source is the sample frame, which includes the structure type (single family home versus multi-unit) for each address, and an indicator for whether the household is located in an urban or rural area.

3 Methods

3.1 Notation and Definitions

Let

- θ_j be the true value of key variable j, where $j = 1, \ldots, J$,
- $\hat{\theta}_j$ be the estimate of key variable j,
- $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_J]^T$ be the vector of true values of key variables,
- $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \ldots, \hat{\theta}_J]^T$ be the vector of estimates of key variables,
- $Q(\hat{\theta}_j)$ be the quality of key estimate j,
- $Q(\hat{\boldsymbol{\theta}}) = f\left(Q(\hat{\theta}_1), \ldots, Q(\hat{\theta}_J)\right)$ be a function that measures the quality of all key variable estimates combined,

- τ be a decision/stopping rule,
- C be a cost function,
- $R_{it} = \begin{cases} 1 \text{ if an interview is obtained for household } i \text{ at contact t} \\ 0 \text{ otherwise} \end{cases}$,
- $\mathbf{X}_i$ be a set of explanatory variables for household i,
- $p_{i,t} = P(R_{it} = 1 | R_{i1} = 0, \ldots, R_{i,t-1} = 0, X_i)$ be the discrete-time hazard or response propensity for household i at time t, given that no interview was obtained at times $t = 1, 2, \ldots, t-1$.

In adaptive design, we have one of two dual goals: (1) $\max_\tau \; Q(\hat{\boldsymbol{\theta}})$ subject to $C = c$; (2) $\min_\tau \; C$ subject to $Q(\hat{\boldsymbol{\theta}}) = q$. Or, alternatively, $\max_\tau \; [Q(\hat{\boldsymbol{\theta}}) - C]$. We want to find the decision/stopping rule, τ, that maximizes quality and minimizes cost. In decision theory terminology, $-[Q(\hat{\boldsymbol{\theta}}) - C()]$ is the loss function to be minimized. The key components to these optimization problems are:

- **Statistical Quality Measure**, $Q(\hat{\theta}_j)$: How "good" is the estimate of the key variable?
- **Aggregate Statistical Quality Measure**, $Q(\hat{\boldsymbol{\theta}})$: How should we combine the properties of the key estimates into one quality score? Some options include importance weighting (do we care about certain measures more than others, e.g. percent uninsured more than percent obese?) or sensitivity weighting (how sensitive is the j^{th} estimate to sample size, the stopping rule, etc.?).
- **Response Propensity**, p_{it}: Response propensity is important in defining stopping rules and in calculating expected cost.
- **Cost**, C: In the simplest setting, cost is a function of the number of contact attempts only. Additionally, losses in the number or proportion of households providing an interview may be incorporated into C.

Through empirical analyses, we examine the following four classes of stopping policies.

1. Stop after $4, 5, \ldots, 20$ contact attempts,
2. Stop after $14, 15, \ldots, 30$ days,
3. Stop after 1, 2, 3, or 4 refusals
   (a) of any type,
   (b) "hard refusal" which includes only the following reluctance categories: "does not want to be bothered," "hangs up or slams door," and "hostile or threatening" (Dahlhamer and Simile, 2009).
4. Stop at contact $t-1$ if the estimate, $\hat{p}_{i,t}$, of $P(R_{i,t} = 1 | R_{i1} = 0, \ldots, R_{i,t-1} = 0, X_i)$ is less than some threshold. We estimate $p_{i,t}$ in two ways:
   (a) Empirical hazard functions computed as the fraction of visits at time t, for cases in group G which have no interview up to time $t-1$, and which are converted to interview at time t. We define the mutually exclusive groups as follows: G_0 : No prior contact; G_1 : Prior contact, no refusals; G_2 : Prior soft refusal(s) only; G_3 : Prior soft and hard refusals; G_4 : Prior hard refusal(s) only, and
   (b) Discrete-time hazard model: $\hat{p}_{i,t} = logit^{-1}(\alpha + \mathbf{X}_i \boldsymbol{\beta})$, where α and $\boldsymbol{\beta}$ are estimated via logistic regression.

3.2 Naïve Empirical Study of Stopping Rules

We first examine the estimates under each rule by a simple method for comparing proportions of two independent binomial samples. We use "time" interchangeably among the various stopping strategies to describe the indices of each. In strategy (1), time $t \in \{4, 5, \ldots, 20\}$ is contact attempt or visit number, in (2) we examine time $t \in \{14, 15, \ldots, 30\}$ days, and in (3) time $t \in \{1, 2, 3, 4\}$ is used to denote the visit during which the t^{th} refusal is observed. Let $\hat{\theta}_{j,1:t-1}$ be the estimate of health measure j computed from the $n_{j,1:t-1}$ people who respond before time t, and $\hat{\theta}_{j,t}$ be the estimate computed from the $n_{j,t}$ people who respond at time t. The method proposed by Rao et al. (2008) recommends stopping at time t if

$$z_t = \left| \frac{\hat{\theta}_{j,1:t-1} - \hat{\theta}_{j,t}}{\sqrt{\frac{\hat{\theta}_{j,1:t-1}(1-\hat{\theta}_{j,1:t-1})}{n_{j,1:t-1}} + \frac{\hat{\theta}_{j,t}(1-\hat{\theta}_{j,t})}{n_{j,t}}}} \right| \leq z_{\alpha/2},$$

where $z_{\alpha/2}$ is the $1 - \frac{\alpha}{2}$ percentile of the standard normal distribution and α is pre-specified.

In addition, we calculate and assess the estimates of each health measure using the statistics and quality measures enumerated in the following list.

1. Treat the estimates from the full NHIS data as the truth, $\boldsymbol{\theta}$. Specifically, each component of $\boldsymbol{\theta}$ is computed as $\hat{\theta}_j = \sum_i w_i \mathbb{1}_{ij}/n_j$ where $\mathbb{1}ij$ is the indicator for the characteristic measured by key variable j, n_j is the number of persons who respond to the question that measures variable j, and w_i is the baseweight for record i.
2. ***Estimate*** $\hat{\boldsymbol{\theta}}$ as in (1), but only for households which satisfy the given stopping rule, and compute the individual quality measures:
3. ***Bias***: $Q_1(\hat{\theta}_j) = (\hat{\theta}_j - \theta_j)$, and
4. ***Relative Bias***: $Q_2(\hat{\theta}_j) = Bias(\hat{\theta}_j)/\theta_j$; the aggregate quality measures:
5. ***Total Absolute Bias***: $Q_3(\hat{\boldsymbol{\theta}}) = \sum_j |\hat{\theta}_j - \theta_j|$,
6. ***Average Absolute Relative Bias***: $Q_4(\hat{\boldsymbol{\theta}}) = \frac{1}{J} \sum \frac{|\hat{\theta}_j - \theta_j|}{\theta_j}$, and
7. ***Maximum Absolute Relative Bias***: $Q_5(\hat{\boldsymbol{\theta}}) = \max_j \frac{|\hat{\theta}_j - \theta_j|}{\theta_j}$. Finally, compute
8. ***Cost***: Let PV_τ be the number of personal visit interview attempts made under policy τ, T_τ be the number of interview attempts made by telephone, $-I_\tau$ be the number of interviews lost (i.e., the number of interviews obtained in the full sample minus the number of interviews obtained under the stopping rule), $-S_\tau$ be the number of screener interviews lost under τ, w_S be a weight assigned to lost screener interviews, and N be the total number of households in the sample. For each policy, we examine the cost function

   $$C(k) = \frac{PV_\tau + T_\tau/20 + k[-I_\tau + w_S(-S_\tau)]}{N},$$

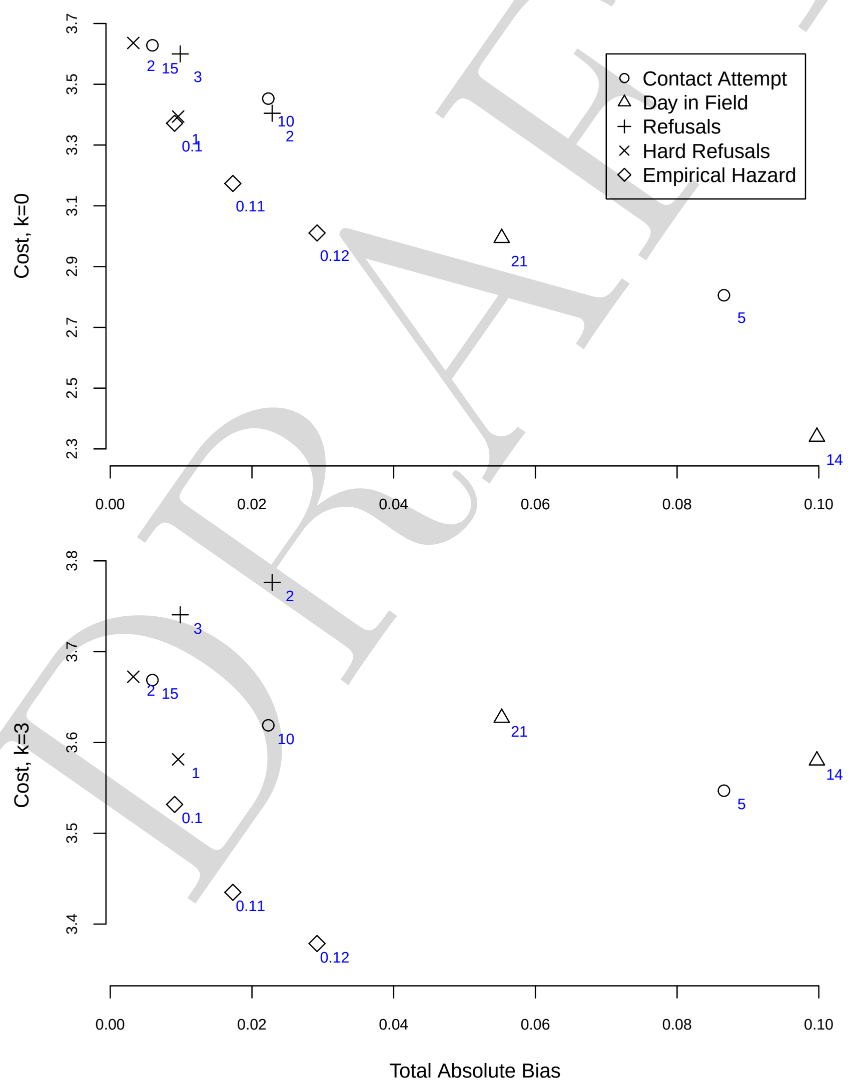
   for $k = 0, 1, \ldots, 5$. Under this cost structure, the unit of cost is a personal visit interview attempt; the cost of a telephone attempt is 0.05, and the cost of a lost interview is k. In the extremes, $C(0)$ only depends on the number of interview attempts, and $C(5)$ assigns the same cost to five personal visit attempts and one lost interview.

4 Results

Examination of the three aggregate statistical quality measures – $Q_3(\hat{\boldsymbol{\theta}})$: total absolute bias, $Q_4(\hat{\boldsymbol{\theta}})$: average absolute relative bias, and $Q_5(\hat{\boldsymbol{\theta}})$: maximum absolute relative bias – under each stopping strategy reveals similar results, and we use $Q_3(\hat{\boldsymbol{\theta}})$ to illustrate these findings. Figure 1 displays the total absolute bias and cost for $k = 0$ and $k = 3$, for a selection of stopping rules recommended by the significance tests. In these plots, the specific stopping times are shown in blue, below and to the right of each plotting character. For instance, the `x`, `o`, and `+` in the upper left corner of the top plot give the estimates of cost and total absolute bias for the policies which stop after two hard refusals, fifteen contact attempts, and three refusals of any type, respectively. In this figure, the "lost interview weight" of the cost function is 0, and so the cost displayed is simply the number of attempts per case (with telephone attempts assigned less weight than personal visit attempts). These three stopping rules have roughly equal cost, and the policy which stops after three hard or soft refusals results in just under 1% more total absolute bias than the policy which stops after two hard refusals. The policy that stops data collection after fourteen days has the lowest cost, and substantially more bias than most of the other policies considered.

In the bottom plot of Figure 1, the costs displayed incorporate both effort (number of contact attempts) and losses in the number of interviews obtained. While the y-axis of this plot does not have the same intuitive meaning as the top plot, examining the cost function with $k > 0$ is useful if we are interested in simultaneously evaluating the effects of various stopping rules on cost, biases, and response rate. With interview losses taken into consideration, the cost of the policy that stops data collection after fourteen days is above that of several alternative policies, and remains highest in total absolute bias. In any plot of cost versus bias, the "best" policies are located in the lower left corner (low cost and low bias). In this figure, the policies that stop pursuing cases based on estimated response propensity thresholds are in the lower left region, and are therefore pursued further. In Figure 2, we examine the hazard functions from which these response propensities are estimated.

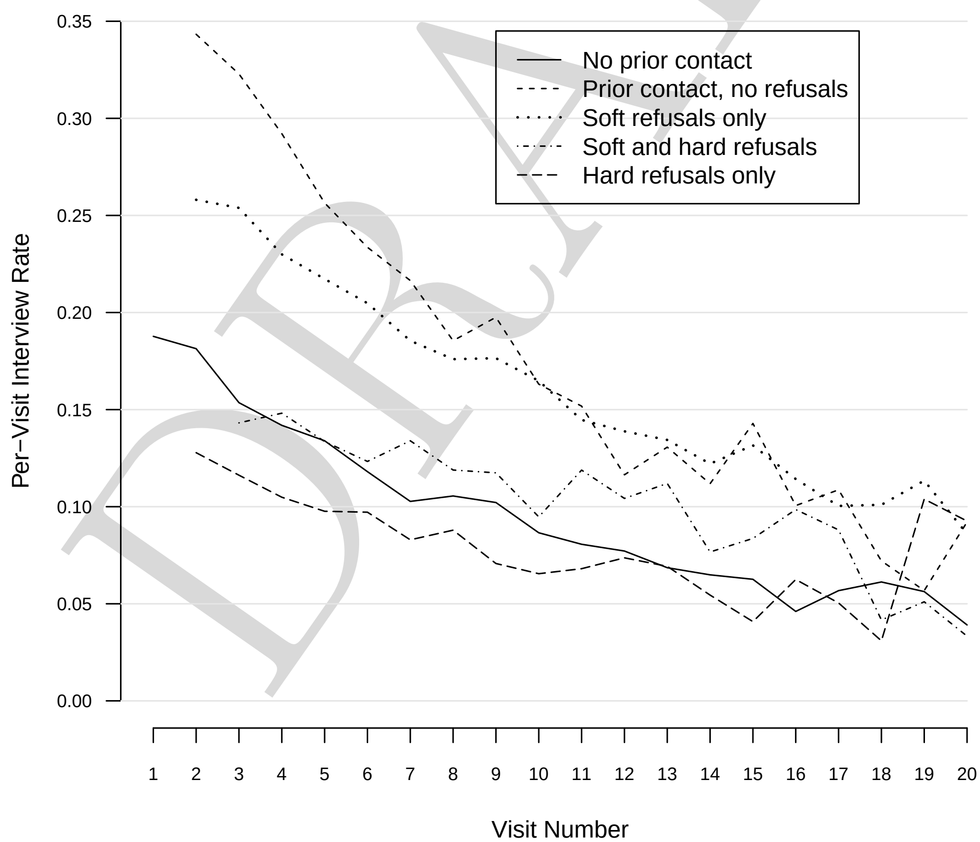
Figure 1: Cost and Total Absolute Bias for Various Stopping Strategies.



The following figure displays the empirical hazard, or per-visit interview rate, for each of the five mutually exclusive, time-varying "refusal" groups – G_0: No prior contact; G_1: Prior contact, no refusals; G_2: Prior soft refusal(s) only; G_3: Prior soft and hard refusals; G_4: Prior hard refusal(s) only – where a hard refusal is defined as in Dahlhamer and Simile (2009) (described in Section 3), and a soft refusal is defined as any refusal that is not a hard refusal. Each hazard function has a decreasing trend; that is, in general, the per-visit interview rate decreases as the number of interview attempts (visit number, t) increases. Overall, cases in the "prior contact, no refusals" and "soft refusals" groups have the highest per-visit interview rates, cases in the "hard refusals only" group have the lowest per-visit interview rate, and the per-visit interview rate of cases with both soft and hard refusals lies in-between.

In addition to the general trends in the hazard functions, there are interesting interactions between time and refusal group. The "prior contact, no refusals" and "soft refusals only" groups have higher interview rates than the baseline "no prior contact" group for all visits. However, while the hard refusal group has the lowest hazard rate for the first twelve visits, interview rates for this group when $t > 12$ are similar to interview rates of cases with large numbers of visits (twelve or more) and no contact up to the given value of t. Also, for most attempts, the hazard function for the "soft and hard refusals" group lies above the hazard for the baseline "no prior contact" group, but for early ($t < 4$) and late attempts ($t > 17$), cases in the baseline group have higher per-visit interview rates. Discrete-time hazard models allow us to incorporate these interactions in our estimation of response propensity, and to examine the relationships between of many additional case characteristics and response propensity. An assessment of these models and their associated costs will be presented at the workshop.

Figure 2: Empirical Hazard Functions for Mutually Exclusive Groups.



References

James Dahlhamer and Catherine Simile. Subunit nonresponse in the national health interview survey (NHIS): An exploration using paradata. In *Proceedings of the 2009 Joint Statistical Meetings*. American Statistical Association, 2009.

R. Sowmya Rao, Mark E. Glickman, and Robert Glynn. Stopping rules for surveys with multiple waves of nonrespondent follow-up. *Statistics in Medicine*, 27:2196–2213, 2008.