Stay Out of the Emergency Room:
A History of Coping and Improvement in the SCF

Arthur Kennickell[1]

[1]Assistant Director, Research and Statistics, Federal Reserve Board, Mail Stop 153, Washington, DC 20551; Arthur.Kennickell@frb.gov

Abstract

Viewed in a larger context, sometimes the worst problems become the greatest opportunities for learning. The Survey of Consumer Finances (SCF) could be seen as a methodological nightmare. Among other things, the survey must describe the highly skewed distribution of wealth, the subject of the survey is generally considered sensitive, the questions are necessarily at least somewhat technical, financial literacy varies across households, understanding of and attention to language more generally may also vary, the interview is long, and interviewers may vary in the extent to which they understand and follow the intended protocols; the joint consequence is that both unit and item nonresponse and other sources of nonsampling error have been important problems. There is an obvious immediate need to cope with such "emergency room" problems in order to have data for analysis, but for the long term it is more important to frame resolution in terms of steps toward prevention and other "public health" measures. Such reframing can lead to continuous improvement.

Key Words: Quality, continuous improvement, field operations, nonsampling error

1. Primary Subhead

Despite advances in the use of "big data," surveys remain a critical tool for scientific research, because of their potential for gathering a broad and coherent set of information from a sample that has a known mathematical connection with the target population. Given a questionnaire that reflects the goals of analysts and a logically defensible sample design, the outcome is the realization of a Monte Carlo process where, in theory, all inherent randomness of estimates is a consequence of the sample design; all valid sample units are reached, those units are "exposed" to the questions and the truth is revealed. Sadly, this is typically far from reality.

Viewed in a larger context, sometimes the worst problems become the greatest opportunities for learning. The Survey of Consumer Finances (SCF), the subject of this paper, could be seen as a methodological nightmare. Among other things, the survey must describe the highly skewed distribution of wealth, the subject of the survey is generally considered sensitive, the questions are necessarily at least somewhat technical, financial literacy varies across households, understanding of and attention to language more generally may also vary, the interview is long, and interviewers may vary in the extent to which they understand and follow the intended protocols; the joint consequence is that both

unit and item nonresponse and other sources of nonsampling error have been important problems.

There is an obvious immediate need to cope with such "emergency room" problems in order to have data for analysis, but for the long term it is more important to frame problem resolution in terms of steps toward prevention and other "public health" measures. Aside from the most basic forms of data processing to assemble an analytical data set and the efforts applied to limit the possibility of re-identifying respondents, all else we do in processing a survey is a necessity imposed by **failure**—in the questionnaire, the sample, preparation for potential problems faced in field work, or the support and execution of the field work. Reframing the issues in this way alters the causes steps toward solutions to be relatively highlighted. The approach also leverages the common psychological aversion to failure.

This paper describes the SCF and its approach to continual evaluation and improvement. The evolutionary result of this approach is a system that has provided the necessary support to allow the survey to remain relevant and flexible enough to cope with changing circumstances. The following section gives general background on the SCF. The third section discusses some of the principal problems in the survey and the necessary "reparative surgery." The fourth section takes these problems and addresses them in the context of plans for eliminating problems. The final session concludes.

2. Background on the SCF

The Federal Reserve Board (FRB) began modern SCF began in 1983, following a twenty-year gap from the landmark Survey of Financial Characteristics of Consumers which was led by Dorothy Projector at the FRB. The SCF had an important methodological redesign in 1989 and it has continued with an eye toward maintaining maximal analytical comparability since then.[1] The survey has been executed as a repeated cross-section every three years. In 2009, a panel re-interview was conducted with respondents to the 2007 survey in order to gather a picture of financial conditions in the Great Recession. Since 1992, the FRB has worked in close collaboration with NORC at the University of Chicago. Since 1995, the survey has been conducted using computer-assisted personal interviewing (CAPI).

The sample for the survey is based on a dual-frame design. A multi-stage area-probability sample provides robust national coverage and a good basis from which to measure financial characteristics that are broadly distributed in the population—such as credit card use. The survey also includes a list sample selected from statistical records derived from tax returns by the Statistics of Income Division (SOI) of the Internal Revenue Service; use of this information is governed by agreements with SOI that strictly control how the information may be used and guarantee protections for respondents selected using that information. This list sample is designed using a proxy for wealth to stratify the population of taxpayers and oversample wealth households (Kennickell (1999)).

The survey is long—it typically runs between 75 and 90 minutes, but may last several hours for someone with particularly complicated finances. The survey covers a wide variety of information on assets and liabilities, supporting and related characteristics, income, labor-

---

[1] See Bricker *et al.* (2012) for an overview of the survey design and recent results.

force participation, pension rights, income, use of financial institutions, attitudes, opinions, and a wide variety of demographic characteristics.

3. Coping with Three Principal Problems

In households surveys, theory can collide with reality in painful ways, largely because very little about a survey is of the nature of an engineering problem that can be mapped out under all contingencies in advance. Human behavior is at the very core of the survey measurement process. The situation is better describes as a type of principal-agent problem in economics.[2] Both interviewers and respondents face complicated incentives that may affect their behavior, and neither of them is easily subject to the sort of monitoring that would be necessary to establish an industrial process control. Respondents often have strong concerns about the control of their own time, and many may not be readily willing to devote the attention necessary to answer technical questions. The questions asked of respondents sometimes may not have an unambiguous answer, or there may be underlying conceptual variability that had not been anticipated. Moreover, there is often inherent ambiguity in words and their structuring in sentences that can lead to a variety of interpretations, and it may not be possible to find a single set of words that have the same effect on all types of people—that is, words themselves may have a distribution of meaning. Trust in the survey process and in the interviewer in particular is often a serious concern, especially in surveys like the SCF where the questions address sensitive matters. A respondent who does not fully accept the confidentiality protections promised will be less likely to cooperate. An interviewer who does not trust in the sincerity of the chain of management are less likely to be open about problems and less likely to induce trust in respondents.

Surveys very often fail in fulfilling the some of the key objectives necessary for the direct scientific validity of the process. Some eligible sample members may not participate, some respondents may not be willing or able to answer some questions, and the answers to some questions may be erroneous or highly questionable. Both fortunately and unfortunately, these failures attract a great deal of intellectual attention. We need highly developed skills with a scientific basis in order to rescue our "injured patient in the emergency room." However, the great attention to these areas tends to distract scientific discussion from the more important goal of eliminating, not just coping with, failure. Before turning to the goal of prevention and mitigation, I will sketch the surgical methods followed in the SCF for three problems; weighting adjustment for unit nonresponse, imputation for item nonresponse, and editing to address errors and inconsistencies in the data. These methods have been developed and refined progressively over 30 years.

The response rate in the area-probability part of the SCF sample has varied in a small range just below 70 percent. The comparable rate in the list sample is more difficult to describe compactly, as a consequence of the differential sampling over strata and some changes in the design over time; it is a reasonably close approximation to say that the response rate ranges from about 45 percent in the lowest wealth-index stratum to about 10 percent in the highest. Typically, we care about unit nonresponse for two reasons: it reduces the efficiency of our sample and it introduces the possibility of selection bias among the set of participants.[3] We might also care about nonresponse as an indicator of the activity taking

---

[2] See Kennickell (2006) for a discussion of agent-principal problems in this context.
[3] See Committee on National Statistics (2013) for a discussion of survey nonresponse, its effects and potential steps toward remediation.

place in the field, an idea elaborated in the next section. Like most surveys, the SCF attempts to mitigate the effects of nonresponse bias by making adjustments through a single set of analysis weights. In common with most other surveys, the SCF has to assume that, in essence, the observe population can be mapped into the sample population under an assumption that observed population differs from the target population only in terms of the proportions of cases in certain post-stratification or raking cells. In principle, nonresponse bias should be treated as an estimate-specific process and addressed specifically, but it is rare that sufficient information is available for this purpose. The unusual feature that the SCF brings to this exercise is the information inherent in the design of the list sample; by design, the list sample give a strong anchor on the distribution of wealth, a core objective of the survey (Kennickell, 1999). A long series of nonresponse-related work on the SCF has indicated a strong negative relationship between wealth and participation.[4] Although many people assume the list sample design for the SCF is simply aimed at increasing the efficiency of measurement of balance-sheet items held disproportionately by the wealthy, it is essential for developing a sufficiently credible to address unit nonresponse.

Nonresponse at the item level in a survey can also introduce important selectivity biases. Some attempts at triage use only complete-case analysis, an approach that can be particularly disfiguring in a wealth survey, where, for example, wealthier respondents by virtue of having more balance-sheet items also have more opportunities to have missing data. In the SCF, the patterns of missing data are not sufficiently similar to enable the use of simple generic methods for coping—such as maximum likelihood estimation. In addition, there are many complicated constraints that may apply for variables or sets of variables and there are "functional" relationships among variables. The SCF employs multiple imputation to fill in a distribution of values for each initially missing value (see Kennickell (1998)). Multiple imputation forces analysts to be more "honest" about what is known, because the variation in imputations makes it impossible to do otherwise. By providing multiple draws from the conditional distribution of the missing data, multiple imputation provides more efficient estimates and also guards against the potential damage from imputed outliers under single imputation. The method applied for most variables in the SCF is, in essence, a type of regression-based approach that proceeds through a defined sequence of variables and iterates through a number of such sequences. An attempt is made to condition broadly on the set of variables that have theoretical relationship to the variable addressed in a given model, as well as variables that might be indicative of effects related to the cooperativeness of the respondent and variables related to the sample design; this information may also include variables that would be too sensitive to release to the public. The approach also allows the dynamic imposition of a wide variety of constraints, including ones derived prior information determined directly from the respondent, as well as ones imposed by logical or institutional factors. To provide maximal information to users who may believe they could perform a more appropriate procedure for their analytical task, the data record includes a set of "shadow variables" parallel to the collected variables that defines the original status of the variable. Since the SCF began multiply imputing missing data in the 1989 survey, a variety of analytic procedures have been included in commercially available software.

In a wealth survey like the SCF, there may be many values that appear potentially implausible, in part because the skewness of the wealth distribution makes it possible for very large outliers to appear. Wealth surveys, by nature of the complexities of the information collected also enable possibilities for similarly complex error. Some such

---

[4] See Kennickell (2007) and references therein.

errors may violated clear-cut logical or plausibility checks, but others may take more effort to detect.[5] The SCF approach incorporates a variety of logical tools, both at the point of data collection and during the data-processing stage with the aim of identifying potential errors. A particular feature of the SCF is the reliance on interviewers to make comments on unusual situation or to clarify situations that may appear inconsistent; such comments may be made at any point during the interview, in a mandatory debriefing that interviewers much complete for each of their completed cases, or in the paradata. The SCF devotes substantial resources to review of each completed interview; beginning with the 2013 SCF, this effort includes staff at NORC, as well as the project team at the FRB that has long undertaken this work alone (see Bricker and Kennickell (2013)). Sometimes the text information from interviewers is sufficient to make a clean correction to the data, but more usually it is only indicative of a direction for deeper investigation of the data. Sometimes reported information is sufficiently unreliable that it must be set to missing and reimputed. Altogether, some aspects of editing clearly make a large difference in the analytical outcomes (see Kennickell (2006)), the effort entailed is a huge effort to address a basic failure in the data-collection process.

4. Planning to Eliminate and Avoid Problems

If others had not been foolish, we should be so.
William Blake, *Marriage of Heaven and Hell*

In the face of the sorts of problems discussed in the previous section, we often agonize and grieve as we work to repair damage, to the extent we can manage. Although there is a reflex to want to make changes necessary to eliminate problems, such effort often does not go beyond the most obvious level, at least until it becomes unambiguous that a major re-design is needed. Moreover, because many aspects of the repair process are intellectually challenging and open to technical innovation, the emergency room often gets more attention than the basic "public health" policies and mindset necessary to avoid or mitigate currently known problems and to maximize resiliency in the face of new problems.

Many years ago I organized a JSM session on multiple imputation, where the late Arnold Zellner was a substitute for an eminent researcher who at the last minute was unable to attend. Zellner had no obvious background in multiple imputation, other than simply being noted as a Bayesian. Nonetheless, he had many very intelligent insights to offer, but none was as electrifying as his opening challenge. As I understood him, he said that the focus on imputation was entirely wrong—we should be focused on the collection of the information we want, rather than methods for coping with failure. Being a very smart man Zellner was hardly naïve about surveys. Rather, the point was one of framing the work of information collection to ensure that the emphasis is always on quality from the beginning, rather than being added on imperfectly later on. As discussed in more detail later in this paper, such a perspective requires openness and a willingness to treat errors as a opportunities to look more deeply into possibilities for understanding and reform. In a complex system like most surveys are, progress must be a stepwise learning-by-doing process, because the full effects of changes often cannot be traced out to an ultimate logical conclusion *a priori*. This approach may be particularly important in a survey of wealth, in the face of continuing evolution of financial behavior. Success depending on a *continuing commitment* to evaluating, learning and improving. Many aspects of this process can take place through the implementation of relatively mechanical steps, but because surveys

---

[5] See Kennickell (1997,2000) for a close examination of the degree of missing data in the SCF.

depends intimately on such a complicated chain of human interactions, building an environment of trust and openness may be even more important. Here, I give some examples of this perspective on the three problem areas addressed in the preceding section.

Why do we have unit nonresponse? Major factors are respondents' suspicion about the legitimacy of the survey, their perceived lack of time or interest, miscommunications between the respondent and the interviewer, or misguided effort in the field work. Some of these and other factors may be correlated with a variety of observables that might provide a basis for weighting adjustments, such as age or income, but these factors are only indirect indicators of the underlying behavior. For example, older people are typically more likely to be at home and available for an interview; in addition to being used for weight adjustment, this might also inform a strategy for contacting respondents—as experienced field staff know well. Since the SCF first examined paradata keypunched from paper screeners in 1992, the study of paradata has guided the search for more effective ways to target individual respondents.[6] In 2001 when the SCF paradata began to be sufficiently to support detailed investigation, analysis showed that there was great diversity in the time until the first contact for all respondents and that the lag was much longer for respondents who ultimately did not complete an interview (figure 1). Based on this work and related results, a decision was made to focus more effort on initial contact and finding alternative ways to convey information when achieving contact exceeded a certain level of difficulty. Figure 2 shows a detailed sketch of the formal contacting developed to provide direction to interviewers. The contacting strategy divides the field work into three phases. The first phase calls for a range of initial contact attempts, and if that effort is not successful a well-crafted brochure with a targeted letter is sent to the respondent via express mail. A large amount of effort is put into the brochure and supporting material to produce a product that is eye-catching yet serious enough to clearly identify and address questions such as legitimacy and motivation for the survey directly.[7] The second phase is a structured follow-up to the mailing. The third phase is devoted to work much more specifically tailored to each case. The result of implementing the strategy was a more uniform application of effort and a more meaningful hope that the ethical responsibility of informing all respondents has been fulfilled. Another benefit was that by organizing the contact attempts in a way that reduced choice-based behavior in the first two phases, it provided a better support for the analysis of paradata free of selectivity effects that are very likely to affect analysis of paradata from unstructured contact attempts. This approach was initially resisted, but in a more simplified form it has remained as an effective part of the routine of SCF field work.
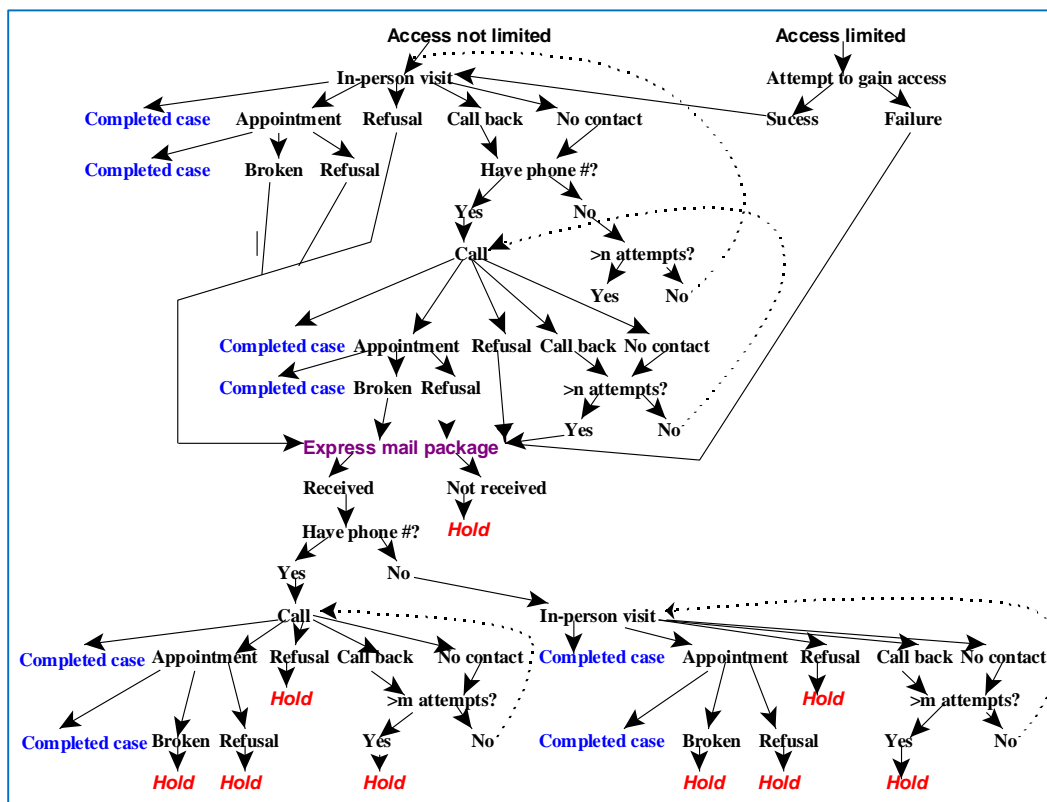
---

[6] See, for example, Kennickell 1998 and 2005.
[7] See Haggerty and Kennickell (2012).

Figure 1: Weeks until first contact, 2001 SCF;
Red line=completed cases; Blue line=incomplete cases.
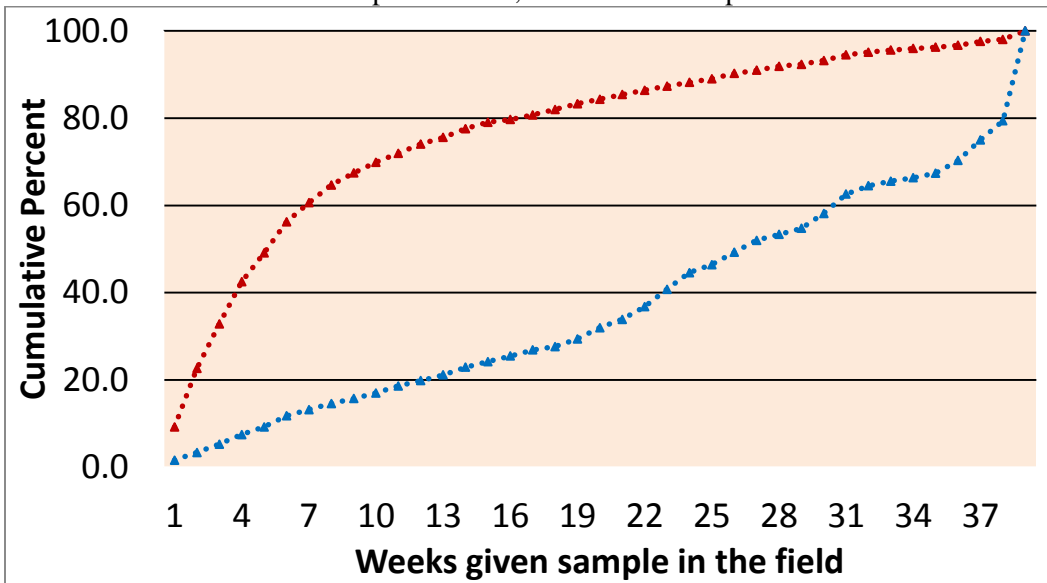

Figure 2: Formal contacting strategy for the SCF.

Why is there item nonresponse? Respondents may not trust the interviewer, the "system" or the government enough to answer sensitive questions; they may not understand the questions; they may not know the answers and be unwilling or unable to consult records;

they may find the questions too difficult to listen to carefully enough to answer meaningfully; the questions themselves may not have unambiguous answers; or the respondents may simply be impatient. The SCF takes a variety of steps to raise the level of trust between the respondent and the interviewer. Interview training focuses strongly on the value of persuasion, patience and clarification. The protocols also allow recourse to a knowledgeable proxy to answer *as if that person were the respondent*, when the respondent is unwilling or unable to answer. The CAPI interview frequently provides for entering respondents' answers in the format the use for reporting—for example allowing periodic payments to be reported in the frequency the respondent prefers or allowing reports of timing of event as an age, a year, or a number of years in the past or future. Second only to the interviewer as a powerful tool in reducing item nonresponse is a routine ("Dollar-Probe") built into the CAPI program to address the reporting of all dollar amounts (figure 3 provides a simplified sketch), where there is relatively often resistance to reporting or uncertainty about the answer.[8] The routine allow respondents to report a single value, or failing that to report their own ranges, ranges from a card or ranges generated in navigating a decision tree (e.g., "Is it greater than X?). At the confirmation step, the computer writes out reported amounts in words for interviewers to play back to respondents in order to ensure the information has been entered correctly

Figure 3: "Dollar-Probe" range and confirmation routine.
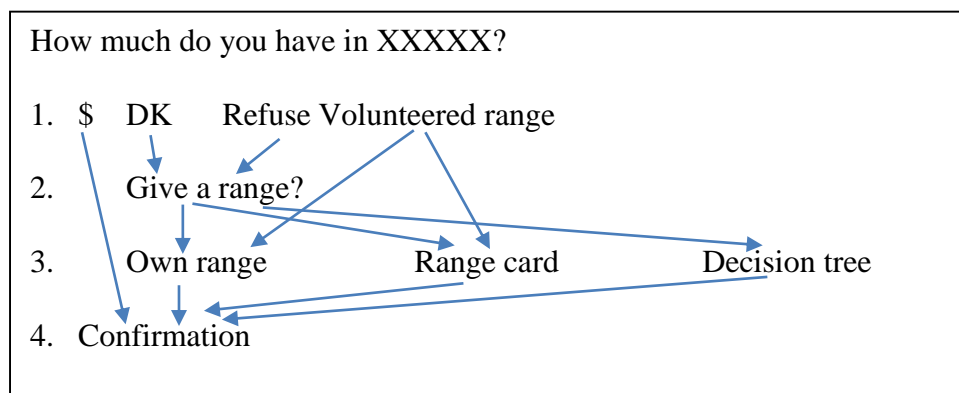
Dollar-Probe was constructed around the standard techniques interviewers were trained to follow before the introduction of CAPI. Interviewers had been trained to probe for ranges, allowing a respondent's range or an entry from a small card containing ranges. When Dollar-Probe was implemented, the data strongly suggested that it altered interviewers' behavior in a strong way. Whereas in the past, relatively many respondents would report they did not know the answer to a question, very few made such a report after the introduction of Dollar-Probe. At the same time, the refusal rate barely changed for most variables. The result has been that "partial answers" from ranges now have an important role in the SCF; the ranges serve a very important purpose in bounding imputations for item nonresponse.[9] Attention is still given to promoting full response, but Dolllar-Probe remains a powerful tool for limiting damage.

Why do we need to edit data, other than to validate legitimate outliers? Editing often reveals that the respondent did not understand a question, the interviewer did not understand the respondent or the question, the interviewer did not follow the appropriate

---

[8] See Kennickell (1997) for a detailed description and analysis.
[9] Each type of range is flagged so that the original, partial information may be extracted.

instructions, the question wording was ambiguous, the underlying concepts in the question shifted unexpectedly, or the situation was too confusing or difficult for either the respondent or the interviewer to cope with. All these reasons point to *communication failure*. Perhaps our questions were not doing what we thought they would and we need to revise, or reframe or re-sequence them, or perhaps we need to add or revise the instructions to the interviewer or alter the approach to training. Very substantial effort in the SCF has been devoted to tracking reporting errors as closely as possible to their sources in order to understand how data collection goes wrong. Sometimes, there is a clear solution that has broad application, but more often the process reveals more subtle problems that are not easily amenable to broad instruction or questionnaire changes.

In the SCF, there is a recognition that the interviewers are the ultimate gatekeepers for data quality—an honored and respected position in the survey. As in almost all surveys, an SCF interviewer is expected to read each question as written and to follow all instructions presented in the interview. Where the SCF approach departs from this norm in a serious way is through the empowerment of interviewers to clarify questions or blocks of questions for respondents—through formal means, such as reference to a glossary that is always available, or by directly intervening to insure that the respondent understands what is being asked and that the answers given are actually relevant to the questions asked. The interview also has the possibility at any point in the interview to enter clarifying information as text, or to provide an overview of problems in the debriefing required for each case. Without adequate training and reinforcement, opening the interview process to interviewers' own interpretations might be a disaster, particularly in a survey as technically complicated as the SCF. Initial training and continuing education on the SCF are central to its success.

Surveys are, in essence, communication engines that connect concepts with answers. The intervening process of conversation is often elaborately mediated between analysts and respondents—analysts know what they want (usually) and only interviewers are available to describe the ultimate interactions with respondents. The interviewers may have the most obvious and important intermediate role, but all other supporting and managing staff can affect the transmission of intent in both directions. Communication by such steps can lead to mangled intent, just as in the classic case of a successively translated and re-translated joke, which ultimately loses all humor. People left without a sense of intended purpose will inevitably and innocently "invent" the guidance they need to accomplish their tasks; this is another road to chaos. Similarly, analysts who have no way to be unaware of any problems associated with administering a given question may sometimes make very inappropriate interpretations of the survey data. Thus, bringing the understanding of the parties into as close as possible alignment is a high priority.

Binding all players in the survey process together in this way has been a continuing goal in the SCF. Trust is an essential ingredient, but it is not sufficient. Finding common tasks that lead to common understandings is the key. A long path of evolution on the SCF has led to increasingly close collaboration. For the 2013 SCF, the most recent one at the time of this writing, an important additional step was taken to involve staff at NORC directly in the process of editing the data.[10]

---

[10] See Bricker and Kennickell (2013) for a discussion of earlier approaches to editing in the SCF, the design of the new process for 2013, and plans for the future.

2. Summary and Conclusion

From the perspective of learning about survey measurement processes, the SCF is a beautiful dream. The problems have driven a long series of research and development aimed at achieving consistent and meaningful measurement and estimation. Among the technical initiatives, the sample design has been specifically tailored to address key nonresponse biases and multiple imputation has been developed for the survey to address missing (or partially missing) data. Technical control of this sort is difficult to achieve in most surveys, but there is at least broad supporting literature at this point. In some ways the great intellectual appeal of such technical procedures has been a distraction. A far more difficult and less well supported problem is managing the chain of human understanding from the conceptual development, through its implementation, to the point of response to the survey. The SCF has driven repeatedly in this direction, and it remains one of the most important areas for further research on the survey.

To remain relevant through time, a survey must adapt—most obviously to changes in the institutional structures in the world, to the words that become current for describing the relevant realities, and changes in the understanding people have of various concepts. But the supporting human processes need to evolve simultaneously through common activities aimed at evaluating, learning and improving. Only by keeping this process open to all participants is there a realistic hope for taking positive steps. Otherwise, the effort could easily degenerate into an empty process of "box checking." Evaluation depends in large part on making parts of the process more measurable, and creativity will always be needed there to provide this important basis.

Devoting our effort toward success should be the dominant perspective. It would be naïve to assume that all problems in a survey can be eliminated, but a change of perspective can sometimes be instrumental in moving in the direction of eliminating them. Accepting—and indeed welcoming—the task of engaging with human behavior will always be important in surveys of people.

Acknowledgements

Finally, I would like to thank the many thousands of SCF respondents, who are the ultimate source of knowledge about what works in a survey.

References

Athey, Leslie and Arthur Kennickell [2005] "Managing Data Quality on the 2004 Survey of Consumer Finances," http://www.federalreserve.gov/econresdata/scf/files/ Athey_Kennickell_101805.pdf

Bricker, Jesse; Kennickell, Arthur B.; Moore, Kevin B., and Sabelhaus, John [2012] "Changes in U.S. Family Finances from 2007 to 2010: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin*, http://www.federalreserve.gov/pubs/ bulletin/2012/pdf/scf12.pdf.

Bricker, Jesse and Arthur Kennickell [2013] "Shared Understanding and Data Quality in the SCF," *Proceedings of the Joint Statistical Meetings*, Montreal, Canada.

Committee on National Statistics [2013] Nonresponse in Social Science Surveys, Panel on a Research Agenda for the Future of Social Science Data Collection; Committee on National Statistics; Committee on National Statistics; Division on Behavioral and Social Sciences and Education; National Research Council, http://www.nap.edu/catalog.php?record_id=18293.

Haggerty, Catherine and Arthur Kennickell [2012] "The SCF: Creating a Consistent Culture of Quality," Proceedings of H2R, Techniques and Methodologies Section, American Statistical Association, Alexandria, VA.

Kennickell, Arthur B. [2008] "The Bitter End? The Close of the 2007 SCF Field Period," http://www.federalreserve.gov/econresdata/scf/files/asa2008complete1.pdf.

_____ [2006] "Who's Asking? Interviewers, Their Incentives, and Data Qualtiy in Field Surveys, http://www.federalreserve.gov/econresdata/scf/files/ iariw20067.pdf.

_____ [2006] "How Do We Know If We Aren't Looking? An Investigation of Data Quality in the 2004 SCF," http://www.federalreserve.gov/econresdata/scf/files/ asa20063.pdf.

_____ [2005] "Darkness Made Visible: Field Management and Nonresponse in the 2004 SCF," http://www.federalreserve.gov/econresdata/scf/files/asa20055.pdf.

_____[2000] "Measuring Data Quality In the 1998 Survey of Consumer Finances," http://www.federalreserve.gov/econresdata/scf/files/quality.pdf.

_____ [1999] "Using Income Data to Predict Wealth," http://www.federalreserve.gov/econresdata/scf/files/wealthincome7web.pdf.

_____ [1998] "Multiple Imputation in the Survey of Consumer Finances," http://www.federalreserve.gov/econresdata/scf/files/impute98.pdf.

_____ [1998] "Analysis of Nonresponse Effects in the 1995 Survey of Consumer Finances," http://www.federalreserve.gov/econresdata/scf/files/asa972.pdf.

_____ [1997] "Using Range Techniques with CAPI in the 1995 Survey of Consumer Finances," *http://www.federalreserve.gov/econresdata/scf/files/rangepap0197.pdf*.