# Action at a Distance: Interviewer Effort and Nonresponse in the SCF[1]

**Arthur B. Kennickell, Mail Stop 153 Federal Reserve Board, Washington, DC 20551**

**Key words:** Nonresponse, interviewer effects, survey operations

Nonresponse is a problem in most surveys. Much effort has focused on defining the theoretical problems with nonresponse, characterizing its particular effects in individual surveys, and in devising ameliorative strategies for analysis (see Groves and Couper [1998] for a wide variety of citations). Not surprisingly, nonresponse has been seen largely as reflecting the characteristics and behaviors of respondents. What has been much less examined is the role of the survey administrative process in arriving at the final set of participants out of a sample of respondents. Groves and Couper [1998] and Kennickell [1999] examined the nature of the interviewer-respondent negotiation as an important contributing factor, but these address just one facet of a complicated set of issues connected with the behavior and management of interviewers.

This paper follows directly from Kennickell [2003], which presented an analysis of the application of interviewer effort and nonresponse in the 2001 Survey of Consumer Finances (SCF).[2] The detailed analysis of call records presented in the 2003 paper indicated that there was a systematic and nonuniform application of effort by interviewers–that in some cases appeared to have carried through into the ultimate patterns of nonresponse. Moreover, the distribution of effort across the sample cases appeared both quite skewed and logistically inefficient. An important consequence of variation in effort is that nonresponse must be treated as a joint function of the interviewers' decisions to apply effort and respondents' decisions to participate. Thus, unless interviewers' behavior can be made exogenous, their perceptions and expectations, the incentives and constraints they face, and the degree to which various aspects of their work are observable potentially all interact with respondent characteristics to complicate greatly the understanding of and proper adjustment for nonresponse. This paper describes one sort of step toward such exogeneity.

The function of the interviewer is much too complex in most instances to be governed fully by predetermined rules, but it may still be possible to take substantial control of this behavioral component of nonresponse. The approach discussed here is a three-phase plan of sample management, where the first two phases take as an explicit goal the equalization of effort across sample cases. This plan is being applied in the data collection for the 2004 SCF. If fully realized, information obtained in the first phases might allow the construction of an explicit strategy of case targeting in a final phase. In this aspect, the approach can be viewed as a type of "responsive sampling" as described by Heeringa and Groves [2004]. Given the level of uncertainty surrounding the effectiveness of the new case management model for SCF and its potential cost implications, it is unlikely that the third phase design actually adopted in the 2004 SCF will be much in advance of the traditional methods of case management at the end of a field period. Rather, the information collected and operational lessons learned will be used to design a more controllable process for the next SCF data collection effort for the in 2007.

The first section of the paper presents a simple model of field effort to guide interpretation of the later discussion. The second section gives a brief overview of the SCF and lays out the case control strategy for the first phase of the field effort. The third section presents information about the progress of the case management system through the first twelve weeks of the field period. The final section summarizes the important points of the earlier parts of the paper and discusses possibilities for the next stages of case control, for ultimate nonresponse adjustments, and for changes in the design and execution of the next wave of the survey in 2007.

## I. A Stylized Model of Survey Operations

This section presents a simple model of the working of a survey sample by field interviewers and develops the general proposal elaborated in the remainder of the paper. Despite the tedious presented notation below, the point is a simple one. A rational interviewer will tend to apply most effort to cases most likely to agree to participate, thus amplifying nonresponse among groups less likely to participate. If the likelihood of participation is related to the variables of interest, then the application of effort will also amplify bias in the distribution of those variables.

Suppose there are $I$ interviewers, and each interviewer $i$ is given an allocation of $A^{i,0}$ cases to work at the beginning of a field period. The number of cases available to be worked may change over time as cases acquire a final resolution or as cases are transferred between interviewers. For expository convenience, suppose no case transfers are allowed and no observations are out of scope. Final resolutions include either completing a case or failing to complete a case because of failures in locating, contacting, or persuading a respondent. At the end of period $t$ of the field effort, $A^{i,t}$ cases remain in the interviewer $i$'s inventory to be worked, $C^{i,t}$ cases have been completed, and $R^{i,t}$ cases have a permanent incomplete status ("refused", "unlocatable," etc.).

Suppose each sample element *j*=1 to J has an equal probability of selection. Each case *j* also has a subjective *a priori* probability of completion $\pi_j$ at the first attempt. With each subsequent attempt to interview the case, additional information is gained. But the nature of the information seen by the interviewers and by their managers differs. The interviewer observes the information set $\boldsymbol{\Omega}_j^i$, which is assumed to be finer than the information set of managers, $\boldsymbol{\Omega}_j^m$. For simplicity, suppose that the information observed about each case is a scalar measure $D_j$, a characteristic associated with the difficulty of completing case *j*. As a result of the informational differences, the subjective probability of case completion from an interviewer's point of view at time t, $\pi_j^{it}|D_j^{it}$, may differ from that of their managers, $\pi_j^{mt}|D_j^{mt}$. That is, interviewers can update their prior probabilities with their local information, but managers can only update their priors using averages within classes observable at their level. Thus, even if managers had an incentive to guide interviewers to offset systematic variations in effort that were a function of expected difficulty, they would have insufficient information to do so with full efficiency.

The vector of variables **X** intended to be collected for each case in the full sample $A^0 \equiv \cup^i A^{i,0}$, the set of all the case assignments to all interviewers, has a sample distribution $\Phi^A(\mathbf{X})$. For expository convenience, A is assumed to contain a sufficiently large number of cases that this distribution is negligibly different from the full population distribution $\Phi(\mathbf{X})$. Let the joint population distribution of X and D be given by $\Psi(\mathbf{X},D)$, where D and **X** have some covariation so that $\Phi(\mathbf{X}|D) \neq \Phi(\mathbf{X})$. Thus, even if effort were applied uniformly to all cases, there would be bias in the distribution of measured **X** because of the implicit selection on D.

Interviewers are generally rewarded for completing many cases in a cost-efficient way. As an extreme but here sufficient example, let interviewer *i*'s wage at period *t* be $w^{i,t}$, which is paid only upon completion of a case. Let the cost of case *j* to interviewer *i* in terms of psychic and material costs be $D_j^i$. Thus, from a risk-neutral interviewer's economic point of view at the beginning of period *t*, the goal is to apply effort to cases $j \in A^{i,t-1}$ to maximize $w^{i,t}E[\dot{C}^{i,t}|D^{i,t-1}] \equiv w^{i,t}\{E[C^{i,t}|D^{i,t-1}] - C^{i,t-1}\}$ subject to the constraint that $E[D_j^{it}|D_j^{i,t-1}] \leq w^{i,t}$ for all cases *j* worked. Assuming that effort requires some time to apply, the obvious result is that each interviewer will rank all cases in $A^{i,t-1}$ in order of $E[D_j^{it}|D_j^{i,t-1}]$ and apply effort to all cases that satisfy the constraint. With a fixed budget B in this very simplified model, the data collection will halt when $\sum_i C^{i,t} = B$ or $E[D_j^{it}|D_j^{i,t-1}] > w^{i,t}$. Thus, interviewers behaving in this way will tend to exacerbate the bias in the distribution of observed X by implicitly conditioning on D.

Many factors are omitted from this model. Interviewers for most social science surveys are not paid by the interview. Interviewers may vary in their abilities, perceptions, aspirations and commitment. The constraints they face may be more complex and there may even be types of gaming behavior across the pool of interviewers that influence the collective outcome of field work.[3] However, the basic point is one that would obviously hold in a wide variety of circumstances.

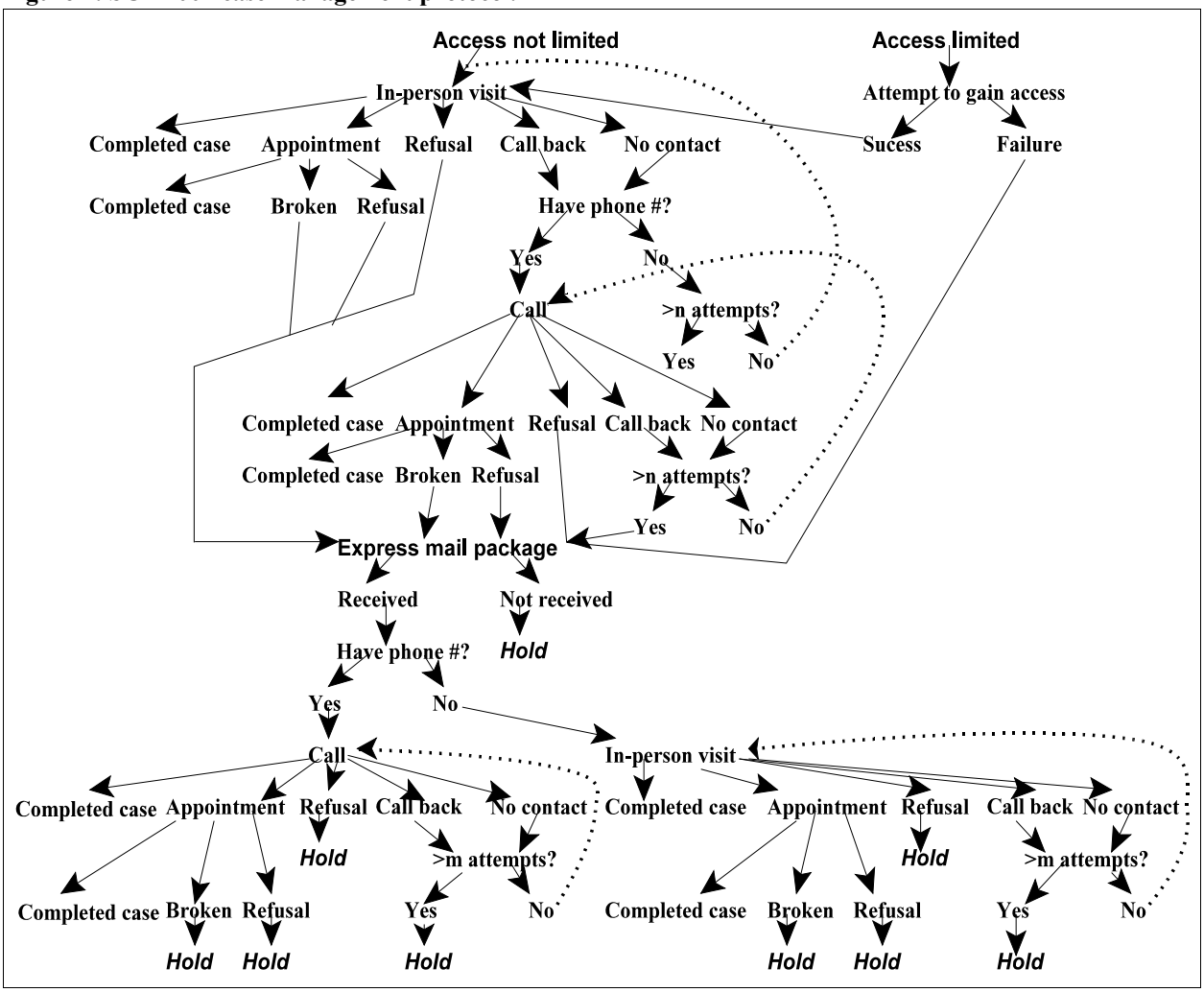
The goal of the ultimate analysts of the survey information is to have data that are unbiased and efficient for their intended statistical purposes. In this instance, this requirement implies a statistically efficient means of offsetting the confounding of the observed distribution of **X** by virtue of implicit stratification on D. For estimation of simple population statistics, weighting adjustments or other model-based adjustments may bring some of the desired result if: (1) there is a vector of characteristics **K** that are related in a knowable way to **X** across the joint distribution of its elements, (2) $\mathbf{K}_j$ is observed for every completed case, and (3) a reliable estimate of the overall population distribution of **K** is available from another source. Of course, for a given number of completed cases, adjustment is statistically inferior to completing an unbiased subset of $A^0$, which might be achieved through alternative allocations of field efforts.

## II. 2004 SCF Case Management Protocol

The SCF is a triennial survey conducted by the Federal Reserve Board in cooperation with the Statistics of Income Division (SOI) of the Internal Revenue Service. Data for the study are collected by NORC, a national organization for research and computing at the University of Chicago. The survey collects information on families' assets and liabilities, their use of financial services, their employment history and associated pension entitlements, and a variety of demographic and attitudinal data.[4]

The survey sample has a dual frame design. One part of the sample is a multi-stage national area-probability sample in which each address is selected with equal probability (see O'Muircheartaigh *et al.* [2003]); this sample provides good coverage of characteristics that are broadly distributed in the population. This sample includes 79 NORC "national frame areas" (NFAs) chosen at the first stage of selection. The other part of the design is a list sample, which is selected by SCF project staff at the Federal Reserve from statistical records derived from tax returns by SOI (see Kennickell [2001]). The list sample is stratified using a "wealth index" (a statistical proxy for actual wealth) and cases with higher values of the index are disproportionately sampled. The list sample cases are restricted to those with addresses within the

**Figure 1: SCF 2004 case management protocol.**




counties comprised by the full set of NFAs. It is anticipated for the 2004 survey that about a third of the expected 4,500 completed interviews will derive from the list sample.

Unit nonresponse is a serious issue for the SCF. The response rate in the area-probability sample is expected to be about 70 percent; for the list sample the rate is expected to range from about 10 percent among the cases with the very highest values of the wealth index to a rate more comparable to that for the area-probability sample among cases at the other end of the distribution. [5] Not surprisingly, much of the methodological research on the SCF has focused on nonresponse (see Kennickell [2000b] and references cited therein).

In earlier waves of the SCF, field work has followed the practices largely shared by the survey research organizations with the highest standards. Interviewers have been trained to follow a protocol that has been standardized in many important ways–techniques to use in presenting material to respondents, ethical considerations, sample maintenance, etc. However, the control of effort on individual cases has been left in large part in the hands of interviewers. Of course, there has been supervision by field managers and, to a much smaller degree, by central office staff, but arguably not enough. Usually global targets have been set centrally, and those targets have been managed by pressuring the workers in the field. Overall response rates have been the most common target, but limited balancing of rates across major sample areas toward the end of the field period has had some importance as well. Work reported in Kennickell [2003] about the 2001 SCF indicates that one result of this relatively laissez faire approach was systematic variation in the application of effort. That variation may have introduced biases while contaminating the understanding of the true response propensities of classes of cases in the sample. The statistical argument for intervention in case

management, thus, is strong. Two other factors also support change. First, ethical considerations argue for an approach that does more to equalize the opportunity of each sample element to be fully informed, whether the respondent ultimately agrees to participate or refuses to do so. Second, escalating costs for the SCF made a reexamination of the methods of production essential.

A new phased model of case management was developed for the 2004 SCF. This protocol, posed to the field as a "contacting strategy," is designed to have break points where response propensities of groups can be evaluated without contamination by variations in the level of effort across cases. Such evaluation offers both the possibility of developing a protocol to target effort in the terminal phases of the survey as well as in guiding post-survey nonresponse adjustments. By design, the approach imposes the requirement that every case be exposed to a roughly comparable level of information, consistent with the ethical goals. The approach also allows more structured management of interviewers. In addition to making costs more uniform, a stronger structure should free field managers to concentrate more clearly on problems that would otherwise be masked and to have more time to coach interviewers.

The contacting strategy as implemented has three phases. In the first phase, there is a prespecified path of actions to be taken for every sample case, depending on the actions of the respondent (see figure 1). In Phase I, a number of attempts are made to reach and interview respondents, while always trying to move effort \toward the use of the telephone (except where the respondent signals that mode to be unacceptable). With the exception of list sample cases in the high wealth-index strata, all cases are offered $20 as a token of respect for their participation.[6] Every case where effort does not yield a complete interview by the end of Phase I is sent, by express mail, a specially designed set of materials with an individually-addressed letter addressing the key point in their nonresponse to that point. Then, in Phase II every case that has received the mailing is subjected to follow-up comparable to the stage before the express mailing.

Cases not completed in the second phase are put into a pool for reevaluation in Phase III. Ideally, this final phase would use information gained in the first two phases to guide the work in the final phase. Unfortunately, there is insufficient flexibility to accomplish this goal for the 2004 survey. In practice, Phase III for the 2004 SCF is expected to rely on standard field procedures to achieve overall case targets with at most some relatively small rebalancing of effort as a result of the evaluation of the earlier phases.

## III. Analysis of Initial Field Work

The phased approach to sample management in the 2004 SCF is intended to force more equal effort across cases, at least through a substantial initial period, than seems to have been the case in earlier years of the survey. As noted earlier, work presented in Kennickell [2003] on the 2001 SCF indicates that some types of cases in the sample for that survey received systematically lower levels of effort than others. The new approach was highlighted in training both of the field managers and the interviewers, and a large amount of effort was taken to clarify the rules. As the field period progressed, the field managers were reminded of their responsibility for enforcing the protocol.

Over the first twelve weeks of the field period, which is the time available for analysis when this paper was written, progress toward the completion of Phase I was much slower than expected. As of the twelfth week, only 78.7 percent of the area-probability cases and 58.3 percent of the list cases had either reached the end of Phase I or had been interviewed before that point; only 53.9 percent of the area-probability cases and 22.3 percent of the list sample cases had either completed Phase II or been interviewed before that point (table 1).[7]

**Table 1: Percent of cases completed Phase I or Phase II or interviewed before each of those points, by subsample; 12th week of the field period.**

| | All | AP | LS |
|---|---|---|---|
| Phase I compl. or case interv'ed. before Phase I | 43.7 | 78.7 | 58.3 |
| Interv'ed before Phase I | 23.8 | 38.3 | 11.7 |
| Phase II compl. or case interv'ed before Phase II | 36.7 | 53.9 | 22.3 |
| Interv'ed before Phase II | 28.9 | 44.4 | 15.8 |

The inter-sample difference in the completion of Phase 1 may be due, in part, to two factors. First, the list sample is geographically more dispersed. There are some scale economies in working the more clustered area-probability segments that managers might reasonably want to exploit initially, but as unresolved area-probability cases become more sparse within areas, the logistical reasons for such a difference are less compelling. Second, because the interviews of list sample cases tend to be much more complicated than those of the area-probability cases, managers may have used the less complicated cases to give less experienced interviewers a chance to develop their skills where errors would tend to be less costly in terms of overall data quality. In addition to the variability of this rate across the two subsamples, the rate varied substantially

**Table 2: Percent of observations completed or reached the end of phase I, by week of field period and by tract income as a percent of area median income; area-probability sample.**

| *Week* | *Tract median income as % of area median* | | | | | |
|---|---|---|---|---|---|---|
| | 0-50 | 50-80 | 80-120 | 120-175 | >=175 | All |
| 1 | 14.1 | 4.1 | 3.8 | 2.6 | 2.4 | 3.9 |
| 2 | 20.9 | 10.9 | 9.1 | 6.8 | 4.8 | 9.2 |
| 3 | 34.4 | 19.3 | 17.7 | 16.3 | 12.1 | 18.1 |
| 4 | 40.5 | 28.3 | 24.9 | 22.8 | 19.4 | 25.4 |
| 5 | 49.1 | 37.5 | 35.9 | 33.9 | 29.1 | 36.0 |
| 6 | 54.6 | 45.4 | 44.0 | 44.9 | 41.8 | 44.8 |
| 7 | 57.1 | 53.4 | 51.7 | 53.7 | 43.0 | 52.3 |
| 8 | 66.3 | 59.2 | 56.8 | 60.0 | 52.7 | 58.2 |
| 9 | 68.1 | 62.5 | 62.0 | 64.7 | 55.8 | 62.7 |
| 10 | 70.6 | 67.7 | 69.0 | 68.7 | 63.0 | 68.5 |
| 11 | 78.5 | 74.9 | 74.6 | 77.2 | 70.3 | 75.3 |
| 12 | 81.6 | 77.2 | 77.3 | 79.3 | 76.4 | 77.9 |

**Table 2: Percent of observations with any action, by week of field period and by tract income as a percent of area median income; area-probability sample.**

| *Week* | *Tract median income as % of area median* | | | | | |
|---|---|---|---|---|---|---|
| | 0-50 | 50-80 | 80-120 | 120-175 | >=175 | All |
| 1 | 46.6 | 33.1 | 28.9 | 24.7 | 24.2 | 29.2 |
| 2 | 84.7 | 62.9 | 54.5 | 51.5 | 52.7 | 56.5 |
| 3 | 91.4 | 75.2 | 70.6 | 71.2 | 69.7 | 72.4 |
| 4 | 92.0 | 78.7 | 80.3 | 81.0 | 82.4 | 80.7 |
| 5 | 94.5 | 84.5 | 84.2 | 86.4 | 83.0 | 85.1 |
| 6 | 95.1 | 88.5 | 87.0 | 89.2 | 84.8 | 88.0 |
| 7 | 95.1 | 91.0 | 90.0 | 90.8 | 85.5 | 90.4 |
| 8 | 98.8 | 94.6 | 92.5 | 92.9 | 91.5 | 93.2 |
| 9 | 99.4 | 95.6 | 94.2 | 94.9 | 94.5 | 94.8 |
| 10 | 99.4 | 95.7 | 95.7 | 95.3 | 96.4 | 95.8 |
| 11 | 99.4 | 97.0 | 96.4 | 98.1 | 97.0 | 97.0 |
| 12 | 99.4 | 97.6 | 97.1 | 98.2 | 97.0 | 97.5 |

across the principal sample areas. One important explanation of this variation is the differences in staffing levels across NFAs.

Some of the differences may also reflect application of effort in inverse proportion to expected difficulty, as suggested by the model in Section II. For the area-probability sample, the data also show a pattern of differentially earlier completion of Phase 1 in census tracts with low median incomes relative to the median income of the larger area (tables 2).[8] For the list sample, the rate tends to be higher earlier in the sample strata where cases are likely to be relatively less wealthy (not shown). Clearly, these categories are not the only reasonable indicators of difficulty, but past

**Table 3: Percent of observations unresolved at a given week of field period that had any effort applied in that week, by week of field period and by tract income as a percent of area median income, area-probability sample.**

| *Week* | *Tract median income as % of area median* | | | | | |
|---|---|---|---|---|---|---|
| | 0-50 | 50-80 | 80-120 | 120-175 | >=175 | All |
| 1 | 46.6 | 33.1 | 28.9 | 24.7 | 24.2 | 29.2 |
| 2 | 60.6 | 43.5 | 39.2 | 38.7 | 38.9 | 40.6 |
| 3 | 44.7 | 43.5 | 42.4 | 44.3 | 38.1 | 43.0 |
| 4 | 42.2 | 38.3 | 40.7 | 42.5 | 34.8 | 40.5 |
| 5 | 42.2 | 42.7 | 38.7 | 43.1 | 39.1 | 40.6 |
| 6 | 32.3 | 38.1 | 34.6 | 40.3 | 38.8 | 36.8 |
| 7 | 25.0 | 42.0 | 36.9 | 40.4 | 28.6 | 37.9 |
| 8 | 54.8 | 39.3 | 34.6 | 39.1 | 43.1 | 37.5 |
| 9 | 24.0 | 31.4 | 33.9 | 36.2 | 35.1 | 33.8 |
| 10 | 25.0 | 28.3 | 32.2 | 28.2 | 26.2 | 30.0 |
| 11 | 40.0 | 41.4 | 35.6 | 37.6 | 37.8 | 37.4 |
| 12 | 24.2 | 28.4 | 27.5 | 26.5 | 24.6 | 27.2 |

experience suggests that they are important. Moreover, they have the advantage of being available for early analysis.

As an indicator of effort, completion of Phase 1 suffers from the fact that some cases are indeed completed relatively easily. The fact that some cases are completed with less effort than others could impose the appearance of a lag on what was otherwise a uniform application of effort. But the fact that cases where respondents were inaccessible or refused initially were eligible to skip directly to the end of Phase 1 should mitigate this distortion. Whatever may be the behavioral dynamics among interviewers and their managers, by the twelfth week of the field period, the variation across these income groups and strata were much less sharp. This result suggests that, the phased protocol was having the desired effect of steering behavior.

Underlying the movement toward Phase 1 completion, there was great variation in the rate at which initial effort was applied to cases in the various income groups and strata. By the twelfth week, 97.5 percent of the area-probability cases (table 2) but only 84.7 percent of the list sample cases (not shown) had been the subject of any sort of action beyond the initial mailing of study materials to the sample address. It is clear from the figures in the table that effort was applied early on disproportionately to the lower-income areas and that initial efforts for other cases followed with a lag; a similar patterns holds for the list sample by strata. Thus, the rate of Phase 1 completion for the poorer groups is not simply an artifact of higher rates of completed interviews among this group; these cases

were also more actively sought out, perhaps because they were thought to be more likely to participate.[9]

Another revealing way of looking at the field effort is in terms of hazard rates for application of effort in a given week. For a given group in a given week, this measure is given as the percent of cases to which any action was applied relative to the set of cases which remained eligible to receive effort (that is, those that were neither completed nor permanently refused). The data show quite variable rates across weeks of the field period for both subsamples. Particularly for the area-probability sample (table 3), the data show a general decline of effort, which reflects, at least in part, the steady reduction in the number of interviewers retained by the project over this period.[10] Throughout, the hazard rate for list sample cases lies notably below that for the area-probability cases (not shown). The first few weeks confirm the pattern seen in tables 2; subsequently effort within each subsample appears to have changed in such a way as to offset much of the initial differences. Statistical analysis of the data using hazard models (not shown) estimated for each subsample separately confirms this impression. The models control for relative median income or sample stratum, administrative area of the field work, MSA status of areas, and week of the field period. For the area-probability sample, the model estimated over the first six weeks of the field period shows systematically more effort directed toward lower-income areas; when re-estimated for the succeeding six weeks alone, neighborhood income has no significant effect. For the list sample, a model for the first six weeks shows relatively less effort in higher strata; this effect in the next six weeks is not entirely erased, largely because effort still flagged for the very wealthiest stratum. This remaining difference for the list sample appears to be more reflective of a desire to send only the very best interviewers to interview the very wealthiest cases than with any other sort of impediment.

If the rules for determining the path to the end of Phase 1 had been followed strictly, very few cases would show more than five attempts before being coded as having reached the end of Phase I, and many should show fewer. In fact, of the area-probability cases that actually reached the end of Phase I without yielding a complete interview, about 38 percent had more than five attempts; about eight percent had ten or more attempts. For the list sample, about 26 percent had more than five attempts and five percent had ten or more attempts. Although it is possible that some of the actions counted as "attempts" were very minor, the result suggests that there may have been some reluctance to move beyond traditional persuasion efforts in the field.

## IV. Discussion and Future Work

This paper presents a formal model of interviewer behavior, which suggests that decisions made by individual interviewers when left relatively unconstrained in fulfilling their measurable responsibilities may lead to systematic distortions in the ultimate set of respondents to a field survey. The driving force in the argument is that interviewers facing production demands will tend to pursue with greatest vigor the set of cases they believe, based on the information they have available, that they can complete most easily. If interviewers' expectations are correct, such behavior would serve to amplify patterns of nonresponse because "difficult" cases would tend to get a lower application of effort.

Statistical results reported in Kennickell [2003] based on the 2001 SCF support this argument and served as the motivation for the work in this paper. To counterbalance some of the differential application of effort and to provide a more measurable basis for managing the sample cases, a phased protocol of case management was developed. In the first phase, a limited number of attempts are made to contact the respondent and gain cooperation; if those attempts fail, the respondent is sent by express mail a package of materials tailored to the perceived obstacle–whether the respondent is difficult to contact, "too busy," concerned about privacy, uninterested, or "anti-government." That mailing was carefully constructed to present clear and compelling reasons for the respondent to participate. Phase II involves a limited number of attempts to follow up on the express mailing. If those attempts are not successful, the case is moved into a "hold" category where it is evaluated for follow up in Phase III.

Ideally, the sample cases would progress expeditiously to completion or to the end of Phase II, so that consistent empirically-founded decisions might be made about follow-up in Phase III. The reality shown in the tables in the preceding section of this paper is that there is a long lag for many cases in this progress and the set of lagging cases shows signs of being nonrandom. Two logistical factors in the 2004 SCF make a lag inevitable. First, the number of interviewers is relatively small. Second, there are economies in working through a given case, rather than attempting to work on every cases in an assignment simultaneously. If the mix of cases that reached the end of the phases is related to choices influenced by perceptions of respondents' characteristics–as seems likely–then the partial information available from the set of cases that completed the phases at a given point may well be misleading as a guide to managing Phase III. Thus, for the 2004 SCF control of effort over the remainder of the field period is likely to be limited to ensuring that all cases at least complete Phase II and that the NFAs with

the very lowest completion rates receive disproportionately higher attention.

Although it had been hoped that the phased design would be of broader utility for the 2004 survey, there are still three distinct benefits. First, an important ethical point has been addressed. It can be argued that people who have declined to participate have been more often fully informed of what it is they are declining. Second, the process of monitoring cases by stages does appear to have changed the behavior of both interviewers and their managers. Monitoring ensures that effort proceeds more evenly than would otherwise appear to have been the case in former waves of the SCF. Third, the fact that interviewers' decision to apply effort in the first two phases has been made exogenous to the resolution of each case as of the *end* of Phases I and II (though the sequencing of effort *within* each of those phases is still uncontrolled) means that unbiased models of nonresponse can be estimated at those points to guide the post-survey nonresponse adjustments. Being largely free of interviewer effects, these models would reflect most fully the nonresponse propensities of respondents. Thus, they offer a chance of identifying and addressing aspects of nonresponse that may be nonignorable.

Although systematic analysis of the Phase I and Phase II data cannot inform the final phase of the 2004 SCF, that information can be used to guide two types of decisions in the 2007 survey, which is expected to follow a comparable phased protocol. First, the information will be useful in projecting field staff needs and expected interviewer productivity. Second, the information will enable a more formal cost-benefit analysis to decide which cases to pursue in Phase III of that survey. Models can reveal where the set of participants has become "unbalanced," a least in some key ways, so the monetary cost of pursuing the "under-represented" respondents can be balanced against the costs in terms of estimation inefficiency or even bias. This point goes to the heart of a little discussed sampling issue that requires additional formal work before proceeding.

Much argument has been made over correct sample design, but relatively little discussion has been devoted to the connections between survey logistics and the ultimate statistics. Nonresponse is an additional stage of sample selection that is typically treated as dependent only on respondents, and without doubt this component is important. However, this paper argues that logistical control is also important and it may be possible to use logistics to overcome some of the statistical limitations induced by reluctant respondents. Sometimes statistical managers make post-selection decisions about subsampling nonrespondent cases for follow-up within the classical framework (see Hansen and Hurwitz [1946]), but until the recent work of Groves and colleagues, few statistical papers have been bold enough to address the possibility of targeting effort in any other way. One important exception is Seymour Sudman [1966], which argues for quotas within probability samples.

Subsampling and more specifically targeted action (whether at the level of individual cases or group quotas) make different assumptions about what is known and what are the necessary adjustments to the ultimate analysis weights. If quotas were based on a model that controlled for all relevant dimensions in a survey, they could be used to "rebalance" a set of participants by controlling effort applied to nonrespondents. So much is rarely known, though similar homogeneity assumptions are invoked in the application of post-stratification (Little [1993]). On the other hand, pure random subsampling of nonrespondents assumes that nothing is known about nonrespondents. This level of ignorance is also unusual; at the least, one could perform stratified subsampling.

More technical apparatus is needed to distinguish carefully between these two approaches. The reality of most surveys is that someone other than statisticians is already making decisions about the application of effort in what is a tangled practical blend of subsampling and quotas, and these decisions are most often ignored both at the level of statistical adjustment and end-use analysis. The hope of this paper is that the structured data from the 2004 SCF will provide sufficient ground to develop a deeper understanding of these issues and the potential for a more structured approach to field work.

## 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. And Mick P. Couper [1998] *Nonresponse in Household Interview Surveys*, John Wiley, New York.

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.

Heeringa, Steven G. and Robert M. Groves [2004] "Responsive Design for Household Surveys," mimeo Institute for Social Research, University of Michigan.

Kennickell, Arthur B. [2000a] "Wealth Measurement in the Survey of Consumer Finances: Methodology and Directions for Future Research," working paper, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

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

Kennickell, Arthur B. [1999] "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. [2001] "Modeling Wealth with Multiple Observations of Income: Dedesign of the Sample for the 2001 Survey of Consumer Finances," working paper, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B. [2003] "Reordering the Darkness: Application of Effort and Nonresponse in the Survey of Consumer Finances," *Proceedings of the American Statistical Association Survey Methods Research Section*.

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

O'Muircheartaigh, Colm, Stephanie Eckman, and Charlene Weiss [2002] "Traditional and Enhanced Field Listing for Probability Sampling," *Proceedings of the American Statistical Association Social Statistics Section*, pp. 2563-7.

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

**Endnotes**

1. The author wishes to thank Leslie Athey, Steven Pedlow, Fritz Scheuren, other NORC central office staff, and interviewers and other field staff at NORC for their help in testing and implementing the case control model described here. The author particularly wishes to thank the interviewers and field staff who have worked on various years of the SCF for the education they have given him in the way that field surveys operate. The opinions expressed in this paper are the responsibility of the author alone and do not necessarily reflect the views of the Board of Governors of the Federal Reserve System. An expanded version of this paper is available at http://www.federalreserve.gov/pubs/oss/oss2/method.html.

2. The 2003 paper, in turn, follows from Kennickell [2000a], which used data from the 1998 SCF to support an argument for a relationship between survey nonresponse and interviewers' behavior consequent on their informational advantages over other actors in field data collection.

3. One important game implication is that where interviewers are paid an amount that varies with their application of time, an interviewer who did not follow the take-the-easiest-first strategy would always look bad relative to those who did so.

4. See Aizcorbe, Kennickell and Moore [2003] for a summary of the data.

5. Unlike the area-probability cases, the respondents in the list sample are offered an initial opportunity to refuse participation by returning a postcard and 13 percent did so in the 2001 SCF.

6. The reason cases in the high wealth-index strata were not offered the monetary token is that past experience suggested offers of money might make such respondents more suspicious and might appear to be an insulting attempt to buy their time.

7. A small number of cases refused participation so strongly before the end of Phase II that effort on those cases ceased. For convenience, those cases are treated here as if they had fully completed Phase II. All calculations presented exclude cases in each subsample that were determined to be ineligible based on the sample design criteria; additionally in the case of the list sample, cases returning the refusal postcard are excluded.

8. The tract-level data derive from the 2000 Census of Population. The larger area used as the denominator in computing relative income is generally the MSA or county in which each tract is located.

9. Economic arguments alone suggest that the offer of a $20 respondent fee would have a stronger effect among such groups.

10. An active effort was made to remove interviewers who failed to meet minimum production goals or who failed to produce data of sufficient quality.