HRS | HEALTH AND RETIREMENT STUDY
A Longitudinal Study of Health, Retirement, and Aging
Sponsored by the National Institute on Aging

Evaluating Strategies for Reducing Field Costs in a Longitudinal Study

David R. Weir & Mary Beth Ofstedal
Survey Research Center
Institute for Social Research
University of Michigan

June 2016

Evaluating Strategies for Reducing Field Costs in a Longitudinal Study

David R. Weir and Mary Beth Ofstedal, ISR, University of Michigan

June 12, 2016

1. Introduction

Cost is a major concern for population surveys. In a world full of inexpensive "big data" the cost of scientifically sampled representative surveys using instruments thoughtfully designed for research requires increasingly difficult justification. Cost containment must become a priority for population research. Innovative sampling methods often summarized as "responsive design" have been used to hold down the cost of fresh cross-section population surveys. The essence of these approaches is to reduce the effort (cost) spent on cases that require a lot of effort but which do not have any greater scientific value than cases requiring less effort.

Longitudinal surveys have not seen the same innovation in design. Clearly, the loss of a participant with prior waves of data is a much greater loss to scientific value than the loss of a potential new participant in a new study. The latter is easily replaced by another less costly new participant, while the former cannot be replaced at any feasible cost. But in truly longitudinal studies the choice is typically not between obtaining an interview and permanent attrition. The choice is between obtaining an interview this wave or not. Indeed, in practice every longitudinal study makes such decisions every wave based on constraints of field period length, ethical concerns, and budget. What most of us do not do is make such decisions in a scientifically informed way, despite the fact that by virtue of longitudinal experience we have more information about the costliness and scientific value of cases than do fresh cross-section surveys. Viewed from that perspective, the basic principles of responsive design can be brought to bear on the management of effort in longitudinal studies to reduce cost and/or improve the scientific value of what is achieved for a given cost.

In Section 2 of the paper, we evaluate several approaches to cost containment using data from recent waves of the Health and Retirement Study (HRS). The HRS is a biennial longitudinal study of the US population 50 and older supported by the National Institute on Aging and the Social Security Administration. Its first wave was in 1992 and it is currently conducting its 13th wave in 2016. It typically achieves response rates of about 88-89% each wave from a mix of ~95% response from prior wave participants and ~30% from those who had previously participated but were not interviewed in the immediate prior wave. Permanent attrition by living participants averages about 1% each wave. Thus, the great majority of non-response in any wave are participants who remain in the study and have a significant likelihood of participation in future waves.

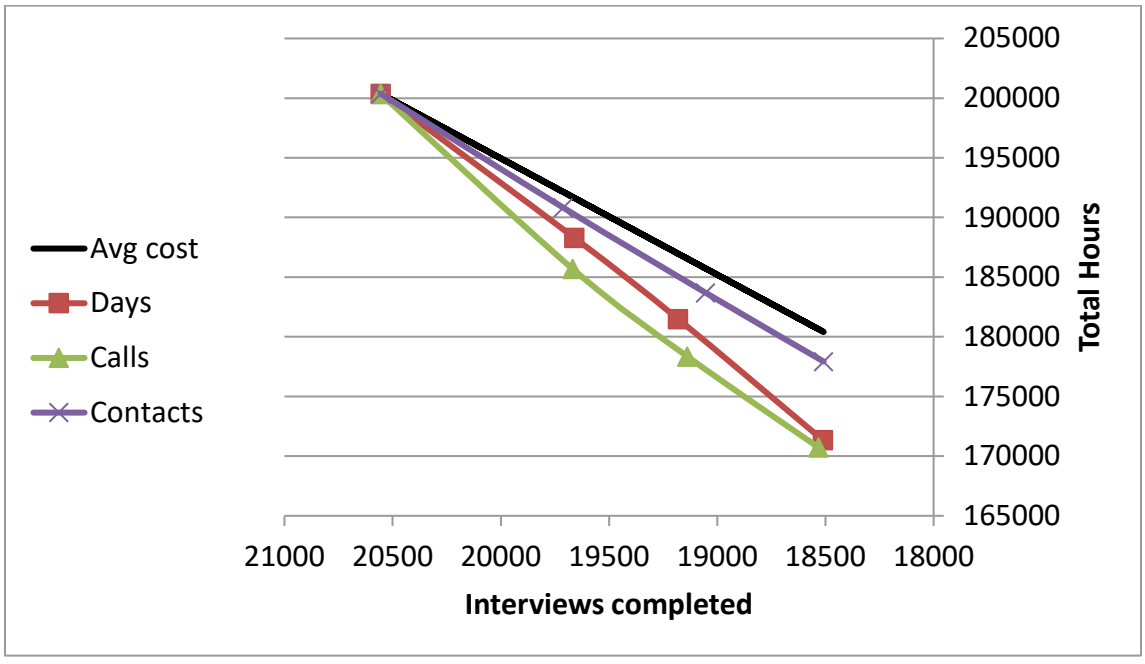
This research is taking place in the very real context of sponsor concern about the cost of HRS. A proposal has been made to consider taking HRS from a biennial to a triennial interview cycle. The effort-limiting strategies cannot produce savings equivalent to the 1/3 reduction from changing periodicity. Savings at that scale require ex ante censoring of the sample. In Section 3, we compare the scientific impact of several alternatives scaled to achieve the same 1/3 cost reduction.

2. Effort-limiting approaches in a single wave

The success of a responsive design strategy depends on two things: variance across cases in cost of seeking an interview, and in particular the presence of an upper tail in cost, and low correlation between cost and characteristics of interest. We evaluate strategies by comparing the cost savings and bias created for a given target reduction in response rate. This was done using data from the 2012 wave of HRS by simulating strategies to hit target (lower) response rates by limiting effort. We targeted 85%, 83%, and 81% response rates. Our three strategies were to 1) limit days, that is, to stop fieldwork when the target RR was reached, 2) limit calls per cases by setting a call limit that reached the target RR, and 3) limit contacts per case by setting a limit on the number of contacts that reached the target RR. We assess cost using an estimate of interviewer hours based on numbers of calls, numbers of contacts, and numbers of interviews, by mode. Most of HRS field costs are proportional to number of interviewer hours. Instrument programming, data processing, and other costs are not considered here.

Figure 1 shows the reduction in effort against the loss in interviews for these strategies. Because these effort metrics are discrete we can't reach exactly identical numbers of interviews by setting targets on different metrics. The average cost line shows the cost saving if each lost interview saved the average effort across all interviews. An efficient strategy should reduce effort by more than this. All three strategies improve on simply dropping some cases ex ante. The most efficient strategy, especially for small reductions in RR, is to limit calls. Somewhat surprisingly, simply stopping the field effort at the target RR does nearly as well. This result is strongly dependent on how HRS manages its field period and may not generalize well to other studies. Finally, limiting contacts is less efficient than the other two strategies.

Figure 1. Reduction in effort to obtain target RR; three strategies compared with average cost



We next examine the impact on key sample indicators. Table 1 shows this for the three strategies for a target RR of 83%. Because this is a relatively small reduction in interviews, the retained cases do not differ much from the total actually interviewed in 2012. Bias can be seen more clearly in the dropped cases. All three effort-limiting strategies have biases in the same direction. All reduce non-white cases more than white, all reduce less-educated and fair/poor health. All reduce greatly the number of cases interviewed via proxies, which is not surprising because proxies are often taken as a last resort in a difficult case. The primary purpose of having proxy interviews is to retain older subjects with cognitive problems. Interestingly, the rates of memory impairment are actually lower for limiting field period or calls per case and are only higher for limiting contacts. Overall, limiting days has the least bias, limiting contacts the most, and limiting calls somewhere in-between.

Table 1. Key Indicators of HRS2012 sample, actual compared to simulated effort-limiting designs

	Actual		**Retained cases (83% RR)**				**Dropped cases**		
Indicator	**(89% RR)**		**Days**	**Calls**	**Contacts**		**Days**	**Calls**	**Contacts**
% Nonwhite	32.6		31.9	31.4	31.2		41.6	48.3	50.4
% Education HS or less	53.1		53.1	53.0	52.8		53.2	54.9	56.7
% with memory impairment	19.3		19.4	19.4	19.0		17.8	17.4	23.7
% in fair/poor health	29.4		29.3	29.2	28.9		31.3	32.1	35.6
% proxy	5.6		4.7	5.1	4.9		18.4	12.3	14.1

The relatively small biases introduced by lowering RR through effort limits could easily be addressed by adjusting sampling weights. One might therefore want to use the increase in weight variance in conjunction with loss of sample size to compare the loss of effective sample size to the savings in cost. In HRS, the first-order determinant of sampling weight is race because of minority over-sampling. Adjusting weights for the loss of interviews with non-white participants would reduce overall weight variance. In fact, despite their lower sampling weights, non-white participants are more highly valued by sponsors and many users. Thus, any approach to studying effects on sampling weights would have to be done separately by race/ethnicity.

3. Options for a Large Cut

A tempting target for substantial reductions in field costs in longitudinal studies is to reduce periodicity, i.e., conduct interviews less frequently. PSID achieved major savings by moving from a one to two-year periodicity. NLSY rebuffed a proposal to move from two to three. As a biennial survey, the logical option for HRS is also to consider moving to a three-year interval. Changing periodicity makes no use of information on the relative costliness of cases or the relative value of an interview. It simply saves the average cost of an interview for every interview lost. We compare periodicity change to alternative approaches designed to achieve the same overall savings. The strategies considered in Section 2 truncate effort on difficult cases but cannot achieve large savings because we must first incur the cost of working cases until they reach the effort limit. To achieve large savings we must set aside some cases ex

ante, either by permanent cuts to the sample or some sort of rotation. As a longitudinal study, we have data on the past costliness of cases that can be used to predict ex ante who will be most costly in the future.

We show here four options for reducing the field costs of HRS by about 1/3 over six years. These were simulated by using data prior to 2010 to make hypothetical cuts for waves 2010, 2012, and 2014. This should be a good approximation to the impact of a similarly-designed cut for 2018, 2020, and 2022. The candidates are:

P: Periodicity change from two to three years. Simulated simply by taking away 1/3 of all interviews, or ½ of 2010 and 2012.
R: Randomly cut the sample by 1/3.
I: Information-based cuts to sample using data on costs and cooperation from 2006 and 2008. The number of cases cut was chosen to arrive at the same total cost reduction over 2010-14 as in options P and R.
S: Drop one **spouse** in coupled households. This preserves the household size of the sample as of 2008.

The number of individual and household interviews from each of these options is shown in Table3.1 below, compared with the actual number obtained combining three waves 2010, 2012, and 2014. Figure 3.1 shows the number of household interviews and Figure 3.2 shows the number of individual interviews. For equal dollars, periodicity delivers fewer interviews than any other option. For individual interviews there is little difference except for the information-based cut to sample, which saves about a third of the number of individual interviews that would have been lost by cutting periodicity. Looking at household interviews clearly the "PSID" model of only interviewing one person per household is most efficient at preserving household participation in the study.

Figure 3.3 shows the components of the enhanced face-to-face interview introduced in 2006. Reflecting the much higher interview count from information-based sampling, the number of observations on blood, physical performance (blood pressure, grip strength, etc), and psychosocial status are substantially higher under that option than any of the others.

Figure 3.4 shows the impact of cuts on number of interviews with subgroups that may be of special interest to specific research questions. Information-based cuts to the sample advantage veterans and slightly disadvantage African-Americans. This option could be modified to cut equally across groups of interest and in any event its effects can easily be corrected through modest changes in weights.

Table 3.1. Number of individual and household interviews 2010-14 with panel as of 2008, under different simulated cuts

	Number of interviews over six years	
	Persons	Households
A: Actual	42,448	30,363
P: Cut a wave	28,271	20,222
R: Cut sample at random	28,266	23,047
I: Cut sample in an informed way	33,016	24,687
S: Cut spouses	28,428	28,428

Figure 3.1. Number of household interviews 2010-14 with panel as of 2008, under different simulated cuts

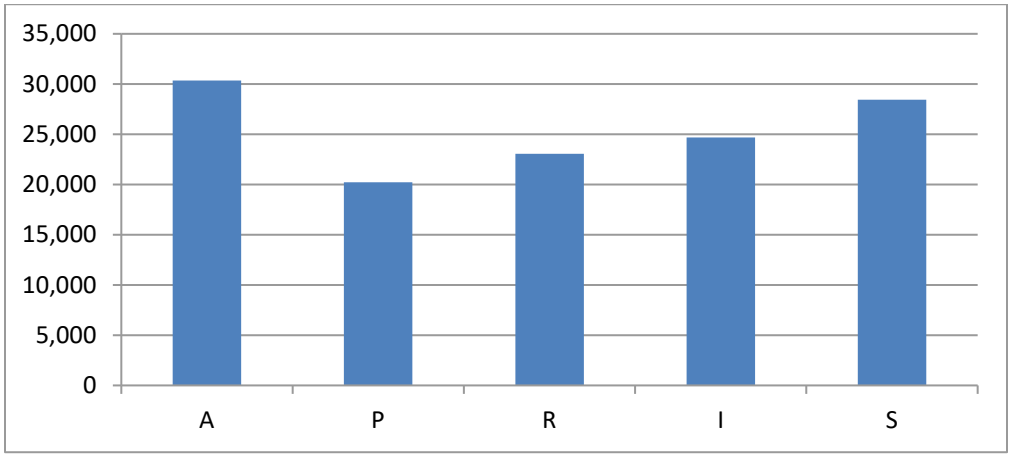


Figure 3.2. Number of individual interviews 2010-14 with panel as of 2008, under different simulated cuts

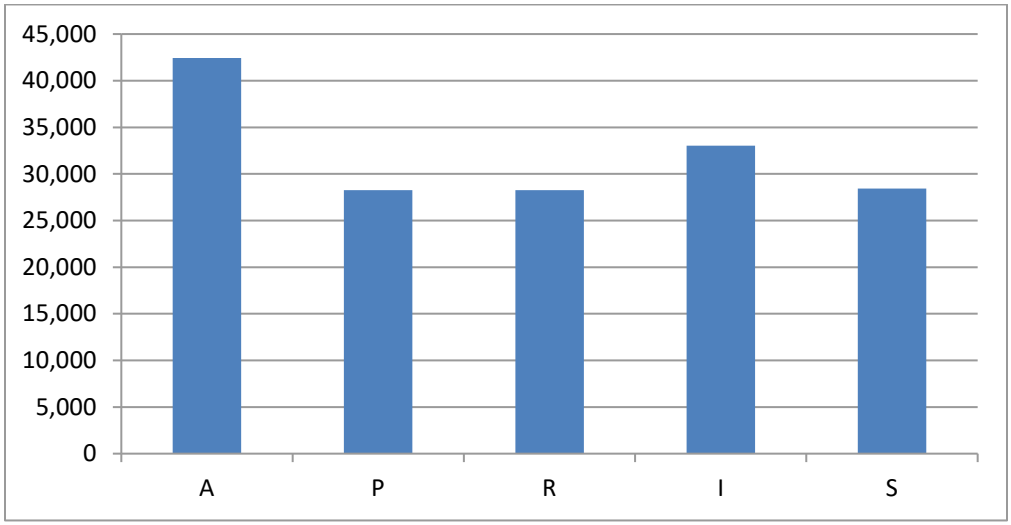


Figure 3.3. Number of biomarker interview components measured 2010-14 with panel as of 2008, under different simulated cuts+

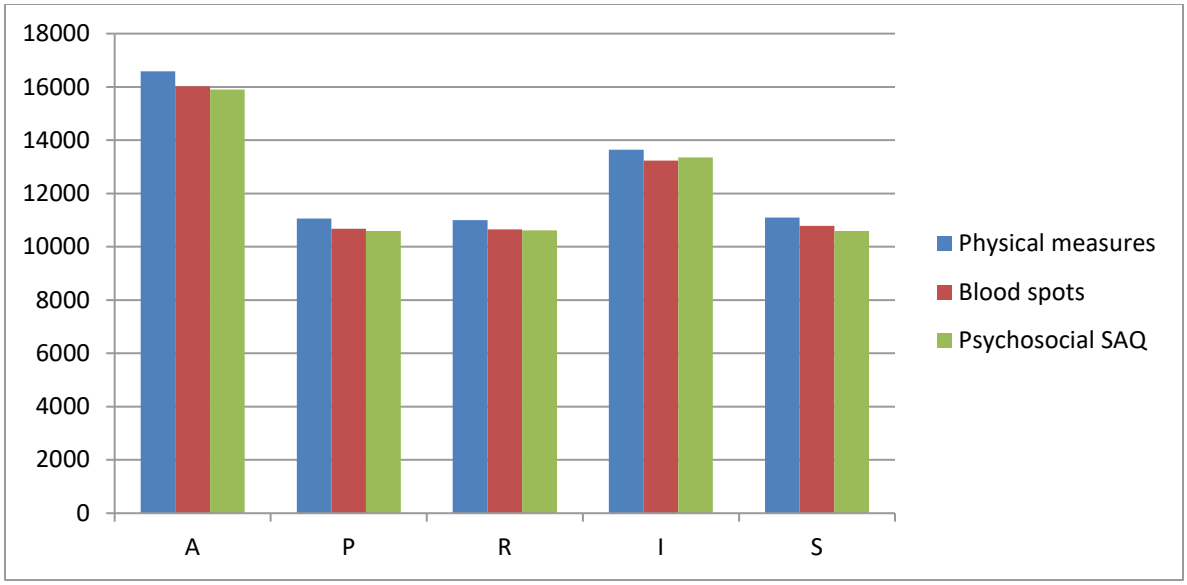
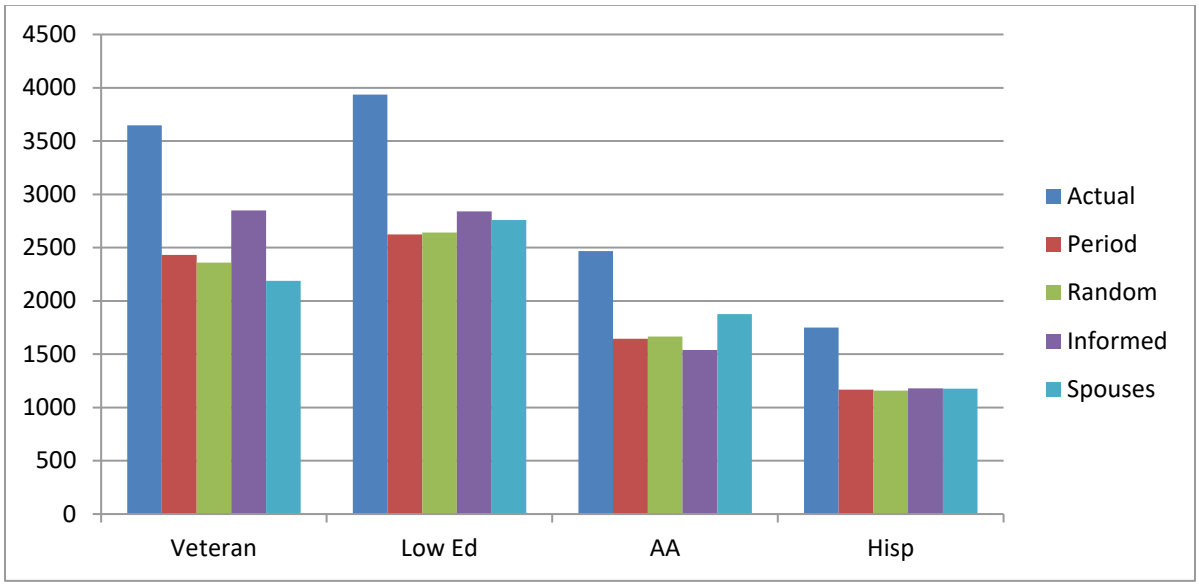


Figure 3.4. Number of interviews by key characteristics

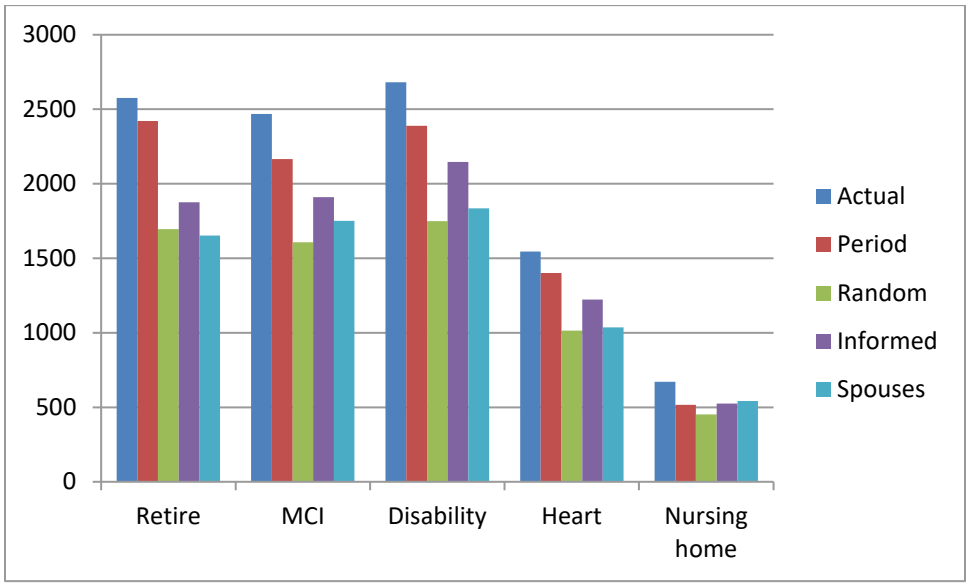


Longitudinal outcomes.

The key advantage of periodicity change over sample cuts is that in principle most longitudinal events are eventually observed when interviews are done at longer delays whereas they are not observed if cases are permanently dropped. This is shown in Figure 3.5 for incident events of retirement, development of MCI, disability, heart disease, and movement into a nursing home. Periodicity still has some negative impact. This is because for some people the event was only registered in 2010 or 2012, only half of which are observed. If the person then failed to give an interview in 2014 we would not record the event. As Figure 3.5 shows that is a relatively small loss compared to the effect of cutting people out of observation for the entire six years.

This advantage of periodicity change over sample cuts disappears if we allow the cut cases to come back at some future wave when funding is restored. This is why periodicity is a preferred approach to a permanent reduction in the study whereas sample cuts are a preferable option for temporary funding shortfalls. And again an informed selection of cases to cut preserves a higher proportion of events than do other alternatives.

Figure 3.5. Number of incident events observed at interviews 2010-14 with panel as of 2008, under different cut options



Conclusions. We find that it is possible to reduce cost proportionally more than the number of interviews by using information available in a longitudinal study to select cases and/or to manage effort on cases. For modest cost reductions, the biases are also modest and correctable by weighting. Large cuts, not surprisingly, create large losses of scientific value.