Key words: Survey field operations, nonresponse bias, paradata

Getting to the Top: Reaching Wealthy Respondents in the SCF

Arthur B. Kennickell
Chief, Microeconomic Surveys
Federal Reserve Board
Mail Stop 153
Washington, DC 20551
Arthur.Kennickell@frb.gov

August 18, 2009

Paper prepared for the 2009 Joint Statistical Meetings, Washington, DC

Abstract

This paper examines the effort devoted to securing interviews with a very wealthy part of the sample for the 2007 Survey of Consumer Finances (SCF). Only about a quarter of the group completed an interview. At the close of the field period, more than a third of this part of the sample was judged by the field staff to be still workable—that is, those cases were neither complete nor final refusals. The evolution of the field work was driven both by the behavior of respondents and the behavior of the field staff. The paper uses the formal data coded in the call records for each case to describe the work. But that information is inconclusive about the factors that drove the work. However, informal notes in the call records do provide a clear picture of the points of resistance among respondents. Although it was difficult to locate, contact, and convince respondents of the legitimacy and value of the survey, it appears that the ultimate constraint in a large proportion of cases was the length of the interview—potentially several hours for these respondents. Examination of the available auxiliary data provides little evidence of nonresponse bias.

Acknowledgements

The views expressed in this paper are those of the author and they do not necessarily represent the opinions of the Board of Governors of the Federal Reserve System or its staff. The author is grateful to staff at NORC, including Suzanne Bard, Catherine Haggerty and Micah Sjoblom in the central office and to Maureen Bonner, Tracie Buie, Ella Kemp, Sandra Pitzer, Nina Walker and many other field staff for their essential contributions in fielding the 2007 SCF. The author is also grateful to staff at the Federal Reserve Board and SOI for their contributions to the design and processing of the survey, particularly Brian Bucks, Gerhard Fries, Barry Johnson, Traci Mach, Kevin Moore and Thomas Petska. Finally, the deepest thanks must always go to the survey respondents, without whom there would be little to discuss here.

Operational decisions and respondent behavior are critical elements in any field survey. Questionnaires and sample designs may be endlessly polished, but once a field survey is set in motion, the overall statistical credibility of the effort ultimately depends on behavioral factors that are tied up in decisions by field staff to locate, contact and persuade respondents and in decisions by respondents to structure their lives to resist intrusions or to resist the entreaties of field staff when there is contact. In general, the two parties party face different incentives and they have different goals. Depending on the outcome of their sometimes interdependent decisions, a survey case may be completed or not. There is no necessity that such an outcome is neutral. Depending on the resulting distribution of incomplete cases relative to the distribution of completed cases, there may be bias in some estimates.

Experience has shown that some groups of survey respondents tend to have higher rates of nonresponse than others. One such group is wealthy households. Although success in obtaining interviews with such respondents faces many of the same obstacles as for other respondents, wealthy respondents appear generally to be far more difficult to contact and to persuade to participate.

This paper examines the efforts devoted to interviewing a set of wealthy respondents in the 2007 Survey of Consumer Finances (SCF) and it looks at some indicators of response bias. The first section of the paper provides background on the survey. The second section presents evidence on some measures of effort applied to the sample and the outcomes of those efforts, as well as some informal information derived from case-level call records. The third section probes potential response bias. A final section concludes and discusses the next steps.

I. Background on the 2007 SCF

The SCF is intended to provide information on household wealth, income and associated variables (see Bucks *et al.* [2009]). The questionnaire explores these topics in detail. Design of an appropriate sample to support the intended analysis of these data must address two special issues. First, wealth in the U.S. is highly concentrated, with about two-thirds held by the wealthiest 10 percent of households and about a third held by the wealthiest one percent (see Kennickell [2009]). Second, experience has indicated

that relatively wealthy households are less likely to participate in surveys (see Kennickell [2005]). The SCF addresses these challenges by using a dual-frame sample that includes a multi-stage area-probability sample in which each case is selected with equal probability, and a differentially sampled stratified list sample. The area sample provides robust coverage of characteristics that are broadly spread in the population. The list sample is designed to oversample wealthy households and it is structured in a way to allow identification of nonresponse related to wealth.

The list sample was constructed by Federal Reserve Board (FRB) staff using statistical records derived from individual income tax returns by the Statistics of Income (SOI) Division of the Internal Revenue Service, under an agreement that provides protections for the privacy of taxpayers and limits the use of the data to statistical purposes. A model of wealth defined in terms of income characteristics and other variables is developed using multiple years of the SOI data; the predicted value of the model is taken to be a "wealth index," which is used to divide the population of taxpayers into seven strata (see Kennickell [2001]).[1] This paper focuses on the sixth and seventh strata (the two wealthiest), which have predicted wealth above the 99th percentile of the distribution of the wealth index. In 2007, the median net worth of the survey participants in the sixth stratum was approximately $50 million and that of the cases in the seventh stratum was about $300 million. The area-probability sample was selected by NORC at the University of Chicago (see Tourangeau *et al.* [1993]).

Data collection was conducted by NORC between May 2007 and March 2008. Only a small number of interviews were completed in 2008; the field period was extended from its expected close in December, largely to accommodate unexpected difficulties that required a number of cases to be returned to the field. The list-sample cases were released for field work about one month after work on the area-probability cases had begun. The overall median interview length was about 90 minutes, but interviews for wealthy respondents could last for several hours.

---

[1] When the sample was selected, members of the *Forbes* list of the 400 wealthiest people in the U.S. were removed. The number discovered in the sample was far smaller than 400, presumably because not every such case may be sampled, the wealth reported in *Forbes* may actually be owned by a number of family members or a trust rather than the named individual, the model may be faulty, or the SOI or *Forbes* data may be in error.

Interviewers were required to record their actions on each case in a system of electronic call records (paradata). Each record contains the date and time of a given action taken on a case, the nature of the action (mailing, in-person attempt, telephone attempt), what person (if any) had been contacted, the outcome of the contact, notes on the event, and a few miscellaneous other data items. Because these records were used to monitor interviewers' performance, they give a reasonably complete and highly reliable inventory of efforts by interviewers for most of the field period.[2] Efforts by higher-level field staff (in refusal avoidance or conversion, targeted mailings, and other means), particularly in the later part of the field period, were much less likely to be documented in the call records; although these actions were clearly quite important in resolving the later interviews, they were also likely to be a very small proportion of the history of a given case. A larger problem, as will emerge later in this paper, is the extent of the content of the records.

Of the 4,422 completed interviews in the 2007 survey, approximately one third derived from the list sample and about a third of those cases are accounted for by the cases in the wealthiest two strata.[3] The response rate for the two top strata combined was 26.2 percent (table 1);[4] fewer than a quarter of the completed cases in this group were conducted in person.[5] Cases that resolved as non-interviews because of a refusal or inability to contact the respondent accounted for another 26.4 percent, postcard refusals were 11.3 percent of the total and 36.0 percent were not resolved as complete or final

---

[2] It is clear from reading the records, that multiple events (usually attempted contacts on the same day) are sometimes combined. In the analysis that follows, all contact attempts on a given day are treated as a single effort. Considerable efforts were devoted to locating many of the list-sample respondents (and a few area-probability respondents who moved after the initial contact); such information normally appears summarized as a single entry made by higher-level field staff.

[3] Neither the sampling rates nor the exact number of observations for the list sample strata can be revealed. But it can be revealed that the seventh stratum is roughly one-third the size of the sixth stratum.

Unlike the area-probability sample cases, the list-sample cases were given an opportunity to opt out of the survey before those cases were released for field work. About 12 percent of the cases in the sixth stratum chose to refuse participation in this way and about half as many in the seventh stratum did so; altogether, there were about 100 such cases.

[4] The response rate in the highest stratum was approximately 12 percent and that in the next-highest stratum was about twice as high. Owing to restrictions on what can be revealed about the individual strata, these two groups are pooled for most of the remainder of the work presented in this paper.

The analysis reported is unweighted. Weighting the results for the the area-probability sample would make little difference in the results. The blurring here of the two list-sample strata is intentional.

[5] The field staff was given the flexibility to conduct interviews in person or by telephone. Telephone interviews were favored in order to control costs, but according to the field staff, wealthier respondents had a strong preference for telephone interviews.

incomplete during the field period (that is, those cases were still viewed workable at the close of the field period).[6] [7]

Table 1: Final disposition of in-scope cases; list-sample strata 6 or 7 and area-probability sample cases, 2007 SCF; percent.

Final disposition	*List sample, stratum 6 or 7*	*Area-probability sample*
Complete, telephone	19.3	24.2
Complete, in-person	6.1	43.2
Complete, proxy	0.8	0.3
Postcard refusal	11.3	NA
Final refusal	19.9	18.1
Final break-off	0.2	0.1
Final refusal by gatekeeper	2.6	0.1
Final unlocatable	0.6	1.4
Unavailable during field period	0.2	0.6
Language barrier (non Spanish)	0.1	0.9
Too ill/handicapped	1.2	1.0
Stopped work	36.0	8.0
Other nonresponse	1.6	2.0
All	**100.0**	**100.0**

In contrast, the area-probably sample had a response rate more than twice as high and a much higher proportion of in-person interviews. The samples also differ strongly in the fraction of unresolved cases at the end of the field period.[8]

The agreement between the FRB and NORC specified several constraints on response. For the area-probability sample, the FRB set a minimum required number of completed cases and a range of allowable response rates, with constraints on the geographic distribution of the completed interviews. For the list sample, the agreement specified a minimum number of completed cases for each stratum. The sample size of each stratum was set based on historical performance in securing interviews.[9]

Whenever a set of cases is worked until there is no possibility of further completed interviews, the prospect of potentially serious selectivity bias induced by field

---

[6] A distinction is made throughout this paper between cases that were *resolved* as incomplete during the field work (most typically, cases that refused participation) and those that were incomplete because they did not have a final resolved status at the close of the field period. Where the unmodified term "incomplete" is used, both resolved and unresolved incomplete cases are intended; unless otherwise noted, the postcard refusals are excluded from the analysis.

[7] Unlike most other respondents, respondents in the part of the list sample considered here were not routinely offered a fee for participating in the survey. The presumption was that anything that could plausibly be offered would seem small and it might also raise suspicion about the legitimacy of the effort. However, some people objected to being given nothing for the sacrifice of their time. Interviewers had the flexibility to mention a fee when necessary. Many times the respondent designated the fee to go to a charity.

[8] Virtually all area-probability cases were actually coded by the field staff as resolved as either completed or incomplete. A set of cases that were given a final resolved noninterivew code only at the very end of the field period is treated here as unresolved.

[9] For the 2007 SCF, a replicate structure was adopted for the sample. To test the operational efficiency of the sample implementation, replicates containing, in total, only 85 percent of the total number of sample cases were released.

operations recedes. Even though a seemingly large fraction of the area-probability cases was not resolved by the end of the field period, it was clear from detailed inspection of the call records for these cases that even with great additional effort and a further extension of the field period, it would have been very unlikely that more than a very small number of additional cases could have been completed (see Kennickell [2008]).[10] The situation for the list-sample cases appears to have been quite different, as discussed variously later in the paper.

II. Effort and Outcome

Often discussion of the response rates of surveys proceeds as if all of the key decisions rested with the respondents. But generally, there is no reason to expect that effort would be applied by field staff toward securing interviews without at least some regard to the particulars of each case. *A priori* particulars may include the location (or uncertainty thereof) of the respondent and the characteristics of area, the identity of the respondent (in the case of the SCF list sample), the available field staff, and other factors. Once any work has begun, additional information becomes available that could reasonably be expected to be used to guide effort—the presence of a physical or human barrier to contact, the respondent's availability (in time and space), previous positive or negative feedback from the respondent about the survey, additional intelligence about alternative ways of reaching the respondent, and a wide variety of other factors.

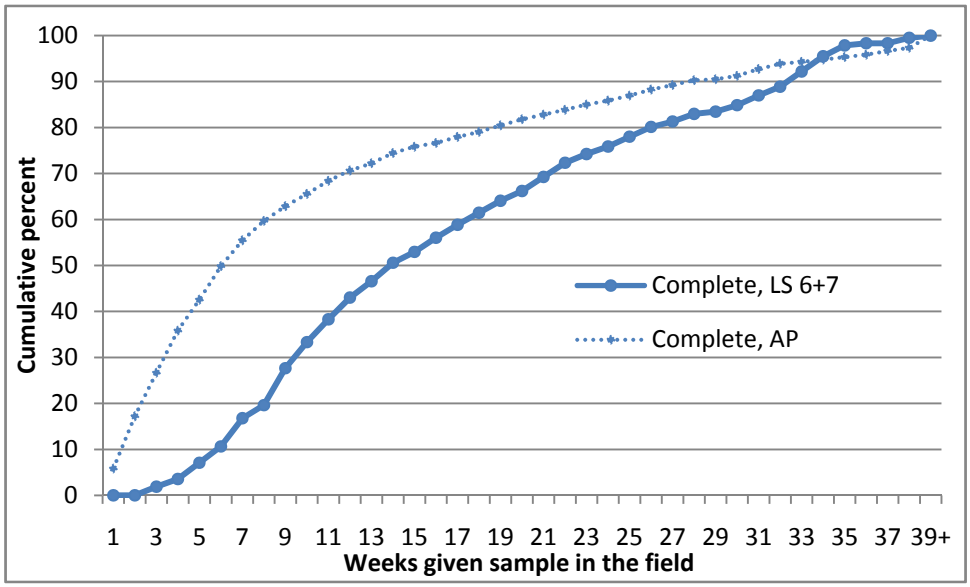
Operational efficiency argues for making as much use as possible of such information. However, this endogenous application of effort seriously complicates the systematic understanding of the progress of cases through the field period, and the interpretation of effort and response in the context of models of respondents' characteristics or behavior; this is particularly so in where there is no systematic indication of the motivation for the decisions made by field staff, as is the case in most surveys.[11] Nonetheless, it is still possible to use the call records to describe the patterns of work on the sample, even if the interpretation of those patterns is open to question.

---

[10] No further distinction is made in the paper between resolved and unresolved incomplete cases from the area-probability sample.

[11] The SCF employs a protocol for managing the contact attempts that is intended to give a more deterministic structure to the initial stages of effort, without overly complicating the survey operations (see

Figure 1: Cumulative distribution of completed interviews; by number of weeks sample released to the field; list sample stratum 6 or 7 and area-probability sample; 2007 SCF.

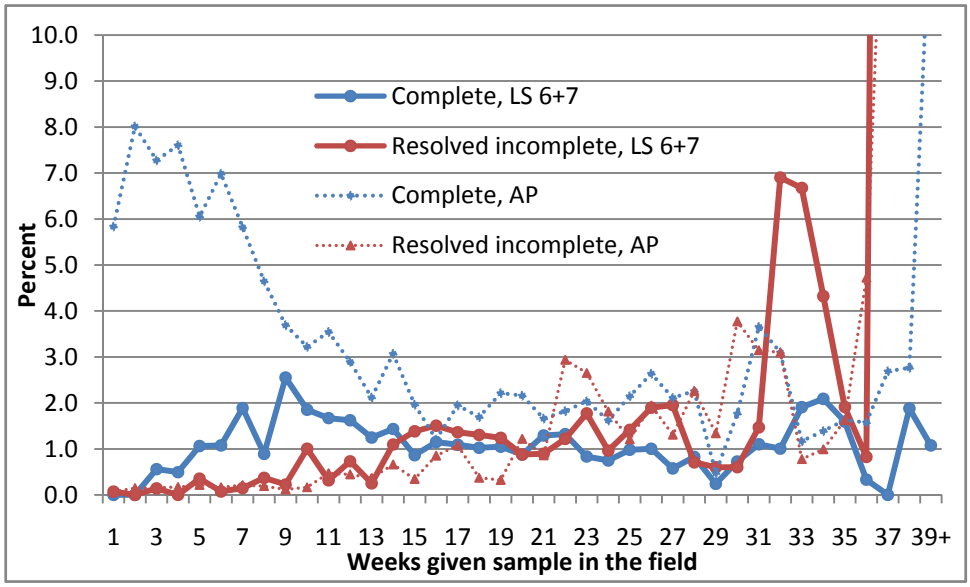


The speed with which work is completed is an important measure for operational cost reasons. The rate of progress was quite different among the area-probability sample cases and list-sample cases considered here. Owing to an expected lag in the receipt of the information needed to contact the list sample, those cases were released to the field about a month after the area-probability sample. But even if the timing of effort is viewed relative to the start of work on each sample, the area-probability sample cases were still much more likely to be completed early in the field period than the cases in the top two strata of the list sample (figure 1). By the 28th week of work on the area-probability cases, 90 percent of the completed interviews from that sample had been obtained. In contrast, this point was not reached for the list-sample cases considered here until the 32nd week of work on that sample. Although the formal records of locating activities are fragmentary, it is clear from discussions with the field managers that such actions account for a substantial part of the slower progress of the list-sample cases. As discussed below in the review of the call record notes, these respondents can be quite hard to find, both because ownership of multiple home is common and because regular travel for other purposes features prominently for many people in the group.

Kennickell [2005]). In practice, the protocol is not implemented cleanly enough to enable direct use of the phases in analysis, at least without incorporating other factors to refine the classification. Nonetheless, the protocol should serve as insurance that at least a credible amount of effort is devoted to every case released to the field.

Another way of looking at the progress of the samples is to transform the underlying case resolution information into hazard rates—that is, the number of cases resolving with a given status (complete or incomplete) as a fraction of all cases remaining unresolved just prior to that point. Hazard rates can make it more straightforward to see shifts in outcomes.

Figure 2: Hazard rate for final resolution of cases as complete or incomplete, by number of weeks sample released to the field; list sample stratum 6 or 7 and area-probability sample; 2007 SCF.



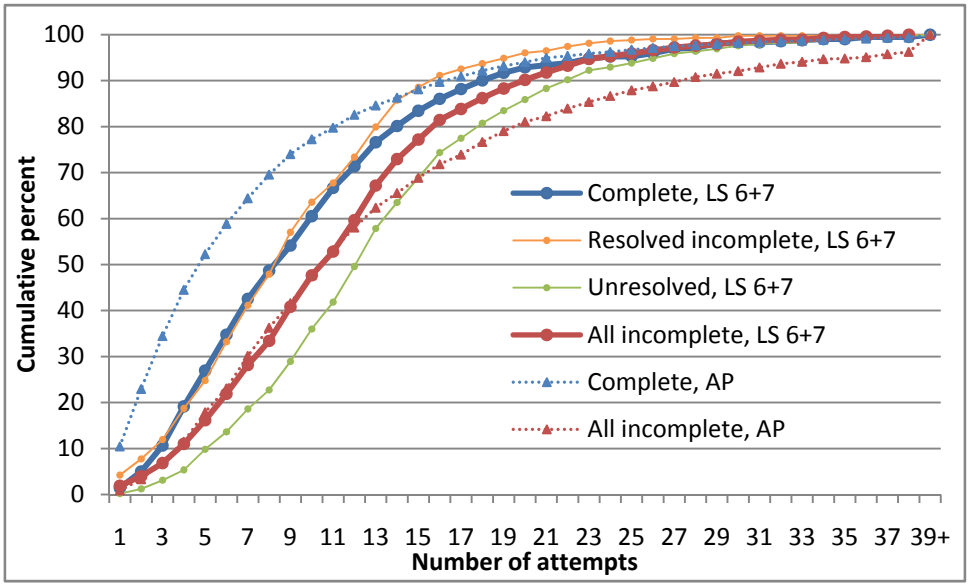
The weekly hazard rate for completing an area-probably sample case at a given point in the field period began at about 6 percent for the first week, rose to over 8 percent, and then declined to about 2 percent after the 16th week until late in the field period (figure 2). During each period, there is also the possibility that a case will resolve as incomplete (typically a refusal of some sort). For this sample, the hazard rate for resolved incomplete cases rose gradually throughout the first half of the field period; it continued to rise on average after that and became more variable late in the field period.

From the release of the list sample to the field, the hazard rate for completion in the top two strata of that sample rose gradually to over 2 percent and then declined to around 1 percent—below the rate for the area-probability sample—until about 32 weeks into the field period. Up to that point, the hazard rates for resolved incomplete interviews were very similar for the two sample groups. As with the area-probability cases, the rates for resolved incomplete interviews were more variable later in the field period.

Underlying the hazard rates is a set of ongoing operational judgments about which cases remained legitimate targets for additional work at each time (the denominator of the hazard rate). Because the field staff decides when a case is to be considered a final resolved incomplete case, their behavior can affect this rate. For example, as the field period progressed, field managers made periodic decisions to cull cases they believed were very unlikely to be completed, in order to allow clearer focus on cases that appeared to offer a hope of completion; some such points were related to decisions about reductions in the size of the field staff. The spikes in the later part of the field period in part reflect such decisions. Although the decisions to accept such cases may have become sporadic, inspection of the detailed call records suggests that cases accepted as resolved incompletes had very little likelihood of completion, regardless of how much additional effort might have been applied.

The *unresolved* cases at the end of the field period might be viewed as just another group of cases with little hope of completion that just happened to be culled at the end of the field period, but the available information suggests a more complicated situation, at least for the list-sample cases, as discussed later in this paper.

Figure 3: Cumulative distribution of number of attempts, for final status of cases as complete, resolved incomplete, or unresolved; list sample stratum 6 or 7 and area-probability sample, 2007 SCF.



There was considerable variation within and across both sample types in the number of attempts expended to secure an interview (figure 3).[12] As one might expect, the distribution of effort for completed list-sample cases in the two strata considered here is strongly shifted to the right of that for the area-probability sample cases, and the distributions of effort for incomplete cases are shifted to the right of those for the completed cases for both sample groups. But it more surprising that the distributions of effort for the two sample groups are nearly identical for the first 60 percent of all incomplete cases; above that, it is the area-probably sample distribution that is more shifted to the right.

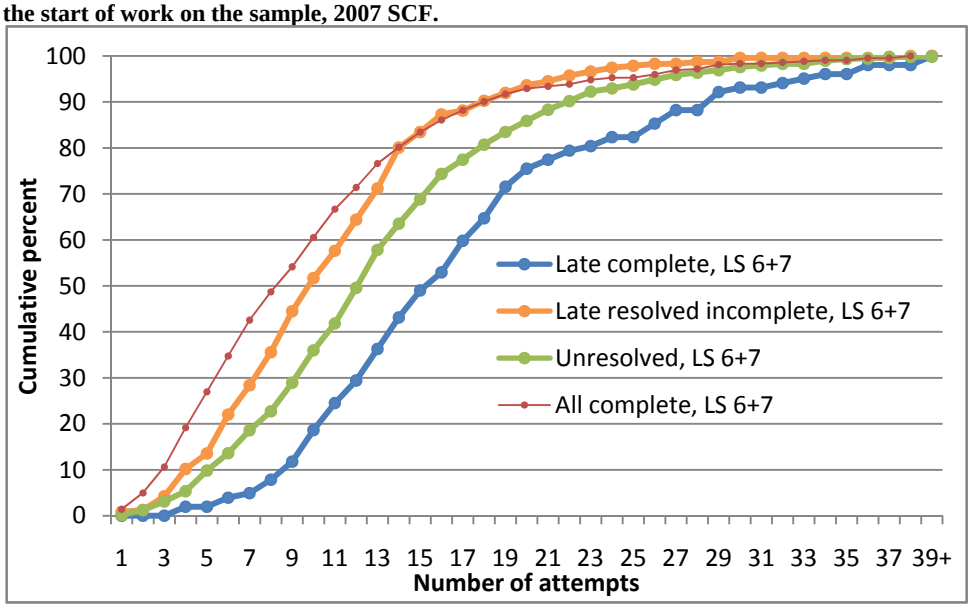
When the incomplete list-sample cases are decomposed into resolved and unresolved cases, the data show that the effort profile for resolved incomplete cases is much like that for complete cases, while that for the unresolved cases shows a higher degree of effort in the intermediate range and a lower degree in the higher range than the incomplete area-probability cases. On the surface, it is not possible to tell whether the differences in distributions reflect differences in the two samples in the propensity of respondents to cooperate, operational decisions, or a combination of the two.

Because unresolved cases were, by definition, not resolved by the end of the field period, it may be less appropriate to compare effort on those cases with ones resolved in the early part of the field period. If the unresolved cases were truly about as likely to be completed as the cases completed toward the end of the field period, then one might expect to see a distribution of effort for them that is more similar to that for those completed cases.

---

[12] Here an attempt is taken to be any action taken on a case other than one involving only recording a comment of information obtained from a search. Multiple actions taken on the same day are treated as one action. The actions include in-person visits, telephone calls (regardless of outcome) and mailings.

Figure 4: Cumulative distribution of number of attempts, for final status of cases as complete, resolved incomplete, or unresolved; list sample stratum 6 or 7 completed after 24 weeks from the start of work on the sample, 2007 SCF.

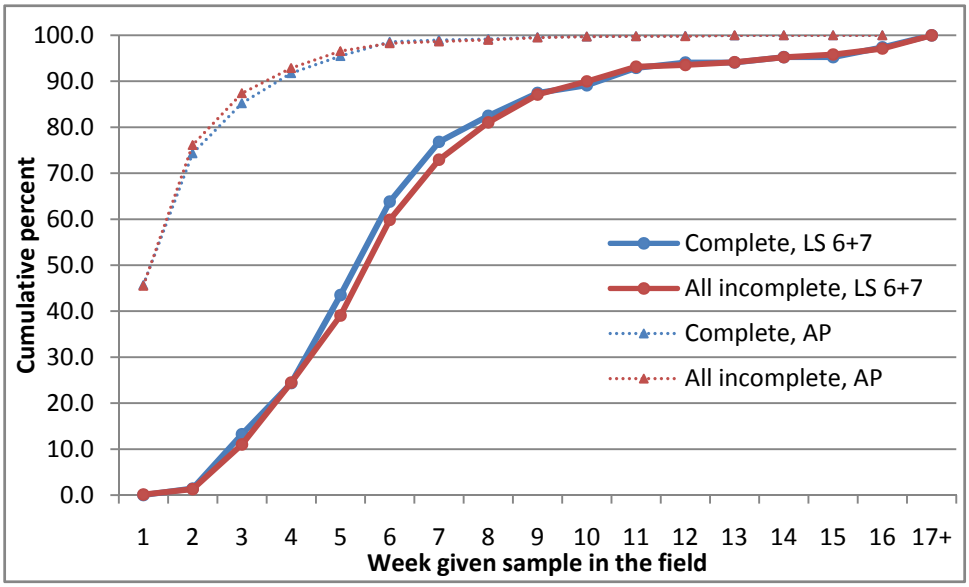


Slicing the data to focus on work done after 24 weeks from the start of work on the list sample (figure 4) shows that the distribution of effort for the later completed cases is, to the contrary, shifted to the right of that for the unresolved cases.[13] Because there is evidence that the field staff under-recorded some efforts at the end of the field period as they worked to complete a case, the level of effort for the completed cases may even be understated on average. The distribution for the unresolved cases does remain right-shifted relative to that for the cases that were resolved as incomplete during this period. What cannot be told from this figure is whether the unresolved cases received less effort than the late complete cases because respondents were not available to be worked on productively or because the field staff had selected other cases to work more intensively, either based on specific information or more or less randomly.

[13] During this period, a final resolution was reached for about a quarter of all the list-sample cases in stratum 6 or 7 that were completed and about 55 percent of the incomplete cases.

However, it is clear that there were operational decisions resulting in a much longer lag for the list-sample cases overall between the release of the sample and the first recorded attempt to secure an interview (figure 5). For about 40 percent of the high-stratum list-sample cases, no attempt had yet been made to secure an interview 6 weeks after that sample had been released to the field. In contrast, attempts had been made on nearly all of the area-probability sample cases at that point. Reportedly, before serious work was allowed to begin on list-sample cases, field managers made great efforts to locate respondents or to obtain telephone numbers with a reasonable hope of being usefully connected to them, and as noted earlier, such work may not appear in the call records. The later start of work on these list-sample cases may also have been influenced by a desire to test interviewers' abilities before assigning them the most complex cases. In addition, there may have been some pressure to maintain the overall level of production early in the field period.

Figure 5: Cumulative distribution of the number of weeks from the release of the sample to the first attempt, for final resolution of cases as complete or incomplete; list sample stratum 6 or 7 and area-probability sample; 2007 SCF.

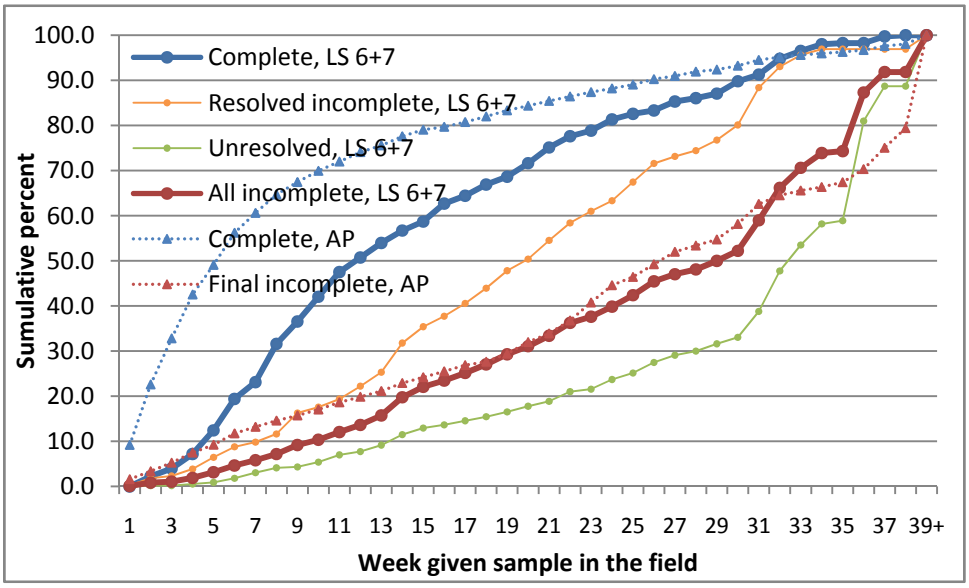


The time from the release of each sample to the first personal contact presents quite a different picture (figure 6).[14] Typically, more time elapsed in trying to contact

[14] A personal contact could be any sort of oral communication between the field staff and any person encountered in the attempt to reach the respondent—including literal or figurative gatekeepers, employees, co-workers, family members, friends, neighbors or others with some connection with the target household or person.

list-sample cases that were completed than was the case for area-probability cases that were completed. But cases that were ultimately incomplete took much longer, and there was relatively little difference between the patterns for such cases in the two samples. Separating the unresolved and resolved incomplete cases changes this pattern. Unresolved list-sample cases show a much greater time to the first personal contact and the resolved incomplete cases show offsettingly less time.

Figure 6: Cumulative distribution of weeks until first personal contact, for final resolution of cases as complete or incomplete, list sample stratum 6 or 7 and area-probability sample; 2007 SCF.



At 24 weeks after the start of work on the list sample, about three-quarters of the unresolved list-sample cases had not yet been contacted personally. Only somewhat smaller fractions of the cases that were completed or resolved as incomplete during this period had been contacted (figure 7). What these data alone cannot tell is whether these results stem from difficulty in reaching the respondent or from a decision to apply less effort. In terms of attempted contacts, the call records show that cases completed in this period received more attention leading up to the first personal contact than the unresolved cases, and the unresolved cases received more attention than the cases that were ultimately resolved as incomplete (figure 8). This pattern would make sense if the field staff had *a priori* expectations that the cases resolved as complete were more likely to resolve in that way. But the pattern could also reflect other decisions about the management of work or unclassified behavior of respondents.

Figure 7: Cumulative distribution of weeks until first personal contact, for cases resolved as complete or incomplete after 24 weeks of work on the sample and cases unresolved, list sample stratum 6 or 7; 2007 SCF.

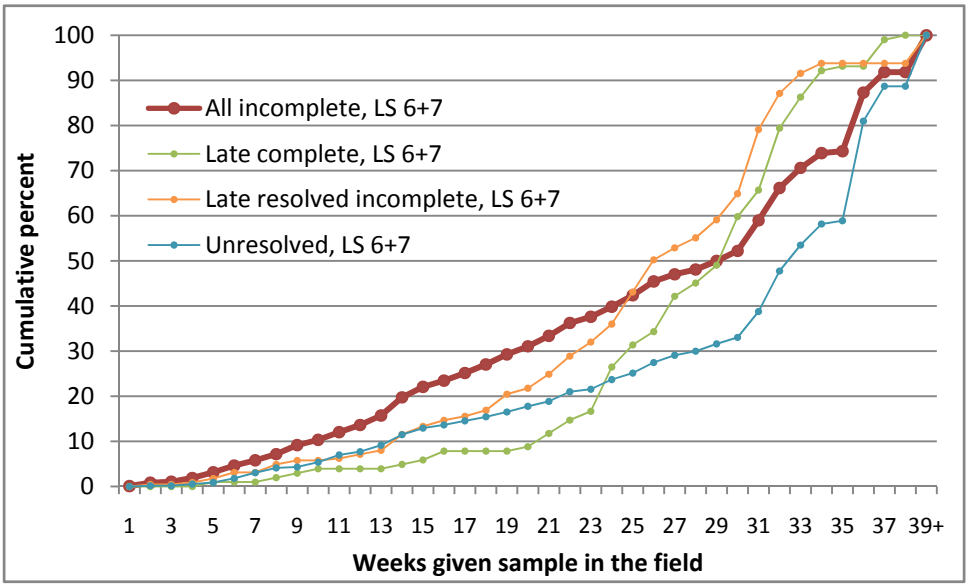
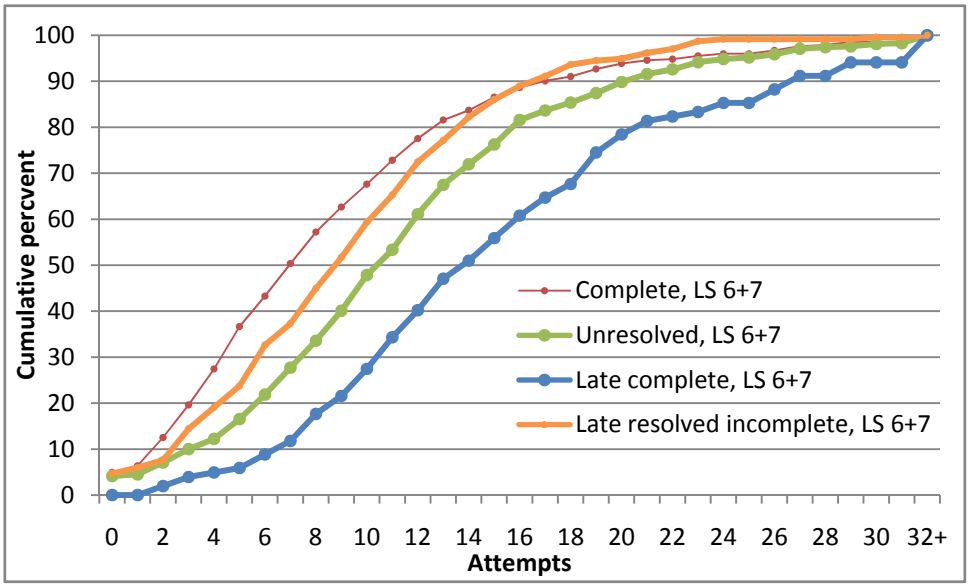


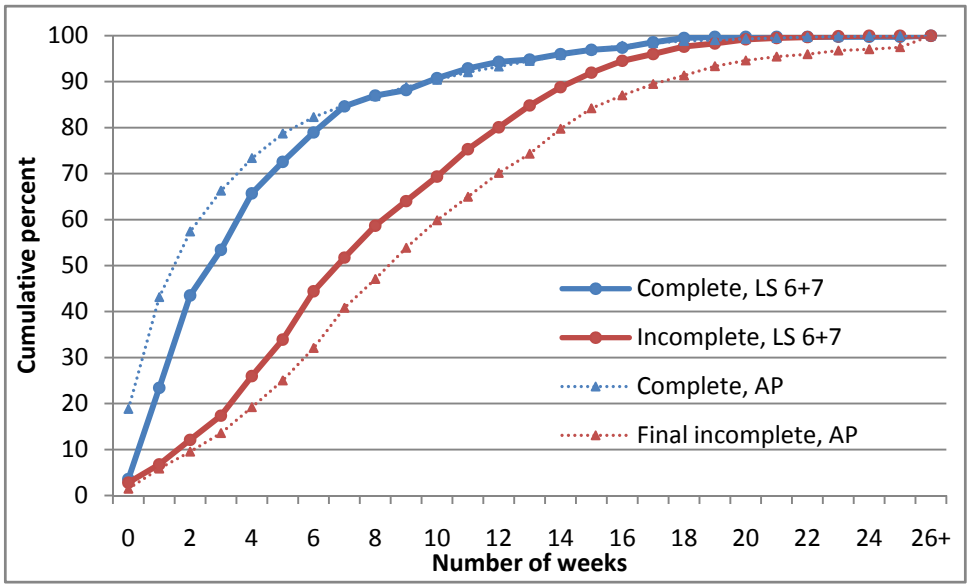
Figure 8: Cumulative distribution of the number of attempts before the first personal contact, cases resolved as complete or incomplete after 24 weeks of work and cases unresolved, list-sample cases in stratum 6 or 7, 2007 SCF.



Once contact was made, there was a very wide array of patterns of activity that would be difficult to summarize meaningfully. But one potentially revealing indicator is the maximum amount of time elapsed between attempts to secure an interview (figure 9). Incomplete cases (resolved and unresolved) tend to have much longer gaps than

completed cases; unresolved cases (not shown) have almost the same pattern as incomplete cases overall. Differences between the distributions for the list-sample cases and the area-probability cases are fairly small. Completed list-sample cases tend to have somewhat larger gaps than completed area-probability cases, but the pattern is reversed for incomplete cases. The gap potentially reflects a number of factors: a "cooling off period" for respondents who expressed resistance, unavailability of the respondent or a local interviewer, deliberate diversion of effort to cases believed to be more likely to be completed, or simply a degree of inadvertence. Whatever the underlying motivations may have been, it is clear that some cycling of effort over cases was an important element of the field work.

Figure 9: Cumulative distribution of the maximum number of weeks between attempts, for final resolution of cases as complete or incomplete, list sample stratum 6 or 7 and area-probability sample, 2007 SCF.



Information from call-record notes

To understand more deeply the movement of cases toward resolution, we need to know what the field staff knew or expected and why they acted as they did (or did not act) across the field period. There are many other ways of slicing the formal measurements in the call records, but as must be clear by now, the formal data are very limited in their power to reach this level of clarity. However, there is more detailed information in the notes that interviewers routinely recorded along with the more formal information coded in the call records. Because this information was intended to facilitate

communication between managers and interviewers and to facilitate transfers of cases among interviewers, there was continuing pressure to record pertinent information about the history of each case. But because these comments are open-ended remarks made at the discretion of the interviewer and without a uniform analytical perspective, it would be very difficult to code the information in a way that would support formal analysis. Nonetheless, the notes do constitute a rich, if fragmentary and selective, body of information for each case on the steps taken and the motivations for some of those steps.

A number of themes emerge clearly from this information. These records suggest that the most important problems overall for the cases in the top two strata of the list sample were related to (1) locating the respondent, (2) contacting the respondent, (3) getting the respondent's attention, (4) addressing the respondent's concerns about the confidentiality the answers or the legitimacy of the survey, and (5) obtaining sufficient time from the respondent to complete the interview. Although there was some overlap of these categories at various points for some cases, the general picture is one of progression through these issues. While there were corresponding issues for all list-sample and area-probability cases, these problems were generally expressed with greatest intensity among the list-sample cases in the top two strata considered here.

Compared with other cases, the respondents for these list-sample cases were much more likely to be away from the address associated with the original sample selection—at a second home, traveling for work domestically or overseas, or living at a new location—and many were contacted at their place of business (or one of its branches). Entirely unlike the case for the area-probability sample, which is based on a sample of household addresses, for many list-sample cases the initial address was the office of an attorney or an accountant or other tax preparer.[15] Because very wealthy people sometimes take pains to disguise their location, the locating efforts occasionally mistakenly identified someone with the same name as the correct respondent, and the mistake was not be uncovered until substantially later. The field managers were heavily involved in locating efforts, and reportedly they did not normally release the cases to interviewers until plausible contact information was available.

---

[15] Because there was no information about on the name of the responsible party within such offices, such addresses were often a dead end in locating the respondent.

Even when an address was obtained, contact with the respondent or anyone directly associated with the respondent was often quite difficult. Gated homes and homes or offices in guarded or otherwise restricted areas were common. Other gatekeepers—most commonly employees or other agents of the respondent—were also very often present, even when access to a physical address was not seriously restricted. Commonly, a large amount of effort was devoted to identifying a confidential assistant to the respondent, and persuading that person to act as a conduit for gaining direct or indirect access to the respondent in order to get a decision about participation; because such staff are often employed for the express purpose of keeping people away, a great deal of skill and persistence was often required to gain even their neutral acquiescence.

Having established a channel of some sort to the respondent, the information about the survey had to compete with a large variety of other distractions, many of which were intense and related to the respondents' conduct of business or other work. Initial mailings introducing the survey were most commonly not recalled by respondents or their gatekeepers. Even subsequent express mailings often had a low level of saliency. Faxes and e-mails were often not received or were blocked or unnoticed. Once any attention was gained, a skilled team of field staff acted in coordination to fan the respondent's interest. Sometimes before they were fully informed, respondents would tell the interviewer that they were not sufficiently "typical" to be of use in the survey. The field staff reported that often, when a respondent understood what was being asked and how the data would be used, the person was sophisticated enough to recognize the importance of the project.

But even motivated and sympathetic respondents often still had serious reservations. The questions on the SCF address topics that very many people consider to be among their most private information. For some in the group of list-sample respondents considered, there are legal agreements that preclude their sharing of financial information with anyone, presumably short of the IRS. Wealthy people are obvious targets for people engaged in identity theft, scams, as well as legitimate fund raising. Considerable effort was devoted to convincing respondents that the study was legitimate, that no one would ask them for money as a result and that their data would be kept confidential.

Even having informed, reassured, and at least minimally gained the respondents' interest, the length of the interview was usually a large obstacle. People in the sample group considered here typically have great time pressure during the work day, and they guard their time outside of work carefully. Over 20 percent of the interviews for this group took more than 2 hours, and some required more than 3 hours. Interviewers were instructed to be honest with respondents about the time commitment required, but they were able to offer to do the interview in as many segments as the respondent might need or want, or to do the interview with a proxy authorized by the respondent. However, less than 1 percent of the list-sample participants designated a proxy and only about a quarter of the list-sample participants had an interview spanning multiple sessions; these proportions did not vary greatly over the field period.

There are also some distinct patterns in the history of cases by outcome and by stage of the field period. Although the five issues noted earlier were present for cases completed in the first two months after the release of the list sample to the field, the intensity of expression was generally low. Respondents for these cases appear to have been relatively easy to contact and persuade, and the call records indicate than of them even gave expressions of active interest. In the relative ease in which these cases were completed, they resemble the area-probability cases that were completed similarly early.

Cases that resolved as incomplete during this time were mostly ones that were adamant in refusing to participate, though in some instances serious health problems precluded participation. One might think that the postcard-refusal stage of the sample implementation would have removed the most resistant respondents from the sample released to the field staff, but it is apparent that many of people had either never received or never read the advance letter. Some others expressed a very strong sense of the value of their time relative to their perceived value of the survey and incredulity about being asked to share their private financial information. Many refused even to listen to the interviewer or read any of the materials and in some cases interviewers clearly suffered a great deal of verbal abuse. Because the cases resolved as incomplete in this early phase generally rejected participation before any meaningful dialog was established, little else aside from the frame data is available to understand the situations and motivations of these respondents. A very small number of the cases accepted as incomplete looked, at

least to the extent identifiable from the notes, little different from other cases that continued to be pursued until later in the field period.

For cases completed after about the first two months, the five basic issues grew more amplified, but the longer a case remained in play, the further along that list of obstacles a case moved. The call record notes for this period include more frequent mention of difficulties in locating respondents—often it turned out that they were away at a summer home—of gatekeepers who were initially less helpful to the interviewer and more protective of the respondent, and of privacy and confidentiality concerns that were expressed more forcefully. In addition, descriptions of the extremely busy lives of the respondents were very common. Repeated in-person visits became an important means of establishing rapport with gatekeepers, who would sometimes be very helpful to the interviewer in getting the respondent's attention. Express mail became an increasingly important means of reaching respondents through gatekeepers or of getting the respondents' attention.

Almost throughout the later part of the field period, interviewers managed to speak directly with respondents who had previously been so protected by gatekeepers that they were completely unaware of the survey. When finally contacted, the reaction of most of these respondents appeared to have been very similar to that of respondents who had been reached earlier in the field period. Some readily agreed to complete the interview, while others expressed strong resistance. Most required additional persuasion.

Often privacy concerns or serious questions about the legitimacy of the survey were raised. In some cases there was not sufficient opportunity to parry respondents' misperceptions. But where some dialog was possible, it appears the field staff was successful in addressing those issues. Increasingly, the overall impression is that the inability or unwillingness of respondents to find time for the interview became the most important issue. Some respondents (or their gatekeepers) insisted that they would not be able to find time for the interview. But it appears to have been more commonly the case that there was a general willingness on the part of respondents to cooperate, subject to their being able to find the time; it is, of course, impossible to evaluate the sincerity of such respondents. Persistence on the part of the interviewers in returning to respondents or their gatekeepers appears to have been the most important factor in finding an opening

in which to do the interview. Strenuous refusal-aversion interventions were successful in reversing, or at least postponing, acceptance of many cases as resolved incompletes.[16]

Among cases left unresolved at the end of the field period, there are clear signs in the call-record notes of resistance to participation, but these notes do not seem strikingly different from those for most of the cases that were completed after the earliest part of the field period. For the great majority of these unresolved cases, it appears that time was also the most important factor.

Given the unpredictability of available time for many in this sample group, a clear focus on finding openings for individual cases might seem most obvious. But the records often show complicated patterns of gaps in work across cases remaining eligible, and most often these gaps were completely unexplained. There some surprisingly long gaps for some cases the respondent apparently expressed a willingness to cooperate when time became available. A general pattern of gaps runs throughout the work on this sample; it may have been that there were undocumented efforts, that the respondents were being given a "cooling off" period, or that such cases simply evaded attention because of an inefficient cycling of attention and effort over too many cases.

III. Potential Response Bias

In addition to having implications for the efficiency of information collection, variations in effort to secure interviews may lead to response bias—that is, meaningful differences in the distributions of characteristics observed for participants and the corresponding unobserved characteristics of nonrespondents. At least for the SCF list sample, extensive information is available from the statistical records derived from tax-return data for the years preceding the survey that were used in selecting this sample. An obvious variable to examine first is the wealth index used in the original sample stratification; that index, which is intended as a proxy for wealth, combines a variety of different variables related to age, income and other characteristics reported on a tax

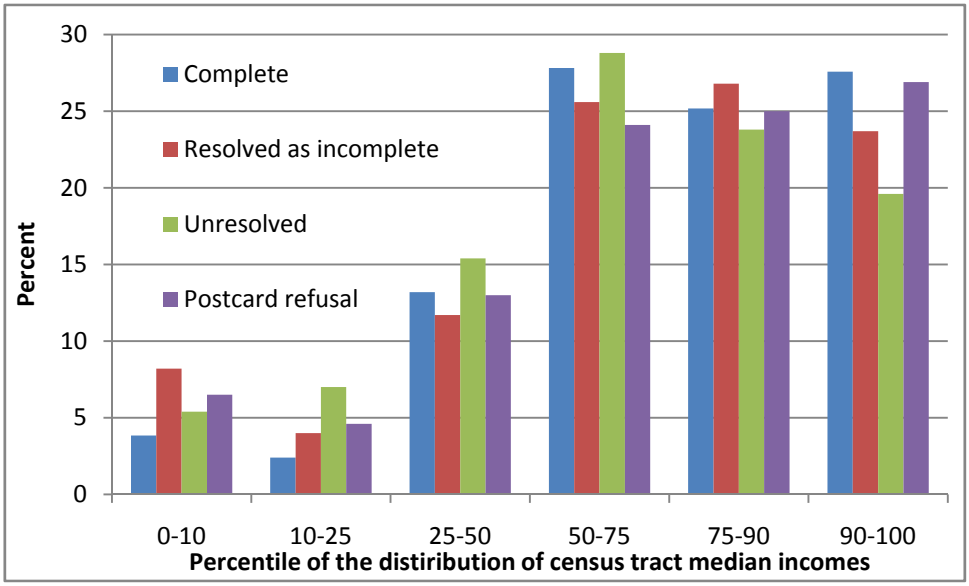
---

[16] Except in instances where respondents objected strenuously to the survey, interviewers were instructed to tell respondents who wanted the interviewer to stop attempting to contact them that they should call a particular toll-free number to get themselves removed from the list of people eligible to be contacted. A respondent could indeed be removed in this way, but the person in charge of this effort had the opportunity to make a last appeal to the respondent. Remarkably, the effort was successful in convincing a large fraction of the respondents who called to participate.

return. Comparison of the distribution of the index across the set of complete cases and each of the sets of incomplete cases—postcard refusal, resolved incomplete and unresolved cases—shows only negligible differences among these groups. Although there may be (necessarily unobserved) difference in the types of assets and liabilities cases in the different groups hold, this result does increase confidence that overall wealth, in at least the univariate sense, is well measured for these groups.

As a part of the design of the list sample, all cases are mapped into the census tract corresponding to the location from which their tax return was filed. Although returns are required to be filed from a home address, it has long been clear from SCF field work that returns are sometimes filed from the office of an accountant, attorney or other address different from the taxpayer's home address. Because so much locating effort is required on the SCF for reasons other than this one, it is not possible to assess the degree of noise in the available measure of location. The distribution of cases over neighborhoods classified by median incomes is very similar for each of the types of incomplete cases (resolved incomplete, unresolved incomplete and postcard refusals) and the completed cases (figure 10). About three-quarters of the completed cases and each of the incomplete cases filed taxes from areas with median incomes above the level of the median over all census tracts. The unresolved cases were somewhat more concentrated in tracts with median incomes just above the median over all tracts and somewhat less concentrated in areas with higher incomes.

Figure 10: Distribution of list-sample cases in stratum 6 or 7, by percentile of the distribution of 2000 Census tract median incomes; by final cases disposition; 2007 SCF.

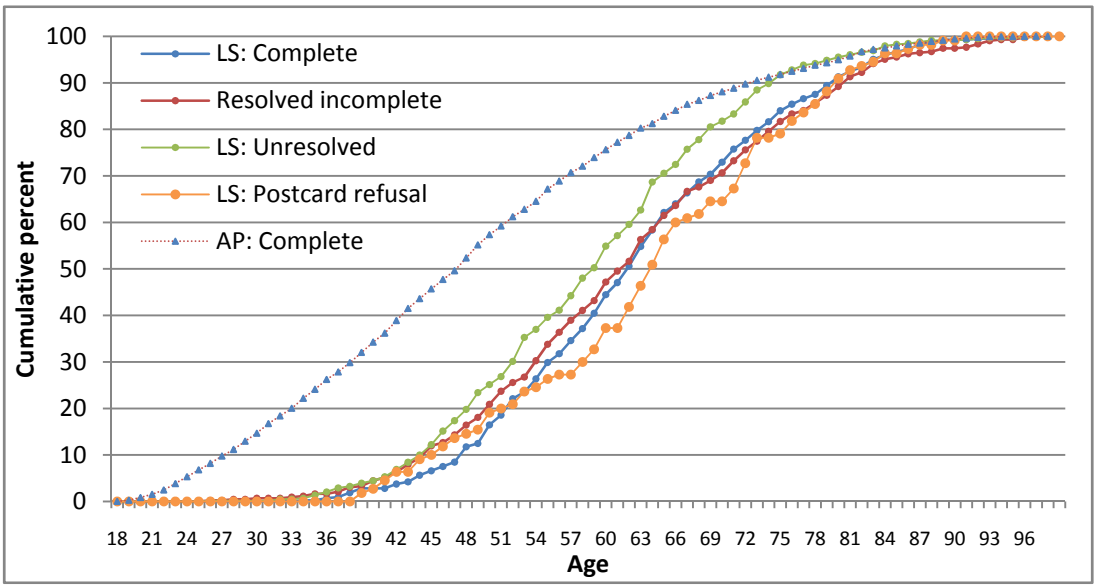


The sample file for the list sample also contains the year of birth of the taxpayers associated with the tax return. Among the list-sample cases considered here, the 2007 age distribution of the primary filer for the cases that returned the refusal postcard is overall relatively shifted toward older ages, while the age distribution for the unresolved cases is overall shifted toward younger ages (figure 11).[17] The distributions for the complete cases and the resolved incomplete cases lie intertwined largely between the other two. But the differences between all these response groups are very small relative to the overall differences with the age distribution for the area-probability sample cases that completed the interview.[18]

[17] In an income tax return that is filed jointly by a married couple, the primary filer is the one whose name is listed first. In a return filed for an individual, that individual is the primary filer.

[18] For the area-probability sample, age is available only from the completed interviews.

Figure 11: Cumulative distribution of age of respondent; for final status of list-sample cases in stratum 6 or 7 as complete, resolved incomplete, unresolved or postcard refusal and of area-probability cases as complete; 2007 SCF.



IV. Summary and Future Research

This paper investigates the progress of field work on a set of very wealthy cases in the list sample selected for the 2007 SCF. The data used in this study are largely the call records maintained by the field staff for each case. Although it is very clear that a substantial amount of work was required to gain the agreement of such respondents to complete an interview, there was also a great deal of variation in the intensity and consistency of treatment across cases. Understanding the underlying dynamics is important both for minimizing bias in the resulting data and for optimizing the operational efficiency of the survey. But these ends are hindered by the inadequacy of the formal coding system used to record effort. The systematic information available covers only the timing and outcome of efforts on each case, along with some formal information about who, if anyone, was contacted. There is richer information in a set of notes recorded in the call records by the field staff about some details of their work, but even that source generally fails to explain the degree to which variations in effort reflected behavioral responses by respondents, strategic decisions by the field staff, or even inadvertence.

Overall, the call notes clearly outline the general problems that interviewers faced with these wealthy cases: locating the respondents, contacting them, getting their attention, allaying their concerns about the confidentiality of their answers and the legitimacy of the survey, and convincing respondents to take the time to do the interview. Although these issues are common for all types of cases in the survey, they are expressed with great intensity for wealthy respondents. When there was sufficient opportunity for the field staff to engage with the respondents, it appears that the respondents' overriding concern was the length of the interview and the scarcity of their free time. And generally, in the last months of the field period, the interview length was by far the most visible point of difficulty. Success with such seriously time-constrained respondents appears to have been a result of persistence in attempting to get an interview and luck in hitting upon a time when the respondent was relatively free.

When all cases are ultimately resolved as either complete or incomplete, the room for operational decisions to affect the distribution of outcomes is limited, though obviously the timing and sequencing of effort may still have strong effects on outcomes.[19] In a case where a large number of cases remained at the close of the field period without a final resolved status, as for the upper strata of the SCF list sample, there is much more room for the behavior of the field staff to shape the distribution of outcomes. Field staff is under great pressure to complete interviews; the obvious resulting rational incentive is for energy to be focused on pursuing, in rank order, cases that have the highest subjective probability of being completed. Such an outcome has something of the outward appearance of a convenience sample selected from the overall sample. Moreover, this incentive complicates interpretation of the role of respondents' information and their behavior in the progress toward the final resolution of cases.

At the close of the 2007 SCF field period, there was a set of completed cases and several sets of incomplete ones—ones that had opted to refuse participation in the survey by returning a postcard before the field work began, ones that were accepted as fully

[19] A more basic point is that the decision to code a given observation as anything other than complete is often subject to some degree of judgment. The relevant standards for accepting a case as a resolved incomplete may differ within a given survey, as well as across surveys and institutions collecting data. Such variation may account for some of the differences observed across surveys in the factors that appear to determine survey participation. It appears rare for a study of nonresponse to consider the role of survey administration as a factor, beyond allowing for the possibility of interviewer-specific effects.

resolved noninterviews during the field period, and ones that were never resolved as complete or incomplete by the end of the field period. Because each of these groups is a substantial fraction of the original sample, concern naturally arises that the completed cases might differ from the others in ways that would bias the results of the survey. As in most surveys, the information available to investigate such potential bias is limited. Nonetheless, data available for the sample show no evidence that within the set of participants from the sample groups considered here, there was bias in the level and distribution of wealth, the distribution of those cases over neighborhoods by income levels, and their distribution by age. The comment data in the call records suggest that the most likely dimensions of bias would tend to be variables systematically related to time constraints, and, to a somewhat lesser degree, concerns about confidentiality. Such factors might be related to particular choices of assets and debts and to other aspects of financial behavior. One way of getting an indication of the likelihood of bias in those dimensions might be to model such choices and test for the significance of long work hours for the respondent and of the interviewer's evaluation of any suspicion expressed by the respondent before the interview.

The large fraction of unresolved cases among the part of the particularly wealthy respondents considered in this paper could be taken to suggest that the efforts of the field staff were operationally inefficient. That is, it could be that the available field staff could not concentrate effectively on so many cases, particularly when a great deal of initial work was usually required to locate and contact respondents. Could the same number of interviews have been completed in the same period of time with a smaller initial sample, should the sample have been culled early in the field period, or was it necessary to have so many cases and to keep so many of them in play until the end of the field period in order to have a chance of reaching busy respondents at a random moment when time happened to be available? The answer depends on whether the observed irregular patterns of attempted contacts were largely deliberate reactions to respondents' behavior, or whether they were largely the result of either random or loosely controlled decisions. If availability of time were random from the perspective of the field staff, then a regular rotation of contact attempts over a set of not-unwilling cases would make sense and the number of cases would be related to the time available and the probability of the

availability of time; but if there were case-specific factors suggesting more or less effort, more varied patterns of effort would be more sensible. Unfortunately, the available data provide only very limited information that could help in discriminating between these alternatives. However, the fact that it was possible to reduce the sample size by fifteen percent from the previous survey without obvious harm suggests that there may be at least some additional degree of inefficiency.

To be more useful for analytical purposes and for purposes of evaluating operational efficiency, call records need to contain clearer indicators of the nature of the activities recorded, some formal information about the objectives of each effort, and an indication of the likelihood of opportunities available for the next step. Any changes in information collected must take account of the fact that these records are viewed by field staff as administrative data; additional information required must be made sufficiently useful to field staff for them to be motivated to provide information useful for research and evaluation. It is expected that the 2010 SCF will expand the set of formal information collected in call records.

Bibliography

Bucks, Brian K., Arthur B. Kennickell, Traci L. Mach and Kevin B. Moore [2009] "Changes in U.S. Family Finances from 2004 to 2007: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin*, v. 95, pp. A1–A55.

Kennickell, Arthur B. [2009] "Ponds and Streams: Wealth and Income in the U.S., 1989 to 2007," FEDS Working Paper 2009-13, Board of Governors of the Federal Reserve System, http://www.federalreserve.gov/pubs/feds/2009/200913/200913pap.pdf.

Kennickell, Arthur B. [2005] "Darkness Made Visible: Field Management and Nonresponse in the 2004 SCF," working paper, Board of Governors of the Federal Reserve System, http://www.federalreserve.gov/pubs/oss/oss2/papers/asa2005.5.pdf.

Kennickell, Arthur B. [2001] "Modeling Wealth with Multiple Observations of Income: Redesign of the Sample for the 2001 Survey of Consumer Finance," working paper, Board of Governors of the Federal Reserve System, http://www.federalreserve.gov/pubs/oss/oss2/papers/scf2001.list.sample.redesign.9.pdf.

Tourangeau, Roger, Robert A. Johnson, Jiahe Qian, Hee-Choon Shin, and Martin R. Frankel [1993] "Selection of NORC's 1990 National Sample," working paper, National Opinion Research Center at the University of Chicago, Chicago, IL.