Preliminary, Comments Welcome

Reordering the Darkness: Application of Effort and Unit Nonresponse in the Survey of Consumer Finances

Arthur B. Kennickell
Senior Economist and Project Director
Survey of Consumer Finances
Mail Stop 153
Federal Reserve Board
Washington, DC 20551
Phone: (202) 452-2247
Fax: (202) 452-5295
Email: Arthur.Kennickell@frb.gov
SCF on the Internet: http://www.federalreserve.gov/pubs/oss/oss2/scfindex.html

Joint Statistical Meeting, San Francisco, CA
August 2003

Abstract

Unit nonresponse in surveys has three key underlying components. Some respondents may be difficult to contact; where contact has been made, others may be resistant to participating; finally, there may be deficiencies or variations in the effort applied to persuade the sample of cases to participate. If any of these factors have systematic effects on response that are correlated with the measurements of interest, then there will be bias. This paper uses data on case administration and other data to consider the distribution of effort across cases in the sample for the 2001 Survey of Consumer Finances (SCF). In the absence of specific guidance, interviewers and field managers will choose the level of effort that is convenient for them to apply to individual cases that is also consistent with whatever general constraints they are given.. As the analysis in the paper indicates, there were both extreme and probably very costly concentrations of effort as well as systematic variations in effort across cases that appear to carry through to overall nonresponse. The paper discusses two types of controlled case management as a means of managing costs and improving data quality. First, equal treatment would be required of all respondents up through a common level of effort. Subsequently, subsampling would be used to mitigate the effects of nonresponse in the first phase.

 The author is particularly grateful for discussions with Leslie Athey, Steven Pedlow, and Fritz Scheuren. Eternal gratitude is due to the NORC and Federal Reserve staff and SCF respondents who made it possible to have data to analyze.

For survey data, unit nonresponse is generally a step into the analytical unknown. Three factors are prominent in explaining nonresponse. First, some respondents are very difficult to contact even to negotiate an interview. Second, when contact has been made, others may be resistant to participating. Third, there may be variations or deficiencies in the application of effort to inform and persuade affect respondents. To the degree that the mechanism behind the nonresponse can be shown to be at least conditionally unrelated to the variables of interest, the serious loss is one of estimation efficiency through a reduced number of observations. Experience suggests that the effects of nonresponse are most often more complicated.

Nonresponse may be addressed in several ways—before, after, or during data collection. Before data collection, interviewers may be trained in techniques to contact respondents and gain their cooperation; it may also be possible to deal with some statistical inefficiencies induced by nonresponse by stratifying on characteristics associated with nonresponse. After data collection is completed, many surveys use post-stratification to adjust for important known deviations of the interviewed population from the target population—differences that may reflect both sampling error and nonresponse.

During the field period, additional application of effort might increase response rates. However, straightforward dedication simply to reducing nonresponse may lead to the application of very large amounts of effort, but without accompanying confidence of reducing nonresponse biases. If contact and response propensities were sufficiently understood and adequate data were available rapidly enough during the field period, one might be able to make more efficient use of resources by targeting specific cases or types of cases.

Unfortunately, it is not straightforward to understand either contact or response propensities—not least because virtually everything we know from the field is affected by the application of effort. "Effort" is not a simple commodity. It is a complex product of the decisions of interviewers, their managers, survey organizations, and survey sponsors, each of whom may face a different set of objectives, constraints, and incentives. Where effort is only partially observed, as in the case of most of interviewers' work, complicated patterns of "accidental" deviations from optimal behavior and even shirking may arise (see Kennickell, 2000a).

This paper focuses on nonresponse and the application of effort in the case of the 2001 Survey of Consumer Finances (SCF), using case-level administrative records along with

information recorded by interviewers for each sample address and tract-level data matched from the 2000 U.S. Census of Population. It follows on earlier work directed toward related goals in Kennickell (1999a and 1999b); recent work reported in Groves et al. (2003) deals with related problems of monitoring and management of field effort. The first section of this paper gives background on the SCF, including information on nonresponse. The next section presents a simple behavioral model of survey organizations. The third section describes the level of effort applied in the 2001 SCF and estimates a model of the likelihood of continued application of effort to cases. The fourth section argues for the development of a contact strategy that would allow better estimation of and control for the respondents' role in nonresponse while allowing greater control over costly field operations. A final section concludes and makes recommendations for further research as well as changes in the administration of the 2004 SCF.

I. Background on the SCF

The SCF is designed as a survey of households' finances and their use of financial services. It is conducted on a triennial basis by the Federal Reserve Board in cooperation with the Statistics of Income Division (SOI) of the Internal Revenue Service. The SCF data used in this paper derive from the 2001 survey.[1] The data for this survey were collected by NORC, a national organization for social science and survey research at the University of Chicago, between the months of May and December of 2001.

Data for the survey are collected using computer assisted personal interviewing (CAPI). The median interview required 79 minutes to administer, though some took considerably longer. Although an atempt was made to make initial contact with every case in-person, 34.6 percent of the completed interviews were conducted by telephone. This use of the telephone very largely reflects the preferences of the respondents.[2]

---

[1]See Kennickell (2000b) for a review of SCF methodology and Aizcorbe *et al.* (2003) for a summary of data from the survey.

[2] By the sample types described below, 51.0 percent of list sample cases and 25.9 percent of AP cases were completed by telephone. Although the data suggest that relatively wealthy households were more likely to interviewed by telephone, substantial fractions of cases were completed by telephone in virtually all the areas where the survey was conducted.

Table 1: Percent distribution of final outcome codes for area-probability and list samples.

	AP	LS
OUT OF SCOPE		
Not a housing unit	4.0	NA
Vacant housing unit	7.2	NA
Seasonal vacant (vacation home)	2.4	NA
Sample incorrect (incorrect listing)	0.5	NA
Deceased (no proxy possible)	0.1	0.4
No eligible R in household	0.1	0.0
Permanently out of the country	0.0	0.2
Other out of scope	0.0	0.0
COMPLETE		
Complete interview, telephone	13.4	13.8
Complete interview, in-person	38.5	13.8
Complete interview, phone conversion	1.7	1.8
Complete interview, in-Person conversion	4.7	1.0
Partially completed interview	0.1	0.1
REFUSED		
Postcard refusal	NA	13.2
Final refusal, conversion attempted	6.6	8.1
Final break-off of interview	0.1	0.1
Final refusal by gatekeeper	0.0	0.1
Final unlocatable	0.2	0.5
R unavailable for field period	0.2	0.8
Language barrier (other than Spanish)	0.6	0.2
Physically or mentally incapacitated	0.2	0.2
Other noninterview	0.2	0.8
CENSORED		
Final stopped work	18.4	43.8
N	4,993	5,026

The sample for the survey is based on a dual-frame design. The first part is a national area probability (AP) sample of households in 100 primary sampling units (PSUs) across the U.S. Units in the sample were selected using a multistage process that yields locally clustered sample cases. Although there is some stratification in this design, the ultimate probability of selection is the same for every case in the sample. The second part of the design is a list sample (LS), which yields a sample of taxpayers in the same PSUs as the AP sample. This sample is selected from statistical records derived from tax returns, using a stratification scheme based on models that approximate the wealth of the underlying tax unit. Strata that correspond to relatively high levels of predicted wealth are progressively oversampled. The dual-frame design allows for good coverage of broadly distributed characteristics—such as home ownership—and others that are relatively concentrated among wealthy families—such as ownership of a closely-held business. In addition, the sample provides a means of damping biases associated with differentially higher nonresponse rates among wealthy families.

The sample for the 2001 survey contained about 10,000 observations approximately equally split between AP and LS cases (table 1). The set of completed cases contained 4,449

observations—2,917 AP observations and 1,532 LS ones. Overall, 14.2 percent of the AP cases and 0.7 percent of the LS cases were ineligible to be interviewed.[3]

The response rate (adjusted for ineligible units) for the AP sample was 68.1 percent.[4] Unlike AP cases, LS cases were given the option of returning a postcard to refuse to participate in the survey in advance of their being approached by an interviewer; 13.2 percent of the sample did so. In addition, a relatively small number of cases were deleted from the sample during a review designed to eliminate members of the *Forbes* list of the 400 wealthiest people in the U.S., and a few other very unusual people.[5] Leaving aside the deleted cases, the LS response rate was 30.7 percent; removing the postcard refusals from the denominator of the estimate increases that rate to 35.4 percent. By far, the largest category of nonresponse was "final stopped work"—21.7 percent of the eligible AP cases and 51.4 percent of the LS eligible cases; these are cases that, in theory, remained eligible for further work at the end of the field period.

II. A model of case management

Through the process of case management, survey cases are resolved into completed or refused status as interviewers try to present information to respondents through a number of actions ("attempts") and respondents arrive at a final determination of their willingness to participate. As work continues, the set of cases with censored outcomes—those remaining at risk to be completed or refused in a future attempt—shrinks (figure 1).

Several factors may complicate this process. First, it is not always possible to reach respondents to provide them with information. Second, there is no firm definition of what constitutes a final determination of unwillingness to be interviewed. For example, some "refusal

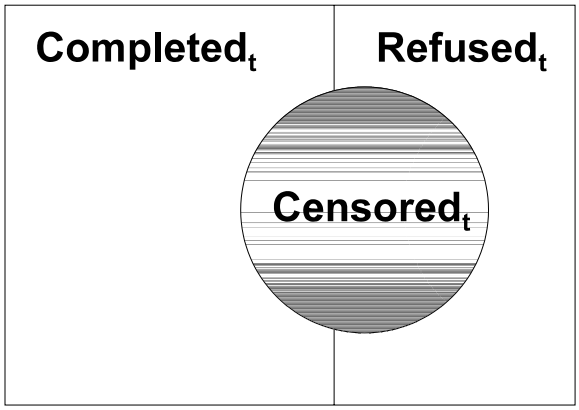
---

[3]Such cases are excluded from all analysis reported in this paper unless otherwise specified.

[4]Final outcome codes indicating refusal or a final termination of work short of a resolution as completed or refused were assigned by interviewers only with the permission of their managers.

[5]Because the analysis reported in this paper is largely concerned with field procedures, the LS postcard refusals and the deleted cases are excluded from the analysis presented in this paper except where specifically noted otherwise.

Figure x: Division of sample observations at time *t* into complete, refused, and censored observations of which an unknown proportion would ultimately resolve as complete or refused if sufficient effort were applied.



converters" are renowned for their ability to persuade people who have repeatedly refused other interviewers. Third, the application of effort is generally a dynamic decision process that is a function of many factors. Some cases may not be worked to the point that an unambiguous resolution is reached, probably because those cases are ones perceived to be either too "expensive" or insufficiently "valuable"—or put another way, because time, money, and interest are limited.

As a stylized framework for the analysis of the SCF data on case resolution, consider the following simple model. Let the population distribution of the characteristics of interest be given by $\mathcal{R}^{\bullet}$; for a sample of size N from the target population let the distribution be given by $\mathcal{R}$. To avoid needless technical complications, assume that $\mathcal{R}$ and $\mathcal{R}^{\bullet}$ are identical. Assume that the scientific goal of the data collection process is to minimize, to the degree possible, the generalized distance between the distribution of characteristics of the final set of participants and that of the full sample by targeting effort to cases still at risk at each point in the data collection period. Nearly every survey has some formal or informal targeting to direct effort, even if it is only high-order response rate goals. For expository convenience, suppose that *a priori* there are *K* observable discrete categories $\mathcal{R}^K$ (where N^k is the number of elements in the *k*th cell and $f^k=N^k/N$) on which we want the population and the cases interviewed to coincide in terms of proportions.

As effort is applied to a case in the field at a given step *t*, the case may resolve as complete or refused, or the ultimate outcome may remain censored. Let the indicator function C_t^i specify the outcome set, where 1=complete, 2=refused, and 3=censored. Let effort on case *i* at point *t+1* be given by W_{t+1}^i, which is taken to equal one if effort is applied and zero otherwise; for simplicity the cost of effort is normalized to equal one per application. From the point of view of the surveyor at step *t* planning effort at *t+1*, the probability function for the set of outcomes for a case *i* given the application of additional effort at *t+1* is expressed as

$\pi(C^i_{t+1} \mid W^i_{t+1}, W^{i*}_t, I^i_t)$, where W^{i*}_t summarizes all effort from the beginning of the field period to step *t*, and I^i_t denotes the information available to the survey agents about the case at step *t*.

The total cost over the field period (t=1 to T) of all efforts over all cases (i=1 to N) must be within a budget W^*. There may also be additional managerial constraints that limit the number and distribution of interviewers available, specific contractual obligations, etc.; these will be ignored here. For each of the categories $\mathcal{R}^k$ at step *t*, there are n^k_t survey participants, a number which may deviate from the desired proportion in each group.

With the information available at step *t* in a field period, the surveyor's problem is to choose a vector of effort $\mathbf{W}_{t+1}$ to be applied in the next period so that in expectation (denoted E_t) over the whole sequence of potential efforts over the remaining field period, the constraints are satisfied:

$$\left(n_t^k + \mathrm{N}^{\,k}_{t+1}\right) \Big/ \sum_{k=1}^{K} \left(n_t^k + \mathrm{N}^{\,k}_{t+1}\right) = f^k, \ \forall k$$

$$\textit{where } \mathrm{N}^{\,k}_{t+1} \equiv E_t\left[\sum_{i=1}^{N^k} \sum_{\tau=t+1}^{T} (W^i_\tau = 1) * (C^i_\tau = 1)\right]$$

$$\textit{and } \sum_{k=1}^{K} \sum_{i=1}^{N^k} \sum_{\tau=1}^{T} E_t(W^{ik}_\tau = 1) \le W^*$$

In this example, the expected response rate and the expected length of the field period are endogenously determined at every decision point.[6] The optimal choice to proceed on any given case is a complex dynamic function of the entire sequence probabilities of success on all observations still at risk at *t*. But if the expectation of these probabilities can be calculated for each case and actions across cases are independent, then an obvious decision rule applies: at point *t*, choose to exert effort on the N^k_{t+1} cases in each group with the lowest expected costs

[6]This discussion abstracts from the integer constraints on the choice of sample elements to which to apply effort and the possibility that a full solution may not exist due to pathological probability distributions, other events that render constraints insoluble, etc.

relative to the likelihood of completion. In early stages, where I_t^i contains little information, it may be rational to apply effort to "learn" about response propensities of various "types" of cases.

Typical practice of survey organizations that aspire to scientific practice may be less deterministic and constrained in ways beyond what the model supposes, but at least some key aspects must tend to carry through over time if analysts insist on representative data and if the organization survives economically. Four points from this stylized framework have important practical implications for the analysis of the call records of a survey. First, the relatively easy (likely) cases should tend to be approached and, on average, interviewed first. Second, the distribution of effort across cases is endogenously determined. Third, over the field period, the cases remaining at risk should become increasingly dense in cases that would ultimately refuse.

A key difference between the assumptions of the model and what is desired in practice is that the equivalent of $\mathcal{R}^k$ is generally very difficult to define—at least in part because the measures of ultimate interest are very often not observable *a priori*, so that proxies must be used. In the case of the SCF, the most common proxies have been response rates in PSUs and at least a minimum level of effort in all areas for the AP sample, and special targeting of respondents by stratum for the list sample.

Three very important practical factors are omitted in the model. First, because the incentives and constraints faced by the different players—interviewers, managers, survey organizations, and sponsors—are not always the same, their views of the optimal application of effort may also differ. Second, actual effort applied in the field is not directly observed by anyone other than the interviewer and perhaps the respondent, and most of what is known is filtered through the interviewer. As a result, the ability to make adjustments to effort is potentially limited by the scope of the instruments managers have to influence interviewers' behavior directly or indirectly. Third, even when the incentives are aligned and some information on effort is observed, it is still often quite difficult, even in principle, to process attempt-level records and related data into a form that could be used by the managers to guide interviewers in achieving a project's goals. Each of these points is deserving of a full treatment in a separate paper.

III Case records in the SCF

SCF interviewers are required to maintain "call records" on all actions taken on each observation in their assignment. Managers, "locaters," and other specialists also may record such data. In addition, cases may be transferred among interviewers. Generally, interviewers enter their call records into their computer based on notes they record while they are working using a paper "face sheet" generated for each case. The primary incentive for interviewers to enter their call records fully and correctly is that this information is used by their managers to judge interviewers' productivity—those who do not enter call records are assumed not to be working. Managers also use such information to assist interviewers in developing strategies for individual cases. Nonetheless, since data entry is burdensome, it is reasonable to expect that there were at least some failures to record efforts exhaustively, but such problems should be small. At the same time, because the managers had many interviewers each of whom had many cases, there was no formal field process model against which effort could be evaluated, and the organization of the call record data made it very cumbersome to evaluate *sequences* of effort on individual cases, comprehensive and consistent management of effort would have been extremely difficult.

The information entered into a call record includes the following: the date and time of the operation noted; whether the action was taken in person, by telephone, or by mail; whether the interviewer interacted with the respondent, some other person, or no one; and a working "disposition code" describing the operation or its outcome.[7] The dispositions codes, in theory, have a loose hierarchical progression from basic descriptions, to indicators of complex engagements, to a final classification as a type of completed case or refusal, their interpretation as part of a sequence of efforts is often complicated by forcing both action and outcome into one code.

Given the May-to-December field period, approximately 210 days is the longest than any observation could have remained "in play," defined here as the days elapsed between the earliest

[7] There were also fields for the entry of name and address information and unformatted general comments, none of which are available to the project staff outside NORC. Such information at NORC is destroyed at the end of the field period. Some other types of case action are purely informational or are continuations of other entries too long to be contained in one entry; such entries are ignored here.

call record of any sort other than an initial mailing, and the last one.[8] The median case remained in play for over three months (table 2), but there is a long right tail of the distribution that runs to the length of the entire field period. For completed and refused LS cases, the distribution of time in play is shifted upward from the distributions for those response groups in the AP sample; in contrast, the distribution for censored LS cases lies below that for the censored AP cases. However, this relatively pure time measure does not give a clear sense of the amount of effort that was actually expended over the period.

Unfortunately, the data in the call records needed for a deeper investigation are flawed, most importantly in that the standards for describing events in terms of disposition codes and other information were not uniform across all interviewers or their immediate managers. In some cases the data recorded may even be seen as internally inconsistent—for example, a case where the disposition code suggests that a respondent refused, but there was no record of a contact with a person. In other cases, there may be multiple reports of a set of events that might better be treated analytically as a single event—for example, an interviewer who made large number of stops at a house over the course of a day of other work in the neighborhood. In general, for this analysis a record was taken to be any type of "attempt" to contact the

[8]The cases resolved earliest were worked by interviewers who attended the first of two main training sessions, which were about two weeks apart.

Table 2: Mean and quantiles of the distributions of the days a case is in play; by final case disposition and sample type.

	All			*Complete*			*Refused*			*Censored*		
	All	*AP*	*LS*	*All*	*AP*	*LS*	*All*	*AP*	*LS*	*All*	*AP*	*LS*
Days in play:												
Mean	95.9	83.7	108.0	53.3	49.0	61.4	121.6	124.6	119.3	148.0	173.1	137.4
P5	4	1	10	1	0	2	54	53	54	100	126	97
P10	9	5	28	4	3	7	75	87	70	114	140	109
P25	37	19	76	14	11	25	106	113	103	130	162	125
P50	108	69	123	41	33	58	124	127	123	147	176	139
P75	144	144	144	86	78	92	142	144	141	170	194	152
P90	171	181	162	121	121	121	161	160	161	188	200	169
P95	182	195	171	141	143	138	172	179	168	197	203	175
P99	201	206	180	176	181	165	187	197	181	206	211	181
Number of obs.	8,597	4,285	4,312	4,448	2,916	1,532	1,002	437	565	3,147	932	2,215

respondent, where the record type indicated that the information related to any type of field event (other than a simple update of an address or a comment) or an appointment, and where the action described was made in person, by telephone, or by mail. An "on-site attempt" was taken as the subset of such actions made in-person. A "contact" was defined as the subset of attempts where the data indicated that someone—not necessarily the respondent—was contacted and the working disposition code was not in obvious conflict; also included are instances where the disposition code made contact inevitable (e.g., "completed cases"). "On-site contacts" were defined analogously to on-site attempts. Because these definitions are relatively loose, they may overstate the level of effort actually applied. Unfortunately, there did not appear to be obvious alternative mechanical definitions that were not clearly overly restrictive, and the important findings appear to be robust to simple perturbations of the definitions.

Obviously, there is a behavioral component in all of the effort measures. The decision to make a visit to a sample address, rather than to reach the respondent in another way or not to apply effort at all, suggests that a decision maker somewhere in the data collection process determined that this was a strategy productive in a sense relevant to them. Contacts usually require both determination on the part of the interviewer, as well as some degree of willingness in the respondent to receive the contact.

The effort expended on SCF cases in 2001 tended to be fairly concentrated. For example among the AP sample, 5 percent of the cases accounted for 18 percent of total attempts, and 20

percent for 47 percent of the total; this sort of disproportion also holds over cases viewed separately by final dispositions. The effort measures all have a long right-hand tail, which is truncated in the statistics presented in tables 3 and 4 at 20 or more. Although these tails may be somewhat exaggerated by the measurement problems noted, close examination of the underlying data suggests that the shape is not badly distorted by error.

For any type of attempt to reach the respondent, the distributions for refusals and censored cases are shifted to the right of the distribution for the completed cases. This difference serves to indicate both that some completed cases were relatively easy to convince—over 20 percent of those who ultimately agreed did so within three attempts of any sort—and that even observations that gave strong signs of refusing were pursued. Overall, the distributions of attempts for final refusals and censored cases are very similar to each other. By sample type, the clearest difference is the tendency for a larger number of attempts to be needed to secure an interview in the list sample than in the AP sample.

On-site attempts are relatively expensive. The distributions of such attempts look very similar across outcome types, implying that the differences across outcomes in all types of attempt are explained by use of less expensive means (mail and telephone) than on-site attempts. LS cases tended to have fewer on-site attempts than AP cases for all ultimate outcomes, perhaps

Table 3: Distribution of all types of attempt and on-site attempts, by final disposition and sample type.

	All types									On-site								
	Complete			Refused			Censored			Complete			Refused			Censored		
Num	All	AP	LS	All	AP	LS	All	AP	LS	All	AP	LS	All	AP	LS	All	AP	LS
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.3	1.0	16.3	10.8	0.0	18.8	11.1	1.4	15.1
1	4.3	5.4	2.2	0.2	0.0	0.4	0.7	0.5	0.8	21.0	17.1	28.5	16.9	0.7	24.1	16.7	4.7	21.7
2	8.2	9.6	5.6	1.7	1.8	1.6	1.9	1.1	2.2	20.1	19.3	21.5	14.2	7.8	17.0	16.9	7.5	20.8
3	10.7	12.2	7.8	3.7	5.3	2.5	3.6	2.3	4.2	14.8	16.0	12.5	12.8	10.6	14.1	13.0	9.4	14.6
4	10.4	11.2	8.8	5.4	4.8	5.8	5.2	4.1	5.7	10.8	11.9	8.5	10.2	11.0	9.4	10.1	11.0	9.7
5	9.4	9.7	8.8	8.2	8.1	8.3	6.9	4.9	7.8	6.7	7.7	5.0	8.4	11.3	5.2	8.0	11.6	6.5
6	8.2	7.5	9.6	8.1	4.4	11.0	6.6	6.5	6.7	4.9	6.1	2.6	6.3	12.4	4.0	5.8	10.0	4.0
7	6.5	6.4	6.7	8.2	9.4	7.2	8.1	5.9	9.0	3.6	4.5	1.8	4.4	9.2	1.5	4.1	9.7	1.8
8	5.9	5.4	6.9	9.6	9.7	9.6	7.2	7.5	7.1	3.0	3.7	1.6	3.2	8.1	2.0	3.2	5.8	2.1
9	4.8	4.2	5.9	6.1	5.8	6.3	7.5	6.8	7.8	2.1	2.9	0.7	1.9	4.8	0.5	2.8	6.7	1.2
10	4.9	4.4	5.9	8.4	8.7	8.1	6.1	6.8	5.8	1.7	2.4	0.4	2.2	3.7	1.1	1.9	4.1	0.9
11	3.7	3.2	4.7	5.6	4.1	6.7	5.7	6.3	5.5	1.2	1.7	0.1	1.8	3.7	0.9	1.4	3.8	0.5
12	3.2	2.8	4.1	4.3	3.5	4.9	5.4	6.5	4.9	0.9	1.2	0.3	1.4	3.0	0.2	1.1	2.4	0.5
13	3.1	3.0	3.1	3.7	3.7	3.8	4.8	4.0	5.1	0.5	0.6	0.1	1.3	3.0	0.7	0.9	2.3	0.3
14	2.7	2.4	3.3	3.9	4.6	3.3	3.8	3.3	3.9	0.8	1.2	0.1	1.2	2.1	0.0	0.5	1.4	0.1
15	2.1	1.7	2.9	4.2	4.6	3.8	3.7	3.8	3.7	0.5	0.7	0.0	0.6	2.8	0.4	0.6	1.6	0.2
16	1.6	1.5	1.6	2.6	2.8	2.5	2.5	1.9	2.7	0.3	0.4	0.1	0.4	0.9	0.0	0.4	1.1	0.1
17	1.6	1.2	2.6	2.0	2.1	2.0	2.5	2.9	2.3	0.2	0.3	0.0	0.4	0.9	0.0	0.2	0.3	0.2
18	1.4	1.0	2.0	2.7	2.8	2.7	2.1	2.7	1.9	0.0	0.0	0.1	0.3	0.9	0.2	0.2	0.5	0.1
19	0.9	0.9	1.0	2.3	2.8	2.0	1.8	2.3	1.5	0.2	0.2	0.0	0.0	0.5	0.0	0.3	1.0	0.0
20+	6.3	6.4	6.4	9.1	11.3	7.4	14.0	20.1	11.6	0.6	1.0	0.0	2.7	0.0	0.0	1.4	3.7	0.0

because telephone numbers may have been more readily available for respondents identified by name.

Differences among the outcome types are stronger for contacts. Relative to completed cases, the distribution for the ultimately refused cases is shifted to the right, but that for the censored cases is shifted to the left. These results suggest that where a respondent could be contacted, refusals were pursued; but where contact was more difficult, effort may have been more likely to be diverted elsewhere. However, because no additional information is available on the actual decision process, the differences could also reflect other decisions of interviewers and managers to pursue cases that were perceived as relatively less costly in a number of ways.

For on-site contacts, the distributions for both final refusals and censored cases are shifted leftward relative to the completed cases. The difference is explained by the very large fraction of the uncompleted cases that was never reached on-site by an interviewer—about a quarter of the ultimate refusals and almost 35 percent of the censored cases.

Table 4: Distribution of all types of contact and on-site contacts, by final disposition and sample type.

	All types									On-site								
	Complete			Refused			Censored			Complete			Refused			Censored		
Num	All	AP	LS	All	AP	LS	All	AP	LS	All	AP	LS	All	AP	LS	All	AP	LS
0	0.0	0.0	0.0	2.5	0.9	3.8	11.7	4.6	14.6	12.0	3.6	27.9	24.8	5.8	39.8	34.6	10.3	44.7
1	18.3	21.3	27.9	6.4	3.0	9.0	14.2	7.2	17.2	34.2	32.2	37.9	22.7	16.1	27.9	24.7	16.4	28.2
2	21.9	22.1	37.9	13.5	11.3	15.2	13.5	9.1	15.3	25.9	28.6	20.6	15.4	17.2	13.9	15.5	17.7	14.7
3	17.4	16.7	20.6	16.3	13.1	18.8	12.0	12.5	11.8	12.6	14.9	8.2	12.5	16.8	9.0	9.3	15.5	6.7
4	11.6	10.7	8.2	11.9	11.5	12.3	10.2	9.6	10.4	6.5	8.2	3.1	9.2	14.7	4.9	5.5	11.4	3.1
5	7.9	7.5	3.1	13.4	14.0	12.8	7.9	8.7	7.6	3.4	4.5	1.4	4.8	7.8	2.4	3.5	8.4	1.5
6	5.9	5.2	1.4	7.4	9.7	5.6	6.7	7.1	6.6	2.3	3.2	0.6	3.6	6.4	1.5	2.2	5.9	0.6
7	4.6	4.6	0.6	7.1	9.0	5.6	5.7	8.9	4.3	1.2	1.7	0.2	1.9	4.4	0.0	1.4	4.1	0.3
8	2.8	2.2	0.2	6.1	6.7	5.6	3.3	4.9	2.7	0.5	0.7	0.1	1.4	3.0	0.2	1.0	3.0	0.1
9	2.4	2.4	0.1	3.9	4.8	3.1	3.3	5.4	2.4	0.6	0.9	0.0	1.0	2.3	0.0	0.8	2.4	0.1
10	1.7	1.8	0.1	2.2	2.8	1.8	2.8	5.0	1.9	0.4	0.6	0.1	0.4	0.7	0.2	0.3	1.1	0.0
11	1.0	0.8	0.0	1.1	1.8	0.5	1.6	2.4	1.2	0.2	0.2	0.0	0.7	1.2	0.4	0.2	0.5	0.0
12	1.1	1.0	0.0	1.7	3.0	0.7	1.4	2.2	1.1	0.1	0.2	0.0	0.3	0.7	0.0	0.2	0.7	0.0
13	0.5	0.5	0.0	1.5	1.6	1.5	1.1	1.9	0.7	0.1	0.2	0.0	0.5	1.2	0.0	0.3	0.8	0.1
14	0.7	0.7	0.0	0.9	1.4	0.5	0.6	1.2	0.4	0.1	0.1	0.0	0.1	0.2	0.0	0.1	0.4	0.0
15	0.4	0.5	0.0	1.1	1.2	1.1	0.8	1.6	0.5	0.0	0.1	0.0	0.3	0.7	0.0	0.1	0.3	0.0
16	0.3	0.3	0.0	0.7	1.2	0.4	0.4	1.0	0.2	0.1	0.1	0.0	0.2	0.5	0.0	0.0	0.1	0.0
17	0.2	0.2	0.0	0.1	0.2	0.0	0.7	1.7	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18	0.2	0.2	0.0	0.4	0.7	0.2	0.3	0.7	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
19	0.3	0.3	0.0	0.2	0.5	0.0	0.3	0.7	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20+	0.8	1.0	0.0	1.6	1.8	1.4	1.6	3.8	0.7	0.0	0.0	0.0	0.2	0.5	0.0	0.0	0.9	0.0

Table 5: Percent distribution of last action on censored cases.

	AP	LS
Seemingly out of scope		
Not an HU	0.3	0.1
Deceased	0.0	0.1
Initial contact difficulty		
HU not accessible	0.4	0.1
Temporarily unlocatable	1.5	5.2
Temp unlocatable--business address	0.0	0.4
Language barrier, Spanish needed	0.8	0.0
Other language barrier	0.4	0.2
New lead/resource	1.3	4.7
Lead/resource failed	3.3	1.8
No contact/no progress	9.5	20.3
Initial contact made		
R/informant to call	0.4	1.7
R/informant requests special letter/call	0.9	2.3
Appointment--exact time set	0.3	0.4
Failure to complete screener		
Broken appointment (not rescheduled)	1.6	1.0
Informant/gatekeeper refusal	4.7	6.7
Refusal conversion in progress	4.1	5.5
Failure to complete questionnaire		
Screener completed	0.1	0.2
No further contact/progress	6.2	3.9
Appointment, exact time set	0.4	0.2
Broken Appointment (not rescheduled)	3.3	1.2
Temporary soft refusal	5.5	8.0
Temporary hard refusal	33.6	24.1
Temporary break-off during interview	0.9	1.5
Temporary unlocatable	1.0	0.2
Soft refusal--conversion in process	5.5	3.2
Hard refusal--conversion in process	12.5	5.5
Break-off interview--conv. in process	0.5	0.2
Final Refusal - No conversion attempted	0.0	0.1
Final Refusal - Conversion attempted	0.3	0.3
Other	0.7	0.7

It is almost impossible to determine, even by close examination of the traces of information remaining for individual cases, how likely the censored cases would have been to be completed had additional effort been applied. The formal model presented above suggests that the censored cases should become increasingly like the marginal refusals as the field period progresses, though overall the two groups may differ. Examination of the disposition code recorded in the call records for the last step taken before attempts were suspended indicates that about 70 percent of censored AP cases and almost half of the list sample cases were behaving in such a way that a permanent refusal was imminent (table 5). Substantial fractions also appear to have been difficult to locate or contact. About 1 percent of AP cases and about 1.5 percent of list sample cases had started some phase of the interview process but broke off the interview and could not be rescheduled to complete it; from the available evidence, it is doubtful that many of these suspensions were made during the actual main interview, but the call record data are insufficient to make any finer discrimination.

For the final refusals, the recent prior case history (not shown) is, unsurprisingly, heavily weighted toward various degrees of refusal. In contrast, an examination of previous call entries for completed cases shows a dominant pattern of appointments and other indications of

cooperation. However, 33.8 percent of completed AP cases refused at some point during their evolution, versus 70.7 percent of ultimately censored cases.

The relative rate of contact with the sample cases appears to be one important difference between the refused and censored cases. About 12 percent of censored cases had no contacts at all, compared with 2.5 percent of the cases treated here as final refusals.

A probit model for each of the two samples was used to search for other systematic differences between the groups of cases that ultimately refused or were censored (table 6). The independent variables represent the aspects of the sample design; regions of the country; and characteristics of the sample address and surrounding neighborhood, some drawn from interviewers' observations and others from census tract-level data matched to the sample data. Estimates of the model are given separately for the AP and LS cases.[9] According to these models, there were some significant differences between the two response status groups. There were significant regional differences for both samples. AP

Table 6: Probit models of propensity to remain censored at the end of the field period, given that the final disposition is either refused or censored; AP and LS cases.

		AP		LS				AP		LS	
							3	-0.111		0.269	.
								0.127		0.190	.
Intercept		1.022	*	0.886	*		1	0.149		0.209	.
		0.478		0.364	.			0.181		0.350	.
LSSTRAT	7	.		0.330	+		0	.		0.042	.
		.		0.171	.			.		0.272	.
	6	.		0.239	+	SPACING	3	0.024		-0.180	.
		.		0.142	.			0.144		0.117	.
	5	.		0.283	+		2	-0.088		-0.153	.
		.		0.146	.			0.093		0.102	.
	4	.		-0.155	.		0	.		0.447	.
		.		0.147	.			.		0.327	.
	3	.		-0.506	#	OBSTACL	0	.		0.376	*
		.		0.151	.			.		0.186	.
	2	.		-0.707	#		1	0.894	*	0.319	+
		.		0.163	.			0.428		0.171	.
PSUTYP	1	-0.017		0.617	.		3	-0.302		-0.282	.
		0.136		0.391	.			0.401		0.209	.
	2	0.041		0.354	#		4	0.514	#	0.057	.
		0.107		0.077	.			0.168		0.126	.
REGION	4	-0.866	#	-0.223	+	P_LE17		-0.017		0.000	.
		0.161		0.133	.	.		0.011		0.006	.
	3	-0.183		-0.453	#	P_GE65		0.007		0.005	.
		0.147		0.127	.			0.008		0.005	.
	2	-0.7627	#	-0.758	#	PERCMED		0.003	+	0.001	.
		0.137		0.124	.			0.002		0.001	.
NYC		-0.582	#	-0.123	.	PHHWPAI		0.015		-0.014	.
		0.218		0.135	.			0.020		0.014	.
LA		0.146		-0.182	.	POWNOCC		0.001		0.002	.
		0.206		0.116	.			0.004		0.002	.
NEIBLDG	0	.		-0.267	.	MAGEHU_		-0.004		0.001	.
		.		0.433	.			0.004		0.002	.
	1	-0.023		-0.284	.	PMINOR		0.010	#	0.007	#
		0.197		0.220	.			0.003		0.003	.
	2	0.135		-0.256	.	PNOENG		-0.001		-0.002	.
		0.207		0.238	.			0.012		0.009	.
BLDGCON	0	.		-0.564	.	PHISP		-0.005		-0.007	+
		.		0.419	.			0.005		0.004	.
	1	-0.227		0.350	.						
		0.255		0.283	.						
	2	-0.294		-0.167	.						
		0.214		0.181	.	N		1,132		2,728	
TYPBLDG	4	-0.604	#	0.042	.	Log(likelihood)		-673		-1,201	
		0.152		0.182	.						

P-values: #: ≤1%, *: ≤5%, +: ≤10%
Standard errors are given in italics below each parameter estimate.

[9]A relatively small set of interviewer observations is missing, and because of errors in some locational data, some tract-level data could not be matched. This limitation applies in approximately the same degree to other models reported in this paper that use these data.

Variable definitions for models shown in tables 6, 8, 9, and A1.

Variables used in tables xx, xx, and A1.
LSSTRAT: LS stratum (1 is omitted category).
PSUTYP: Overall urbanization of PSU: 1=non-MSA, 2=non-self-representing MSAs (self-representing MSAs is the omitted category).
REGION: Region of the country: 2=north central, 3=south, 4=west (northeast is the omitted category).
NYC: Observation located in the New York City PSU.
LA: Observation located in the Los Angeles PSU.
NEIBLDG: Types of buildings in the neighborhood of the sample address: 0=interviewer did not see the sample address, 1=all residential, 2=mostly residential (omitted category is half or more nonresidential).
BLDGCON: Condition of unit at sample address relative to others in the neighrorhood: 0=interviewer did not observe sample address, 1=others better, 2=about the same (omitted category is others not as good).
TYPBLDG: Type of building at sample address: 0=interviewer did not see sample address, 1=mobile home, 3=building has 2 to 9 units, 4=building has 10 or more units (omitted category is single-family building).
SPACING: Spacing of units in neighborhood: 0=interviewer did not observe the sample address, 2=21 to 100 feet apart, 3=greater than 100 feet apart (omitted category is 20 feet apart of less).
OBSTACL: Obstacle to reaching the sample address: 0=interviewer did not observe the sample address, 1=doorman or guardhouse, 3=other "gatekeeper" at the sample address, 4=locked lobby or locked gate (omitted category is no such obstacle).
P_LE17: Percent of census tract population age 17 or less.
P_GE65: Percent of census tract population age 65 and older.
PERCMED: Median income of the census tract as a percent of the area median income.
PHHWPAI: Percent of households in the census tract receiving public assistance.
POWNOCC: Percent of housing units in the census tract that are owner occupied.
GE5UNITS: Percent of housing units in the census tract in buildings with 5 or more units.
MAGEHU: Median age of housing units in the census tract.
PMINOR: Racial and ethnic minorities as a percent of the population of the census tract.
PNOENG: Percent of the census tract population that speak English "poorly" or "not at all."
PHISP: Percent of the census tract population reporting Hispanic origin.

Variables used only in table xx:
DAYS: Number of days from the first attempt on a case to the current attempt.
ATTEMPT: Number of attempts made from the first attempt on a case to the current attempt/.
CONTACT: Number of contacts made from the first attempt on a case to the current attempt.
EVCONT: An indicator for whether a case has been contacted at all as of the current attempt.
EVREF: An indicator for whether the respondent has ever been uncooperative as of the current attempt.

cases in large apartment buildings were more likely to be censored than to be recorded as final refusals; those living in buildings with a locked lobby or doorman, in neighborhoods with relatively high incomes or with larger proportions of non-Hispanic minorities were more likely to remain censored. For the LS cases, the observations in the sample strata more likely to be wealthy were more likely to be censored, as were those who lived in a building with a doorman, cases where the sample address was not observed, and those who lived in neighborhoods with larger proportions of non-Hispanic minorities. LS cases in neighborhoods with larger proportions of Hispanics were less likely to have censored outcomes. None of the differences have an obvious explanation, but the significance of so many factors indicates the presence of some underlying decision structure that may have varied across field managers and interviewers.

If the application of effort to cases were either random or independent of the expected outcomes, the empirical hazard rates for cases at risk being completed or permanently refused at each application of effort might be used to estimate the expect cost of a given response rate and the length of a field period. But as is no doubt abundantly clear at this point, the choice and the outcome are interrelated—the choice to apply effort comes before an interaction with the respondent, and this choice is affected by the subjective probability of completing an interview. One might model jointly the decision to pursue a case and the likelihood of its completion. However, because all the variables that are available for analysis might well enter into both processes (other than the outcome, there is no systematically available information that became available only after each attempt was made), such a model cannot be not statistically identified

(except through functional form restrictions). Nonetheless, useful information may still be gained by closer examination of effort and nonresponse.

If response probabilities are well assessed and effort is allocated rationally and without constraint on the distribution of cases within monitored outcome grous, one would expect that as more effort is devoted to a sample, the relatively easy cases would be interviewed early, the very resistant cases would refuse firmly, and the remaining cases would become increasingly rich in those that are inclined to refuse; consequently, an increasingly large fraction of cases pursued should ultimately refuse firmly. But at the surface, the data show a different pattern. Over the course of attempts during the field period, the proportion of all cases at risk at each point that ultimately refuses is roughly constant between about 12 and 15 percent (table 7); this result also holds separately for the AP and LS cases. At the same time, the proportion of cases ultimately completed declines gradually as the rate for cases that are ultimately censored rises; throughout, AP cases have a lower fraction of ultimately censored cases and a correspondingly higher

Table 7: Number of cases at risk and percent of all cases at risk ultimately having disposition codes indicating that they were completed, refused or censored; by number of attempts, on-site attempts, contacts, and on-site contacts.

	Attempts								Contacts							
	Any type				On-site				Any type				On-site			
#	At risk	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk	Comp	Ref	Cen
1	8,598	51.9	11.7	36.4	7,863	53.2	11.4	35.4	8,185	54.3	11.9	33.7	6,714	58.3	11.2	30.5
2	8,384	50.9	12.0	37.1	6,233	52.0	11.7	36.3	6,861	53.0	13.3	33.8	4,193	57.1	12.6	30.3
3	7,943	49.1	12.4	38.4	4,665	50.3	12.5	37.2	5,331	49.9	14.5	35.6	2,403	51.8	15.4	32.8
4	7,314	46.8	13.0	40.2	3,467	48.6	13.2	38.2	4,021	47.0	15.2	37.8	1,427	48.0	17.3	34.7
5	6,633	44.6	13.5	41.9	2,572	47.0	13.7	39.3	3,068	44.8	16.1	39.2	875	45.5	17.7	36.8
6	5,914	43.0	13.7	43.3	1,938	46.8	13.9	39.2	2,337	43.8	15.3	40.9	564	43.4	19.1	37.4
7	5,257	41.4	13.9	44.7	1,477	46.7	14.0	39.3	1,791	42.5	15.9	41.6	359	40.1	19.8	40.1
8	4,629	40.7	14.0	45.3	1,146	46.3	14.1	39.6	1,337	41.5	16.0	42.5	244	38.1	20.9	41.0
9	4,045	40.1	13.5	46.3	880	45.3	14.7	40.0	1,049	41.2	14.7	44.1	179	40.2	20.7	39.1
10	3,532	39.8	13.7	46.4	677	44.8	16.1	39.1	804	40.7	14.4	44.9	119	38.7	22.7	38.7
11	3,038	39.1	13.2	47.6	521	43.6	16.7	39.7	619	40.5	15.2	44.3	88	33.0	26.1	40.9
12	2,639	38.8	13.1	48.1	407	43.2	17.0	39.8	515	40.2	16.1	43.7	69	31.9	23.2	44.9
13	2,286	38.5	13.3	48.2	319	42.3	17.2	40.4	405	39.3	16.0	44.7	55	30.9	23.6	45.5
14	1,963	37.9	13.6	48.5	258	44.6	15.9	39.5	332	40.7	15.1	44.3	37	32.4	21.6	45.9
15	1,687	37.0	13.6	49.4	195	40.5	14.9	44.6	270	37.8	15.2	47.0	28	28.6	25.0	46.4
16	1,435	36.9	13.0	50.0	149	38.9	15.4	45.6	215	38.6	14.0	47.4	20	30.0	20.0	50.0
17	1,263	36.5	12.8	50.8	121	38.0	15.7	46.3	182	38.5	12.6	48.9	13	15.4	15.4	69.2
18	1,092	35.5	12.8	51.7	102	37.3	14.7	48.0	151	40.4	14.6	45.0	12	8.3	16.7	75.0
19	937	34.8	12.1	53.1	91	39.6	13.2	47.3	130	40.0	13.8	46.2	10	0.0	20.0	80.0
20	818	34.9	11.0	54.1	75	38.7	16.0	45.3	105	37.1	15.2	47.6	10	0.0	20.0	80.0

completion rate than LS cases. The pattern is remarkably similar for on-site attempts and all types of contacts. In contrast, for on-site contacts, the refusal rate rises in tandem with the rate of ultimate censorship over the range where there are a non-negligible number of cases still at risk.

Overall, it is clear that continued effort yields an increasing share of ultimate non-interviews (refused and ultimately censored cases) along with a declining payoff in terms of completed cases. The unexpectedly flat refusal rates for the first three measures may reflect reluctance of interviewers and mangers to "give up" on cases, even when the probability of success appears low; among other things, they may think that some of those cases might be "needed" later to meet production quotas.

The choice whether to continue exerting effort on a case is clearly a key factor in the determination of outcomes and the distribution of cases within outcomes. One way of extracting systematic information about the choice to continue effort is to frame the decision as a hazard model. In such a model, the unit events are the elements of the sequences of decision across all cases remaining at risk whether to expend further effort on the case, or to leave the outcome permanently censored. Once a case is completed, refused or permanently censored, it adds no further elements. The choice element is to pursue a case further or to allow the outcome to be permanently censored. Because the model has only two choices, estimation may be performed using a simple logit model of the stacked sequences of decisions.

A separate model was estimated for each sample type using interviewer observations, census tract-level data, and case administration data derived from the call records (table 8). The case administration variables include information specific to each decision point: the number of days a case was in play as of the previous attempt, the number of prior attempts made, an indicator variable for whether contact had ever been made previously, the number of prior contacts, and an indicator for whether the working dispositions codes record any prior refusal by the respondent to participate. Both models show strongly that more days in play, greater numbers of contacts, and a prior refusal lower the frequency with which cases were followed. For the AP cases, the positive effect of the number of prior attempts on the likelihood of continued attempts, which probably captures the repeated calls necessary to make initial contacts, is quickly offset by the negative effect of days in play; the fact that the number of

attempts is not a significant factor for the LS cases suggests that there were deeper differences either in the way in which such cases were worked or in the reactions of respondents.

Even with the administrative controls, other variables also show evidence of significantly different applications of effort across cases.[10] There were strong, but different, geographic effects for the two samples. For the AP cases, those living in mobile homes were less likely and those in apartment buildings were more likely to be followed than those living in single-family homes; those living in areas in areas with larger Hispanic populations were more likely to be followed than those living in other neighborhoods, but the converse was true for cases in neighborhoods with higher fractions of people with limited skills in speaking English; cases in neighborhoods of moderately widely-spaced houses were more likely to be pursued than either cases in more densely or sparsely built neighborhoods. Cases in areas with higher levels of income or higher proportions of people aged 65 and older were less likely to be pursued; in light of the positive correlation of both age and income with wealth, these two

Table 8: Probit models of propensity to follow an observation; AP and LS cases.

		AP		LS				AP		LS	
						SPACING	0	.		-0.241	
								.		*0.149*	
Intercept		5.664	#	4.276	#		3	-0.032		0.088	
		0.575		*0.230*				*0.115*		*0.080*	
LSSTRAT	7	.		-0.174	+		2	0.168	*	-0.027	
		.		*0.093*				*0.084*		*0.074*	
	6	.		-0.3047	#		0	.		-0.121	
		.		*0.055*				.		*0.140*	
	5	.		-0.799	#	OBSTACL	0	.		-0.134	
		.		*0.054*				.		*0.130*	
	4	.		0.603	#		1	0.281		-0.241	*
		.		*0.089*				*0.312*		*0.112*	
	3	.		0.611	#		3	-0.768	+	0.157	
		.		*0.098*				*0.419*		*0.162*	
	2	.		0.844	#		4	-0.037		-0.058	
		.		*0.128*				*0.214*		*0.101*	
PSUTYP	1	-0.228	*	-0.413	*	P_LE17		-0.015		-0.005	
		0.108		*0.183*				*0.015*		*0.006*	
	2	-0.104		0.140		P_GE65		-0.024	*	0.008	*
		0.080		*0.098*				*0.011*		*0.004*	
REGION	4	0.217		-0.322	#	PERCMED		-0.005	#	0.000	
		0.155		*0.060*				*0.002*		*0.000*	
	3	-0.097		0.118	*	PHHWPAI		-0.007		0.023	
		0.106		*0.056*				*0.029*		*0.016*	
	2	0.389	#	0.075		POWNOCC		0.006		-0.005	*
		0.119		*0.064*				*0.005*		*0.002*	
NYC		0.136		-0.242	#	MAGEHU		0.009		-0.000	
		0.155		*0.056*		.		*0.005*		*0.002*	
LA		-0.394	*	0.162	#	PMINOR		-0.002		-0.006	#
		0.169		*0.061*				*0.004*		*0.002*	
NEIBLDG	0	.		-0.384	+	PNOENG		-0.039	+	-0.014	
		.		*0.213*				*0.020*		*0.008*	
	1	-0.189		0.071		PHISP		0.030	#	0.015	#
		0.121		*0.0935*				*0.008*		*0.003*	
	2	-0.243	+	0.182		DAYS		-0.018	#	-0.016	#
		0.130		*0.119*				*0.002*		*0.001*	
BLDGCON	0	.		0.686	#	ATTEMPT		0.036	#	-0.007	
		.		*0.216*				*0.010*		*0.005*	
	1	0.255		-0.237	+	CONTACT		-0.057	#	-0.034	#
.		*0.164*		*0.139*				*0.014*		*0.008*	
	2	0.014		-0.067		EVCONT		0.160		0.042	
		0.120		*0.095*				*0.120*		*0.038*	
TYPBLDG	-4	0.319	+	0.022.		EVREF		-0.846	#	-0.621	#
		0.172		*0.141*				*0.095*		*0.032*	
	-3	0.362	#	0.011							
		0.132		*0.157*							
	-1	-0.629	#	0.227		N		32,573		40,975	
		0.179		*0.288*		Likelihood ratio		3,185		10,190	

P-values: #: ≤1%, *: ≤5%, +: ≤10% Standard errors are given in italics below each parameter estimate.

[10]Omission of these administrative variables causes very little qualitative change in the other estimates.

factors could be taken to imply a bias in key SCF estimates made with the AP sample alone.[11] One might expect that barriers to contacting the respondent would have a substantial effect, but only the presence of a "gatekeeper" (typically, an employee of the respondent, rather than a literal gatekeeper) has a significant deterring effect on following AP cases.

For the LS cases, there are significant differences in the likelihood of following cases according to their sample stratum, with the strata most likely to be wealthy having the lowest propensity to be followed. Such cases living in a house in worse condition than others in their neighborhood were less likely to be pursued; where the interviewer did not observe the neighborhood, cases were more likely to be followed. Where there was a doorman, LS cases were less likely to be pursued. Those living in neighborhoods with higher fractions of people aged 65 and older and those with higher fractions of Hispanics were more likely to be followed; those in neighborhoods with higher fractions of owner-occupied housing or higher fractions of all types of minorities were less likely to be followed.

Although the sketchy data available for respondents and nonrespondents make it very difficult to coax out a clear structural interpretation of the decision making process in pursuing cases, the models do suggests that there were systematic patterns in the allocation of effort. If variations in effort are not offset by opposite variations in the frequency with which respondents are persuaded to complete an interview, then the distribution of outcomes would be skewed away from the population distribution.

---

[11]In practice, the effect of any bias should be mitigated in estimates using the full sample. Post-stratification is used to align the sample age distribution to population totals. By design, the SCF weights are constructed such that the upper tail of the distribution of wealth is largely determined by the LS, for which good external income controls are available (see Kennickell and Woodburn [1999]).

A simple probit model of case completion using all observations and the same non-administrative variables shows that some of the systematic effects in the application of effort remain, but there are also other effects that more likely reflect the difficulty of contacting and persuading respondents (table 9). For the AP cases, two key factors on which the models agree are lower effort and response among respondents who have a "gatekeeper" or who live in neighborhoods that have relatively high median incomes. For the LS cases, the key agreements are lower effort and response among cases in the strata most likely to be wealthy and among those living in neighborhoods with higher proportions of minorities.

IV. Alternative case management strategies

The ultimate goal of survey field operations is to collect data that represent the target population as efficiently and with as little bias as possible. Unfortunately, it is generally highly

Table 9: Probit models of propensity to complete an observation; AP and LS cases.

		AP		LS				AP		LS	
							3	0.148	*	-0.019	
								0.067		0.121	
Intercept		0.340		-0.071			1	0.044		-0.565	*
.		0.261		0.252				0.094		0.276	
LSSTRAT	7	.		-0.587	#		0	.		-0.846	#
		.		0.131				.		0.127	
	6	.		-0.259	#	SPACING	3	0.126	+	-0.007	
		.		0.099				0.074		0.075	
	5	.		-0.083			2	0.015		-0.232	#
		.		0.100				0.050		0.067	
	4	.		0.064			0	.		-0.051	
		.		0.102				.		0.173	
	3	.		0.069		OBSTACL	0	.		-0.789	#
		.		0.106				.		0.117	
	2	.		0.229	*		1	-0.174		-0.442	#
		.		0.113				0.199		0.117	
PSUTYP	1	0.238	#	0.342			3	-0.685	#	-0.566	#
		0.072		0.213				0.264		0.179	
	2	0.146	*	0.105	*		4	-0.214	*	-0.324	#
		0.058		0.054				0.091		0.091	
REGION	4	0.005		0.153	+	P_LE17		0.010	+	0.008	
		0.085		0.089				0.006		0.005	
	3	-0.010		0.212	*	P_GE65		-0.002		-0.001	
		0.077		0.085				0.004		0.003	
	2	-0.088		0.054		PERCMED		-0.002	*	0.000	
		0.073		0.085				0.001		0.000	
NYC		0.357	#	0.157	+	PHHWPAI		0.003		0.007	
		0.114		0.090				0.010		0.010	
LA		-0.106		-0.180	*	POWNOCC		-0.001		-0.001	
		0.119		0.090				0.002		0.002	
NEIBLDG	0	.		-0.678	#	MAGEHU		0.001		0.004	*
		.		0.240				0.002		0.002	
	1	-0.177	+	-0.090		PMINOR		0.002	+	-0.004	*
		0.101		0.146				0.001		0.002	
	2	-0.186	+	-0.216		PNOENG		-0.021	#	-0.007	
		0.106		0.157				0.007		0.007	
BLDGCON	0	.		1.371	#	PHISP		0.002		0.000	
		.		0.244				0.003		0.003	
	1	0.342	#	0.104							
		0.132		0.173							
	2	0.172		0.129							
		0.113		0.127		N		3,988		4,238	
TYPBLDG	4	0.177	*	0.123		Log(likelihood)		-2,323		-2,408	
		0.081		0.123							

P-values: #: ≤1%, *: ≤5%, +: ≤10% Standard errors are given in italics below each parameter estimate.

unlikely that every respondent selected will agree to participate. In the absence of specific guidance, interviewers and their managers will perform an 'implicit stratification" of the sample through their decisions to apply effort to the set of cases available to be worked throughout the field period. Thus, a very pressing question is: What guidance on individual case management can we give to interviewers and their managers to help them reach the statistical goals of a survey? To respond, we need both a framework for classifying cases in terms that reflect the statistical objectives of the survey and a mechanism for transmitting sufficiently precise information to and from the field.

Previous SCF efforts late in the field period have typically been targeted to even out large differences in response rates across PSUs for AP cases, and to achieve specific numbers of completed interviews within the sample strata for LS cases. Detailed investigation of nonresponse issues in the SCF led to the use of various post-strata at the weighting adjustment stage to address a set of potential biases (Kennickell and McManus, 1993), and that practice has been refined over time (Kennickell and Woodburn, 1999). However, there has never been any previous effort in the SCF to develop a more detailed case management plan to address potential bias and efficiency issues *during* the field period. In essence, interviewers and managers were allowed to pursue their own "convenience" in case management.

The information available to make management decisions about a sample is often patchy at best. When a sample is first selected, some characteristics of the elements may be known from the sample frame; some additional information might be matched from other sources. As interviewers visit the cases, call records are generated and some other information about the cases is revealed; some such information may be captured systematically. However, much "local" information may be so idiosyncratic as to be difficult to use systematically or insufficiently salient to be noticed in all relevant instances by all interviewers. Because local information may sometimes bear on the evaluation of interviewers' performance, it may also be important to consider ways to manage interviewers' incentives so that they would be willing to reveal such information. As field operations progress, more detailed information becomes available on the set of sample elements that actually complete an interview.

Ideally, in order to reduce bias or estimation variance, cases would be classified dynamically through the field period on the basis of all available information into ones that

should be disproportionately targeted and those that should not. Loosely speaking, cases believed to be "like" existing cases or to show low variability in terms of *a priori* unobservable characteristics within important *a priori* observable groups would be subsampled, and cases believed to be "unlike" existing cases or relatively variable in *a priori* unobservable characteristics within important *a priori* observable groups would be targeted with relatively more effort. Informational and cost limitations inevitably force a compromise.

There are very many possible formal strategies. Each strategy (including the one of allowing interviewers to persist in traditional minimally guided behavior) entails some sort of "model" of what is known and controllable in a sample. In the classical sampling perspective of Hansen and Hurwitz [1946], at some point in a field process, uncompleted cases are randomly subsampled. By forcing effort more intensively onto a smaller number of cases, the idea is that more could be learned about the nonrespondent population (reduced bias) at the cost of some direct variance inflation, but with lower mean squared error if the level of subsampling can be calibrated sufficiently. Depending on ultimate response goals and differences in quantity and reliability of the available information, one might extend this model to differential subsampling rates for different subpopulations. Sudman [1966] offers another perspective. As in the Hansen and Hurwitz model, there is an initial probability sample that experiences nonresponse. Here subsampling is performed ideally using the probability of nonresponse; those with lower probabilities are oversampled and those with higher probabilities are undersampled. Usually the operational implication is taken to be the generation of "quotas" for field staff of certain classes of cases. Although bias reduction would lead to a direct reduction in mean squared error, the direct variance implications of the subsampling are not straightforward, but depend on the interpretation of the operation. A classical interpretation implies variance inflation through increased variability of weights, while a strong model-based interpretation assuming a credible mapping from groups of participants to nonrespondents would not necessarily imply any such variance inflation.

There are many other arguable approaches to subsampling. For example, as suggested earlier, one could use traditional stratification arguments to sample differentially observable groups discovered during the field period to have strongly differing variances for key variables. If much is known about the nonresponse mechanism when a sample is first selected, differential

sampling at that point (or the creation of reserve replicates to allow more control of differential sampling later) could lead to efficiency improvements and bias reduction. Clearly, there are many other possibilities blending many of these arguments and others. All subsampling plans should also be examined in light of post-survey adjustment, such as post-stratification, that might otherwise be made or be made to larger effect in the absence of subsampling.

One factor which may conflict with straightforward sample management plans is the drive to make a credible level of effort to inform every selected sample element of the nature of the survey and the respondent's role in the process. Informed refusal (at least taken to the limits of something less than a "hard" refusal) seems as large an ethical concern as informed consent. Furthermore, the effect of a lower standard of work on the behavior of interviewers and their managers could also be serious. The key sense of legitimacy field staff require to persuade respondents might be undermined. Perhaps more seriously, by signaling to interviewers a lower importance of interviewing cases in general, it seems almost certain that new selectivity effects would be induced on survey participation.

However, a structured initial case management plan entailing significant work on all cases could serve reasonably as the first part of a two-phase sample management plan. In the first phase, all sample observations in the original sample (holding aside the issue of possible replicate management) would be subjected to a specified level of effort which would play out through a series of alternative branches depending on the difficulties in contacting or persuading respondents to participate. There are two important informational benefits of enforcing this phase of work uniformly. First, because the endogeneity between the application of effort and expectations of success would be broken, it would be possible to make more meaningful estimates of respondents' propensity to cooperate. Second, more uniform and reliable case-specific information would be available. Together, this information could be used to target resources in a second phase to achieve a bias reduction or an improvement in statistical efficiency. With sufficient information and resources, such targeting could proceed dynamically through the remaining field period. A very important side benefit of the phased approach is that effort should become more predictable and controllable, and thus, costs should also become more predictable and controllable.

It is not possible with the available data to test the phased approach in a literal sense, and the cost of a controlled field experiment far exceeds the budget that is available to the SCF at this time. It is, however, possible to perform a rough simulation. For convenience, attention is restricted to the AP sample and a very simple model of sample management. Ideally, the first phase would be specified in terms of the sequences of effort appropriate to the problems associated with the full range of cases—i.e., difficulty of contact, intervening vacations, broken appointments, etc. Unfortunately, the existing call record data are too weak to make such a specification feasible. As a working approximation, eight attempts of any type is taken to be the outer limit of the first phase. Within this level of effort for AP cases in the 2001 SCF, 57.4 percent of cases acquired a final disposition—45.9 percent of all cases were completed (67.4 of all completed cases), 4.4 percent of all cases were permanently refused (43.5 percent of all such refused cases), and 7.1 percent of all cases were left permanently censored (32.8 percent of all censored cases). For purposes of this example, all cases permanently censored within this range of attempts are treated as final refusals, and thus unavailable for further attempts.

For simplicity, assume that income and location, which stood out in the overall response model for AP cases reported above, are the critical factors determining nonresponse and that respondents and nonrespondents are sufficiently homogeneous within the cross-product of these variables that they can be considered the same in distributional terms. Thus, in this model, the role of subsampling is to adjust the proportional allocation of the sample across this cross-product space. Three income groups are specified in terms of the median income of the census tract where the sample element is located relative to the median income of the larger area; the break points for the groups are set at 80 percent and 120 percent. Location is specified in terms of whether the element is located in a self-representing PSU, another MSA, or a non-MSA area. The nine cells that form the cross-product of the income and location groups are the controlled

Table 10: Number of all cases, number resolved at ≤8 attempts, number unresolved at ≤8 attempts, number completed at ≤8 attempts, response rate at ≤8 attempts, additional cases needed to achieve 68.1% response rate, estimated response rate by completion, inverse of estimated response rate, number of cases to subsample, and subsampling rate; by type of PSU and median tract income relative to local area.

Area *Med. tract Inc as % of local median*	*All cases*	*Resolv. ≤8 att.*	*Unres. ≤8 att.*	*Comp. cases ≤8 att.*	*Resp. rate ≤8 att.*	*Add'l cases needed*	*Est'd final resp. rate*	*Inverse est'd resp. rate*	*Sub-. samp.*	*Sub-samp. rate*
SR MSA										
MI≤80%	165	115	50	91	55.2	21	81.8	1.223	26	52.2
MI≤120%	348	187	161	149	42.8	88	63.5	1.575	139	86.1
MI>120%	455	249	206	176	38.7	134	57.4	1.743	233	113.3
Other MSA										
MI≤80%	658	440	218	362	55.0	86	81.6	1.226	106	48.4
MI≤120%	989	568	421	461	46.6	213	69.1	1.447	307	73.0
MI>120%	609	321	288	245	40.2	170	59.7	1.676	284	98.8
Non-MSA										
MI≤80%	119	72	47	63	52.9	18	78.5	1.274	23	48.9
MI≤120%	443	252	191	215	48.5	87	72.0	1.389	120	63.1
MI>120%	493	254	239	203	41.2	133	61.1	1.638	217	91.0
All	4,279	2,458	1,821	1,965	45.9	949	68.1	1.468	1,394	76.5

groups (table 10).

As a guide to the necessary calculations, the lowest income group in the self-representing PSUs is taken as an example. The full sample for this group contained 165 observations, of which 115 were resolved as complete, refused or permanently censored by eight or fewer attempts, leaving 50 cases as a basis for further work. Completed cases amounted to 91 of the resolved cases, an response rate of 55.2 percent. To achieve an overall response rate of 68.1 percent (the overall level achieved for the AP sample in 2001), an additional 21 cases are needed. A key figure needed for this subsampling is an estimate of the final response rate for the cell in the absence of subsampling. Assuming the overall average response rate ultimately reaches 68.1 percent and that the rate for each cell as of eight attempts increases proportionally over the remaining field period, the forecast for this cell is 81.8 percent—the actual rate times the ratio of 68.1 to 45.9, the overall average rate as of eight attempts. Thus, to achieve in expectation the 21 additional interviews, 1.222 times as many (26) cases would need to be approached, a subsampling rate from the cases remaining after eight attempts of 52.2 percent.

In this example overall, there is large variability in the subsampling rates. For one group, the highest income group in the self-representing MSAs, the rate exceeds 100 percent; in such

cases, one would either accept a corner solution at 100 percent and make *ex post* adjustments through weighting, or release an additional sample replicate reserved for this possibility. The second phase could be stopped short of completion to re-target effort. At each such step, a calculation like that shown in table 10 would guide the application of effort for the next sub-phase.

Obviously, this model is overly simple. If the factors explaining nonresponse were available at the time of sampling as they would be in the example, it would make more sense to rely on initial stratification and perhaps more minimal adjustments within the field period based on additional information. However, the largest issue is that the root causes of nonresponse are not normally so well-defined. Usually the most one could hope for is a moderately predictive model based on proxies for the key factors.

One very important issue in moving from a model in which variations in effort are largely ignored (though probably not statistically ingorable) to one in which effort is systematically controlled is that the model of control becomes observable and must be defended directly. The Hansen and Hurwitz model of subsampling, which has the advantage of not requiring any assumptions about the distribution of nonrespondents, can nest fully within the original probability structure of a sample; pursued with sufficient vigor, this approach may reduce some nonresponse biases. But often at least something is known about the sources of nonresponse, and if that information is sufficiently reliable, one should be able to gain by incorporating it into the sample management.[12] One way of incorporating such information is to start with the framework of the Hansen and Hurwitz model and subsample disproportionately as required to offset nonresponse along dimensions believed to be important for nonresponse. If at the end of the field period the evidence is believed to be strong that the subsampled population differed from the earlier respondents in key ways, then a classical subsampling-adjusted weight could be taken to apply. At the other extreme, if the populations within the subsampled groups

---

[12]Use of such information might reduce the need for post-survey weight adjustments to deal with key sample imbalances, where similar assumptions are required. In post-stratification (Little [1993]), for example, one assumes essentially that members of an adjustment cell are a random sample from the full set of that conditional population. If such a distinction could be made at the weighting stage, it seems only reasonable to suppose it might also be made during a field period is the necessary information is available at that point.

were believed to be identical, then no such adjustment would be required. In practice, something intermediate seems more likely to reflect reality, but a formal framework would be need to be developed to support the choice of an optimal intermediate adjustment.

V. Conclusion and future research

Typically, we care about nonresponse because of its implications for bias and inefficiency in the estimation of key survey statistics. Nonresponse is a joint product of the degree to which respondents can be persuaded to participate in an interview and the amount of effort expended in the effort to gain cooperation. One root of such problems is in the respondent, and thus cannot be controlled directly. Generally, persuasion and information come to respondents from an interviewer, or from materials sent to respondents. But unless supplemental information is available, variations in the effort spent in persuasion would be indistinguishable *ex post* from variations in respondents' behaviors.

This paper uses data from the administration of the 2001 SCF to look at the distribution of effort in that survey, and it attempts to draw some conclusions for future practice. Several things emerge clearly. First, there was very substantial variability in the efforts devoted to cases; this variability appears to exceed any reasonable bounds of simple measurement error in the administrative records. Second, the application of effort is correlated with some potentially important characteristics of respondents, even when there are controls for the level of difficulty. Finally, although there is insufficient information to disentangle fully the application of effort and respondents' reactions to effort, the data do indicate that variations in effort have consequences for the distribution of outcomes. The results of the investigation suggest that other surveys might also benefit from a systematic evaluation of variations in effort and its implications for nonresponse.

Many factors may be important in characterizing the most pressing dimensions of nonresponse in a given survey. Generally, some such factors are very difficult or impossible to observe directly, and the structure that makes sense of all the factors is not known. In the absence of such information and structure, a very large number of strategies for the application of effort might be equally appropriate. One might simply push for the highest response rate possible, in hopes that this approach, applied over time, would yield at least time series

comparability, if not reduced bias in any given period. However, any approach that falls short of specific instructions to interviewers and their managers on how to work the sample cases risks introducing selectivity effects in the set of participants; such a realized sample would inevitably have aspects of a convenience sample. One might apply effort in proportion to the degree of respondents' resistance, though such an approach would very likely imply declining to interview some very "easy" cases and pursuing strong refusals to the point of harassment. Alternatively, if one could develop proxies for some important dimensions of nonresponse, then those proxies might be used systematically to target effort in a staged fashion over the field period.

This paper addresses some of the problems of targeted and phased effort and proposes a general approach. The first phase would lead every sample case released to the field through a process designed to inform respondents to a degree that would allow them to make an informed decision to participate or an informed initial decision to decline participation (that is, additional persuasion or refusal conversion would remain an option). This initial phase might be further controlled through the use of sample replicates that would be released as needed to meet the statistical goals. The second phase would operate more indirectly through control of the sample. A variety of ways of subsampling potentially mixed with differential initial sampling are discussed. Each approach turns at least implicitly on a model of the process that generates nonresponse and what might be done to alter the composition of the nonrespondent population. The implications of subsampling for bias reduction and inflation of estimation variance depend on the interpretation of the model.

Targeted and phased effort of the sort described here has two particularly large potential benefits. First, if the targets are meaningfully related to important nonresponse factors, this approach should tend to reduce bias and perhaps increase some aspects of estimation efficiency. Second, by providing a more structured approach to interviewing practice, it would make field activities more predictable—and most likely, more controllable—as well as ensuring that every case receives a credible minimum exposure to effort. As a subsidiary benefit, forcing a minimum level of effort on every case makes it possible to estimate meaningful models of nonresponse uncontaminated by differential effort and these models could be used to guide further work.

It is hoped that field work for the 2004 SCF will be able to proceed in a two-phased fashion: including a phase of specified effort on all cases and a phase of sample management to reduce nonresponse biases. The clearer administrative information required to implement such a strategy will also be useful in a post-survey evaluation of the 2004 procedures and the design of more refined procedures for the 2007 SCF sample.

Bibliography

Aizcorbe, Ana M., Arthur B. Kennickell, and Kevin B. Moore [2003] "Recent Changes in U.S. Family Finances: Evidence from the 1998 and 2001 Survey of Consumer Finances," *Federal Reserve Bulletin*, v. 89 (January), pp. 1-32.

Groves, Robert M., John Van Hoewyk, Frant Benson, Paul Schultz, Patty Maher, Lynetter Hoelter, William Mosher, Joyce Abma, and Anjani Chandra [2003] "Using Process Data from Computer-Assisted Face to Face Surveys to Help Make Survey Management Decisions," Paper presented at the 2003 Meetings of the American Association for Public Opinion Research.

Hansen, Morris H. and William N. Hurwitz [1946] "The Problem of Non-Response in Surveys," *Journal of the American Statistical Association*, v. 41 (December), pp. 517-529.

Kennickell, Arthur B. [2000a] "Asymmetric Information, Interviewer Behavior, and Unit Nonresponse," working paper, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B. [2000b] "Wealth Measurement in the Survey of Consumer Finances: Methodology and Directions for Future Research," paper presented at the annual meetings of the American Association for Public Opinion Research, Portland, May, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B. [1999a] "Analysis of Nonresponse Effects in the 1995 Survey of Consumer Finances,"*Journal of Official Statistics*, v. 15 no. 2, pp. 283-304.

Kennickell, Arthur B. [1999b] "What Do the 'Late' Cases Tell Us? Evidence from the 1998 Survey of Consumer Finances," paper presented at the International Conference on Survey Nonresponse, Portland, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B. and Douglas A. McManus [1993] "Sampling for Household Financial Characteristics Using Frame Information on Past Income," *Proceedings of the Section on Survey Research Methods*, Annual Meetings of the American Statistical Association, San Francisco, CA (with Douglas McManus).

Kennickell, Arthur B. and R. Louise Woodburn [1999] "Consistent Weight Design for the 1989, 1992, and 1995 SCFs, and the Distribution of Wealth," *Review of Income and Wealth* (Series 45, number 2), June, pp. 193-215.

Little, Roderick A.J. [1993] "Post-Stratification, a modeler's perspective," *Journal of the American Statistical Association*, v. 88 (September), pp. 1001-1012.

Sudman, Seymour [1966] "Probability Sampling with Quotas," *Journal of the American Statistical Association*, v. 61, pp. 749-771.

Table Ax: Mean and quantiles of the distribution of the days a case is in play and attempts; for out of scope cases by sample type.

	All	*AP*	*LS*
Days in play			
Mean	25.4	23.3	67.6
P5	0	0	7
P10	1	1	21
P25	5	5	34
P50	14	13	59
P75	29	27	98
P90	63	56	134
P95	106	99	149
P99	173	170	181
Attempts			
Mean	4.0	3.8	7.8
P5	1	1	2
P10	1	1	3
P25	2	3	4
P50	3	3	7
P75	4	4	10
P90	7	7	15
P95	10	10	22
P99	18	18	25
Number of obs.	745	711	34

Table Ax: Number of cases at risk and percent of all cases at risk ultimately having disposition codes indicating that they were completed, refused or censored; by number of attempts and on-site attempts and by sample type.

#	All types												On-site											
	All				AP				LS				All				AP				LS			
	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk
1	51.9	11.7	36.4	8,598	68.1	10.4	21.5	4,279	35.7	13.0	51.3	4,297	53.2	11.4	35.4	7,863	68.2	10.4	21.4	4,234	35.5	12.6	51.9	3,610
2	50.9	12.0	37.1	8,384	67.0	10.8	22.2	4,118	35.3	13.1	51.6	4,244	52.0	11.7	36.3	6,233	65.3	11.1	23.6	3,658	33.1	12.5	54.4	2,562
3	49.1	12.4	38.4	7,943	64.9	11.4	23.7	3,820	34.4	13.4	52.2	4,101	50.3	12.5	37.2	4,665	61.3	12.0	26.7	2,979	30.8	13.5	55.7	1,680
4	46.8	13.0	40.2	7,314	62.1	12.1	25.8	3,419	33.4	13.8	52.8	3,874	48.6	13.2	38.2	3,467	57.2	13.0	29.8	2,377	30.0	13.5	56.5	1,088
5	44.6	13.5	41.9	6,633	59.2	12.9	27.9	3,033	32.3	14.0	53.7	3,583	47.0	13.7	39.3	2,572	53.8	13.8	32.3	1,878	28.3	13.4	58.2	692
6	43.0	13.7	43.3	5,914	56.7	13.2	30.1	2,670	31.7	14.1	54.2	3,229	46.8	13.9	39.2	1,938	52.7	13.8	33.4	1,492	27.0	14.4	58.6	444
7	41.4	13.9	44.7	5,257	54.6	14.1	31.4	2,372	30.5	13.7	55.8	2,873	46.7	14.0	39.3	1,477	51.5	14.0	34.5	1,180	27.5	13.9	58.6	295
8	40.7	14.0	45.3	4,629	53.0	14.0	33.0	2,089	30.5	14.0	55.5	2,531	46.3	14.1	39.6	1,146	51.6	13.9	34.5	924	24.1	15.0	60.9	220
9	40.1	13.5	46.3	4,045	52.2	13.7	34.0	1,821	30.1	13.4	56.5	2,216	45.3	14.7	40.0	880	49.8	14.4	35.8	741	21.0	15.9	63.0	138
10	39.8	13.7	46.4	3,532	51.4	13.9	34.7	1,610	30.1	13.7	56.3	1,918	44.8	16.1	39.1	677	49.2	15.5	35.2	579	18.4	19.4	62.2	98
11	39.1	13.2	47.6	3,038	50.7	13.4	35.9	1,382	29.4	13.1	57.5	1,653	43.6	16.7	39.7	521	47.3	16.3	36.5	455	18.2	19.7	62.1	66
12	38.8	13.1	48.1	2,639	50.1	13.8	36.1	1,213	29.1	12.6	58.3	1,423	43.2	17.0	39.8	407	46.2	17.1	36.7	357	22.0	16.0	62.0	50
13	38.5	13.3	48.2	2,286	49.9	14.4	35.8	1,057	28.7	12.4	58.9	1,226	42.3	17.2	40.4	319	45.1	16.8	38.1	286	18.2	21.2	60.6	33
14	37.9	13.6	48.5	1,963	47.9	14.8	37.2	916	29.1	12.5	58.4	1,045	44.6	15.9	39.5	258	46.6	16.0	37.4	238	20.0	15.0	65.0	20
15	37.0	13.6	49.4	1,687	46.5	14.6	38.9	796	28.5	12.7	58.8	889	40.5	14.9	44.6	195	42.7	14.6	42.7	178	17.6	17.6	64.7	17
16	36.9	13.0	50.0	1,435	46.4	13.7	39.9	692	28.1	12.4	59.5	741	38.9	15.4	45.6	149	39.9	15.9	44.2	138	27.3	9.1	63.6	11
17	36.5	12.8	50.8	1,263	44.7	13.5	41.8	617	28.6	12.1	59.3	644	38.0	15.7	46.3	121	38.9	15.9	45.1	113	25.0	12.5	62.5	8
18	35.5	12.8	51.7	1,092	44.2	13.3	42.4	547	26.7	12.3	61.0	544	37.3	14.7	48.0	102	36.7	14.3	49.0	98	50.0	25.0	25.0	4
19	34.8	12.1	53.1	937	44.2	12.7	43.1	480	25.0	11.4	63.6	456	39.6	13.2	47.3	91	38.9	13.3	47.8	90	100.0	0.0	0.0	1
20	34.9	11.0	54.1	818	44.3	11.6	44.1	422	24.8	10.4	64.8	395	38.7	16.0	45.3	75	37.8	16.2	45.9	74	100.0	0.0	0.0	1

Table Ax: Table Ax: Number of cases at risk and percent of all cases at risk ultimately having disposition codes indicating that they were completed, refused or censored; by number of contacts and on-site contacts and by sample type.

#	All types												On-site											
	All				AP				LS				All				AP				LS			
	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk	Comp	Ref	Cen	At risk
1	54.3	11.9	33.7	8,185	68.9	10.4	20.7	4,232	38.8	13.6	47.6	3,953	58.3	11.2	30.5	6,714	69.4	10.3	20.4	4,053	41.5	12.6	45.8	2,661
2	53.0	13.3	33.8	6,861	65.0	12.0	23.0	3,531	40.2	14.6	45.2	3,330	57.1	12.6	30.3	4,193	64.7	12.0	23.3	2,891	40.2	13.9	45.9	1,302
3	49.9	14.5	35.6	5,331	59.9	13.6	26.4	2,754	39.2	15.5	45.3	2,577	51.8	15.4	32.8	2,403	57.0	14.7	28.2	1,817	35.7	17.6	46.8	586
4	47.0	15.2	37.8	4,021	55.6	15.1	29.3	2,095	37.5	15.4	47.1	1,926	48.0	17.3	34.7	1,427	51.6	16.7	31.6	1,166	31.8	19.9	48.3	261
5	44.8	16.1	39.2	3,068	51.9	16.2	31.9	1,645	36.5	16.0	47.6	1,423	45.5	17.7	36.8	875	48.0	17.2	34.9	757	29.7	21.2	49.2	118
6	43.8	15.3	40.9	2,337	49.4	15.8	34.8	1,283	36.9	14.7	48.4	1,054	43.4	19.1	37.4	564	45.0	18.7	36.3	513	27.5	23.5	49.0	51
7	42.5	15.9	41.6	1,791	47.1	15.7	37.1	1,023	36.3	16.1	47.5	768	40.1	19.8	40.1	359	41.1	19.8	39.1	338	23.8	19.0	57.1	21
8	41.5	16.0	42.5	1,337	45.3	15.8	38.9	766	36.4	16.3	47.3	571	38.1	20.9	41.0	244	39.1	20.2	40.8	233	18.2	36.4	45.5	11
9	41.2	14.7	44.1	1,049	45.0	14.7	40.4	627	35.5	14.7	49.8	422	40.2	20.7	39.1	179	41.3	19.8	39.0	172	14.3	42.9	42.9	7
10	40.7	14.4	44.9	804	43.7	14.6	41.7	487	36.0	14.2	49.8	317	38.7	22.7	38.7	119	39.5	21.1	39.5	114	20.0	60.0	20.0	5
11	40.5	15.2	44.3	619	42.9	15.6	41.5	378	36.9	14.5	48.5	241	33.0	26.1	40.9	88	34.1	24.7	41.2	85	0.0	66.7	33.3	3
12	40.2	16.1	43.7	515	42.6	15.7	41.7	324	36.1	16.8	47.1	191	31.9	23.2	44.9	69	32.4	23.5	44.1	68	0.0	0.0	0.0	1
13	39.3	16.0	44.7	405	41.6	14.1	44.3	262	35.0	19.6	45.5	143	30.9	23.6	45.5	55	31.5	24.1	44.4	54	0.0	0.0	0.0	1
14	40.7	15.1	44.3	332	42.3	13.5	44.1	222	37.3	18.2	44.5	110	32.4	21.6	45.9	37	32.4	21.6	45.9	37	0.0	0.0	0.0	0
15	37.8	15.2	47.0	270	39.7	13.0	47.3	184	33.7	19.8	46.5	86	28.6	25.0	46.4	28	28.6	25.0	46.4	28	0.0	0.0	0.0	0
16	38.6	14.0	47.4	215	39.3	12.7	48.0	150	36.9	16.9	46.2	65	30.0	20.0	50.0	20	30.0	20.0	50.0	20	0.0	0.0	0.0	0
17	38.5	12.6	48.9	182	39.4	11.0	49.6	127	36.4	16.4	47.3	55	15.4	15.4	69.2	13	15.4	15.4	69.2	13	0.0	0.0	0.0	0
18	40.4	14.6	45.0	151	42.3	12.5	45.2	104	36.2	19.1	44.7	47	8.3	16.7	75.0	12	8.3	16.7	75.0	12	0.0	0.0	0.0	0
19	40.0	13.8	46.2	130	42.7	11.2	46.1	89	34.1	19.5	46.3	41	0.0	20.0	80.0	10	0.0	20.0	80.0	10	0.0	0.0	0.0	0
20	37.1	15.2	47.6	105	40.3	11.1	48.6	72	30.3	24.2	45.5	33	0.0	20.0	80.0	10	0.0	20.0	80.0	10	0.0	0.0	0.0	0

Table A1: Probit models: propensity to complete and interview and propensity to refuse by postcard, using the full list sample including postcard refusals; propensity to refuse by post card, using the full set of list sample refusals.

		Full list sample		List sample		refusals	
		~~Complete~~		~~Returned~~ postcard		~~Returned~~ postcard	
Intercept		-0.629		-1.439		-0.295	
		0.148		*0.183*		*0.295*	
LSSTRAT	7	-0.678		0.141		0.201	
		0.116		*0.139*		*0.231*	
	6	-0.331	#	0.103		0.117	
		0.087		*0.114*		*0.188*	
	5	-0.125		0.101		0.220	
		0.088		*0.116*		*0.194*	
	4	0.004		0.209	+	0.027	
		0.091		*0.118*		*0.193*	
	3	0.012		0.298	*	-0.053	
		0.094		*0.121*		*0.195*	
	2	0.111		0.379	#	-0.015	
		0.100		*0.126*		*0.201*	
PSUTYP	1	0.332	+	-0.031		0.575	
		0.186		*0.222*		*0.428*	
	2	0.072		-0.138	*	0.159	+
		0.048		*0.058*		*0.095*	
REGION	4	0.033		0.088		-0.039	
		0.056		*0.068*		*0.116*	
	3	0.112	+	0.210	#	-0.014	
		0.060		*0.071*		*0.120*	
	2	0.035		0.074		-0.475	
		0.061		*0.073*		*0.117*	
P_LE17		0.011	#	0.001		0.003	
		0.004		*0.005*		*0.008*	
P_GE65		-0.003		0.006	*	0.010	+
		0.003		*0.003*		*0.005*	
PERCMED		0.000		-0.001	*	-0.001	
		0.000		*0.000*		*0.001*	
PHHWPAI		-0.001		0.021	*	0.013	
		0.009		*0.010*		*0.013*	
POWNOCC		-0.001		0.002		0.004	
		0.001		*0.002*		*0.003*	
MAGEHU		0.005	#	0.003	+	0.006	*
		0.002		*0.002*		*0.003*	
PMINOR		-0.002		-0.007	#	-0.004	
		0.002		*0.002*		*0.003*	
PNOENG		-0.002		-0.001		-0.006	
		0.007		*0.008*		*0.013*	
PHISP		-0.003		0.002		-0.003	
		0.003		*0.003*		*0.005*	
N		4,898		4,898		1,204	
Log(likelihood)		-2,941		-1,896		-799	