

Discussion Papers

Lessons Learned
from the
National Survey
of America's Families

Tim Triplett

Laura Wherry

The Urban Institute

06-04

November
2006



Assessing
the New
Federalism

*An Urban Institute
Program to Assess
Changing Social
Policies*

Assessing the New Federalism is a multiyear Urban Institute project designed to analyze the devolution of responsibility for social programs from the federal government to the states, focusing primarily on health care, income security, employment and training programs, and social services. Researchers monitor program changes and fiscal developments. Olivia Golden is the project director. In collaboration with Child Trends, the project studies changes in family well-being. The project aims to provide timely, nonpartisan information to inform public debate and to help state and local decisionmakers carry out their new responsibilities more effectively.

Key components of the project include a household survey and studies of policies in 13 states, available at the Urban Institute's web site, <http://www.urban.org>. This paper is one in a series of discussion papers analyzing information from these and other sources.

The *Assessing the New Federalism* project is currently supported by The Annie E. Casey Foundation, The Robert Wood Johnson Foundation, the W. K. Kellogg Foundation, The John D. and Catherine T. MacArthur Foundation, and The Ford Foundation.

The authors thank the following people for their thoughts and comments: Natalie Abi-Habib, Greg Acs, Michelle Beadle, Marty Burt, Jeff Capizzano, Randy Capps, Rebecca Clark, Ken Finegold, Linda Giannarelli, Renee Hendley, Jenny Macomber, Adam Safir, Elaine Sorensen, Sharon Vandivere, Kevin Wang, Doug Wissoker, and Steve Zuckerman. They also give special thanks to Olivia Golden, Susan Kellam, Jenny Kenney, Rob Santos, and Sheila Zedlewski for reading, editing, and commenting on various drafts of this report.

The nonpartisan Urban Institute publishes studies, reports, and books on timely topics worthy of public consideration. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, its funders, or other authors in the series.

Publisher: The Urban Institute, 2100 M Street, NW, Washington, DC 20037

Copyright © 2006. Permission is granted for reproduction of this document, with attribution to the Urban Institute.

Contents

Executive Summary	v
Survey Overview	2
Major Accomplishments	7
Challenges Faced	31
Recommendations	44
Appendix	52
References	53
About the Author	56

Executive Summary

The *Assessing the New Federalism* (ANF) project of the Urban Institute and its partner, Child Trends, analyzed the experiences of low-income families and children over the past decade during major shifts in the nation's social welfare policies. The cornerstone of the ANF project was the National Survey of America's Families (NSAF), a survey of the economic, health, and social characteristics of children, adults under the age of 65, and their families.

The NSAF charted new territory by asking new questions, devising new methods of collecting data, and developing advanced estimating techniques. Over the course of a decade, NSAF data has generated nearly 500 ANF publications, plus dozens of journal articles, book chapters, research presentations, and many published and unpublished analyses by public users of the data. The \$40 million spent to conduct three rounds of the survey marked an unusual commitment of more than a dozen private philanthropic resources to a large survey and proved an efficient use of funds. The Bureau of Labor Statistics, by comparison, spent \$36 million in 2002 to conduct one year of the Current Population Survey (CPS).

This report summarizes the pioneering steps and major accomplishments of the methodology used to complete the NSAF, while identifying key challenges and important lessons for future household surveys. This report aims to present this information so audiences with a range of perspectives—including survey designers, researchers and academics, funders, and policymakers—may draw new insights in survey techniques from the NSAF experiences.

Conducted in three rounds—1997, 1999, and 2002—to gather information on more than 100,000 people and more than 40,000 families across the country, the NSAF gives researchers the tools to track national trends during that period, drawing on unusually detailed and comprehensive information about low-income parents and their children. In addition, the NSAF

provided significant samples in 13 states with a broad range of fiscal capacity, indicators of child well-being, and approaches to government programs. Using the NSAF, a researcher can compare outcomes for low-income families living in states with very different circumstances and policies.

State snapshots were not possible before 1997, since other major national household samples—such as the Current Population Survey—were not state-representative. While the CPS and the National Health Interview Survey (NHIS) had large representative samples, neither met the project’s needs because of limitations on content and the sample size of low-income families.

Future major surveys intended to inform low-income policy should build on what the NSAF taught those most closely involved in its design, implementation, and analysis. Close collaboration between policy and survey design experts proved essential, as did having flexibility to experiment when designing the best ways to enhance survey response. Both strategies work especially well in a multi-wave study like the NSAF, which offers the opportunity to learn from the experience of one wave to design the next. Specific to NSAF, both strategies were enhanced by the absence of procedural barriers and by the flexibility of a nongovernment survey. However, both lessons could be modified for a government setting.

Framing the Right Questions

NSAF allowed for the development and fine-tuning of questions that contributed to the policy debate in many ways. Some questions revealed new information about topics previously not well covered by any survey, including past welfare receipt, frequency of visitation by noncustodial fathers, and summer child care arrangements. The close involvement of policy experts in survey design helped ensure that the most central issues were addressed, often in innovative ways.

The 1997 NSAF provided the first national post-welfare reform picture by detailing the characteristics and well-being of adults and children who had left welfare within a two-year

window. And, the 2002 NSAF remains the most recent source on how these families fared. As the only national data source on noncustodial fathers, the NSAF spurred several states to introduce legislation to boost the earned income tax credit (EITC) as a reward for paying child support.

Some questions improved on approaches previously used in other surveys, demonstrating that old ways of asking questions were providing inaccurate information. A major example is health insurance coverage, where the 1999 NSAF round included a new question to verify lack of insurance. The estimates using the new approach were widely considered more reliable than estimates using earlier approaches. In fact, the CPS adopted the NSAF methodology in 2000 by adding a verification question to their health insurance protocol.

Some questions responded to emerging policy concerns and legislative developments. Because of the multiple survey rounds and the nongovernmental flexibility, questions could be added quickly as the policy environment changed. For example, in the second round, questions to assess respondents' awareness of the State Children's Health Insurance Programs (SCHIP) and Medicaid programs for children were added to obtain needed information on program implementation nationwide.

Experimenting for Maximum Reach

Experimentation on techniques to improve survey response spurred strong response rates and new knowledge that can improve other surveys. Wide arrays of strategies were tried over the three waves of the survey, with some of the tests structured as randomized experiments. For example, different financial incentive strategies were tried on a predicator sample (a small random portion of the overall sample released at the beginning of data collection) for the third wave of the NSAF and during a pilot-test to the third wave.

Another example was the specific protocol for interviewer call attempts to contact households. The strategy was revised after each round based on analysis of calling records to determine what patterns (for example, calling every day for a week, then holding for a week, then trying again for seven days) were most effective. As a result of this analysis and learning, the NSAF was able to contact more households in a shorter amount of time in the third round than in the first two, and it generated useful knowledge for the survey research field.

Major Achievements

The NSAF enhanced the policy debate on low-income families by defining—for the first time—large samples of low-income children and their families at the national and selected state levels. Six major attributes of the survey contributed to this enhancement:

- *Sample size and oversampling of low-income children and families.* The NSAF successfully surveyed a broad swath of low-income families with children, a population never fully captured before. While the Census' American Community Survey now captures a large sample size of low-income families, it does not convey the depth of information on child well-being. NSAF's large sample size permitted researchers to study family structure, living arrangements of children, child care arrangements, and welfare program participation among low-income families in ways not possible before the NSAF.
- *Focal states.* A critical goal of the NSAF was to provide information that could inform state policy toward low-income families, which meant zeroing in on policy-relevant subgroups. Doing this in 50 states would have been prohibitively expensive. After considerable discussion, only 13 focal states were selected.¹ As a result of this strategy, the NSAF enabled state-level analysis of the experiences of low-income families with children in states with different policies, demographic make-ups, and histories.

¹ These states were Alabama, California, Colorado, Florida, Massachusetts, Michigan, Minnesota, Mississippi, New Jersey, New York, Texas, Washington, and Wisconsin.

- *Dual-frame sampling design.* Concerns that low-income households might not have telephones led to an innovative approach to sampling that involved two separate sampling frames: a random digit dial (RDD) telephone sample frame (for the majority of the interviews) and a supplementary area sample conducted by in-person interviewers for households without telephones. The NSAF provided in-person respondents with cell phones to answer the questions, so the results would be more comparable to those of the RDD respondents. This represented one of the first uses of this approach on a large national survey, and the lessons learned should be helpful in future uses of this design.
- *Wide linkage of policy findings.* The NSAF's focus on a broad range of topics enabled researchers to link policy findings across disciplines. Over its duration, the NSAF provided critical data to understand the changes in program participation and well-being that accompanied changes in the social safety net. The challenge was to develop a broad instrument that respondents would complete. Some questions were asked only of a random adult in each household rather than all sampled adults.
- *Detailed data on living arrangements.* The NSAF contributed detailed data on household composition and the living arrangements of children. For example, the NSAF allowed researchers to distinguish between children living with their unmarried biological parents from those living with a biological parent and an unrelated cohabiting partner.
- *Online dissemination.* The creation of online statistical analysis tools, which included an online tutorial, made using the NSAF data very approachable. The NSAF online data files were first made publicly available in February 2004. Since then, about 3,600 people have registered to use the online public-use files. This number understates the total number of people who have used the online public-use data files; people may share their login with others in their organization, and those who registered to download the public-use data before February 2004 were not required to re-register to use the public-use files. Foundations in New York and California provided funds for local trainings to encourage use of the data, and an award-winning sociology textbook uses examples from the NSAF online statistical tools.²

² *Hands-On Sociology*, 3rd ed., by William Feigelman and Yih-Jin Young. This book uses the NSAF online statistical tools to teach beginning sociology students data analysis. *Hands-On Sociology* won the 2005 ICPSR Prize Competition for Best Instructional Module or Instructional Innovation in the Social Sciences and Social Science History.

Remaining Challenges

Despite the NSAF's advancements in data collection and analysis, the survey highlights five remaining challenges:

- *Area sample.* The dual sampling frame had a smaller-than-anticipated impact on final survey estimates. Designed to augment the telephone survey with a sample of low-income households without telephones, the design did not greatly expand coverage of low-income families. Still, the dual sampling frame did add to confidence that the NSAF had done everything possible to reach these households. Also, lessons learned about the dual sampling frame may be useful for a different reason in the future, given a changing telecommunications culture with younger and wealthier households that choose to communicate solely through wireless telephones. But as a tool for making the survey more representative of low-income families, some rethinking will be required.
- *Screening for low-income families.* Using a short screening interview to determine household size and family income was effective in increasing the number of low-income family interviews. However, asking income at the start of interview can increase the number of respondents refusing to take part in the study. An additional problem comes from the inaccuracy of income reports when a single income question is used to screen households.
- *Seasonal differences.* Data collection over the summer complicated later analysis of the NSAF, since both education and child care arrangements for families differ from those held during the school year. Adjustments to the survey instruments and the construction of separate weights for analysis were necessary.
- *Efforts to increase survey response.* Concerns that efforts to increase survey response might bias the results were carefully monitored and investigated throughout the study. With incentives, such as financial ones, becoming an increasingly important tool in dealing with nonresponse, these concerns about bias will be a bigger challenge in the future.
- *Questionnaire design.* NSAF questionnaire development had to balance comprehensibility with accuracy, a particularly challenging task given that many low-income family respondents lacked high school degrees. Survey developers tried to anticipate the respondent's interpretation of each question, familiarity with the terminology, and ability to

recall information. For example, the multiplicity of the various types of child care arrangements and income benefits programs confounded the terminology. For some families, the term Head Start conjures up child care settings that are not part of the official government Head Start program.

Recommendations for Future Surveys

- *To inform policy on low-income families, a large sample size is essential.* A sample size similar to or larger than the NSAF's is necessary to collect sufficient information for detailed analysis of children from low-income families at both the national and state levels. Plus, it would take a significant cut in sample size to make the survey much less costly. At that much smaller level, the sample would no longer allow for key policy analyses—the primary contribution of the NSAF.
- *Explore the possibility of using survey weights to account for coverage issues.* The area sample gave people without telephones a chance to be included in the study, but the costs associated with collecting these additional interviews were disproportionately high. The world has changed and so have the types of people and families that are not reachable by phone. Any future survey needs to think about who would be missed in a telephone survey and how can they be accounted for.
- *Explore the possibility of screening for low-income households using administrative records to get a larger sample.* While the NSAF successfully prescreened households by asking income during the screener interview, this procedure lowered participation, increased costs, and was subject to respondent error. These three problems combined with the improved availability of information in today's telephone sampling frames suggest that one may be able to oversample low-income households by oversampling telephone exchanges that are known to have a high percentage of low-income households.
- *Increase survey participation with new methods.* The NSAF experience indicates that it is increasingly more difficult to convince people to complete survey interviews. Since more people are refusing surveys, sophisticated approaches to refusal conversion are needed, such as only trying to convert a sample of those who initially refuse, which worked well in the third wave of the NSAF.

Lessons Learned from the National Survey of America's Families

The National Survey of America's Families (NSAF) responded to a need for empirical data on major changes in social policy at the federal and state levels. Plans for the survey developed in the early to mid-1990s as Congress pushed for more federalism across many different areas of social policy. While not all the anticipated decentralization actually happened, by the time President Clinton signed the Personal Responsibility and Work Opportunity Reconciliation Act in 1996, survey questions were formulated to monitor and analyze the well-being of American children and families across health care, income security, social services, and job-training programs for low-income Americans.

The multiyear *Assessing the New Federalism* (ANF) project at the Urban Institute, carried out in partnership with Child Trends, was charged with monitoring the experiences of low-income families and children during these major changes in the nation's social welfare policies. A key element of the ANF project was a survey designed to explore national trends and links between state policies and state-level estimates of child, nonelderly adult, and family well-being indicators. The ANF project strived to ensure that state-level policymaking activities benefited from the latest research and data on cash assistance, child care, child welfare, child support, health insurance coverage, Medicaid, and the State Children's Health Insurance Program, along with whether these programs forged into an effective safety net. The NSAF gave ANF the data to do this.

A national household survey, the NSAF provides detailed information on the economic, health, and social characteristics of children, adults under the age of 65, and their families. The survey was administered three times—in 1997, 1999, and 2002. In each round, interviews were conducted with over 40,000 families, yielding information on more than 100,000 persons under

the age of 65. The survey sample is representative of the civilian, noninstitutionalized population under the age of 65 in 13 states and the balance of the nation.

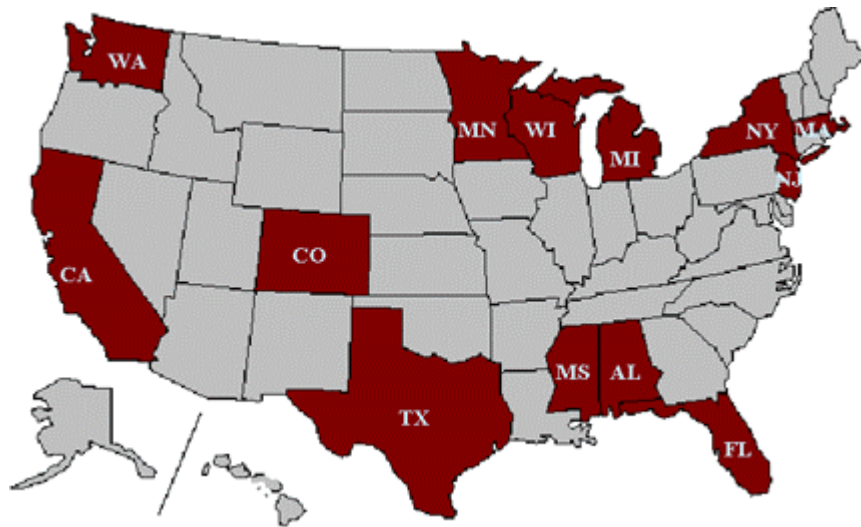
Survey Overview

Selection of Focal States and Sample Size

The goal of the NSAF sample design was to yield interviews representative of children, adults under the age of 65, and their families in individual states and the balance of the nation, with oversamples of low-income families and families with children.³ The costs of completing enough interviews in all states would have been prohibitive. Instead, ANF selected 13 focal states: Alabama, California, Colorado, Florida, Massachusetts, Michigan, Minnesota, Mississippi, New Jersey, New York, Texas, Washington, and Wisconsin. Collectively these 13 states account for over half the U.S. population, and they represent a broad array of government programs, fiscal capacity, and demographic characteristics. Beyond the targeted study areas, the sample design also included a sample from the balance of the nation to permit accurate estimates for the United States.

³ In the 1997 and 1999 rounds of the NSAF, Wisconsin was targeted for particularly intensive study, with separate large samples drawn for Milwaukee and the rest of the state.

Figure 1. Targeted NSAF States



In each round of data collection, more than 40,000 families completed at least one full-length interview. The three rounds of data collection provided detailed social and economic information for 220,821 adults and 104,709 children. The adult sample was limited to adults under 65 years old, unless a person 65 years or older was identified as the primary care taker for a sampled child.

As with virtually all household surveys, some important segments of the population (e.g., the homeless) could not be sampled because of their living arrangements. A small fraction of the sample consisted of “linguistically isolated” households, where no one in the household spoke either English or Spanish. Families in these living arrangements were not interviewed.

Screening for Low-Income and Child Households

Since many of the programs and policy changes that the ANF project monitored were expected to have the greatest impact on low-income families, the NSAF sample design included oversamples of families with incomes below 200 percent of the federal poverty level and families with children under 18 years old. To increase the final sample size of low-income families and families with children, the NSAF subsampled, and then eliminated from the survey,

both high-income households (those with incomes above 200 percent of the federal poverty level) and adult-only households.⁴ This method of subsampling to eliminate households, rather than increasing the overall sample size, conserves survey administration costs.

Dual-Frame Sampling Procedure

Two separate sampling frames were used to accomplish NSAF sample design goals. The majority of NSAF interviews were completed under a random digit dial (RDD) sample frame, which yielded interviews in households by telephone. The RDD approach was adopted as a cost effective means to collect data, as opposed to an in-person interview approach. Telephone interviewers located in centralized facilities conducted all interviews using computer-assisted telephone interviewing (CATI). Westat,⁵ a premier survey research firm, helped with the survey design and completed all interviews and data collection.

Since households without telephone service contain a disproportionate number of low-income children, a supplementary area sample was conducted in person for those households without telephones. Nationally, Giesbrecht, Kulp, and Starer (1996) estimate that about 20 percent of families in poverty have no telephone and that about 10 percent of families with one child age 3 or under have no telephone. Relying on the RDD sample frame alone could have produced biased estimates, particularly for persons in low-income families.

Using a multistage area probability sample to identify households without telephones, the sampling frame used for the area sample consisted of census block groups with a relatively high number of nontelephone households, based on 1990 census estimates.⁶

⁴ The subsampling rates for higher income households varied by focal state to account for the different poverty rates and household compositions in each state. All sampled households without a land-line telephone, however, were considered eligible for interview regardless of family income or presence of children.

⁵ Westat was chosen to help with the survey design and collect the data through a competitive bid procedure.

⁶ Wireless telephone-only households were treated as nontelephone households.

Interview Structure

Interviews were conducted in two stages for all households. First, a short, five-minute screening interview was administered to identify the appropriate respondent(s) in the household, identify households with low incomes and children, and determine eligibility for the extended interview. The extended interview was divided into 16 sections by content. The questionnaire was designed to begin with less sensitive topics and then progress to more sensitive topics. It also assumed that parents would be more likely to talk about their children first rather than themselves, therefore questions about children's health and education were asked before questions about household membership and relationships. A comprehensive list of survey topics is shown in the appendix.

The NSAF extended interview ranged from 25 to 50 minutes long, depending on whether the questions were intended to ask about a single adult or about children in the household. The goal was to keep the entire survey—the screener and the extended interview—under one hour for each interview to reduce respondent fatigue and preserve quality of response, as well as to decrease overall rates of nonresponse.

Procedures for Selecting Persons within a Household

Household members were subsampled to reduce the length of time to complete interviews with each household. If there were multiple children under age 6, one was randomly selected to be a “focal child.” Similarly, only one child age 6 to 17 was sampled in a household. The separate sampling of children less than 6 years old and children 6 to 17 years old was completed to ensure that data collected could be used to assess the well-being of both younger and older children.

For each focal child, the individual identified as the most knowledgeable adult (MKA) in the household for the child was interviewed about the child. During the MKA interview, additional data were collected about the MKA and about the MKA's spouse or partner, if that

person was living in the same household. Serving as a proxy, the MKA provided all data collected about the spouse or partner.

Two other within-household respondent selection steps were used in all three rounds of data collection. Other adults living in households with children (adults who were not the MKA of any child in the household) and adults in adult-only households were also subsampled. Self-response was required for sample adults. During the interview with a sample adult, additional data were collected about the sample adult's spouse or partner, if living in the same household.

Data Collection

Data collection for each round began in early February, around the time individuals receive the income information needed to prepare income tax returns. Like the CPS, the NSAF asked about previous-year income. Data collection ended approximately nine months later in early November. The length of the data collection period had a small but measurable impact on the quality of information collected on family income. The interviews collected in the final four months of data collection produced slightly higher estimates of family earnings. This is consistent with Census Bureau findings that indicate a longer recall period tends to increase respondents' estimates of actual earnings (Pedace and Bates 2001). By starting the NSAF data collection process in February, the survey was able to reduce this recall bias for the majority of interviews, since 60 percent of extended interviews were completed during the first four months of the study.

Questionnaire Content

The content of the NSAF questionnaire was designed to be broad and to cover the many topic areas followed by the ANF project. Designers relied on questions from existing surveys as much

as possible to maintain comparability with other surveys, while new questions were added to measure the effects of anticipated policy changes.

Content items in the NSAF included the following:

- Household composition, demographics
- Health status, insurance, access, usage, knowledge, and awareness of programs
- Employment, earnings, income, poverty status, child support receipt and payments
- Welfare, Food Stamp, and other program participation
- Child care arrangements and costs
- Child and family well-being measures (housing hardship and food security)

Weights

Responses to NSAF items were weighted to provide approximately unbiased aggregate estimates for each study area and for the country as a whole. The weights were applied to all survey items in an effort to

- compensate for differential probabilities of selection for households and persons;
- reduce biases occurring where nonrespondents have different characteristics than respondents;
- adjust, to the extent possible, for undercoverage in the sampling frames and in the conduct of the survey, and
- reduce the variance of the estimates by using auxiliary information.

Additional detail on the development of survey weights is available in NSAF methodology report #3 in all three survey rounds.

Major Accomplishments

This section summarizes where the survey broke new ground and what was accomplished.

Large Sample Size Allowed for Low-Income and Child Oversamples

Completing a large number of interviews was perhaps the greatest achievement of the NSAF sample design. The large sample sizes made it possible to obtain reliable survey estimates for persons and families below 200 percent of the federal poverty threshold, and much of the NSAF analysis included comparisons between high- and low-income families. Gathering detailed information on families and their experiences, the survey permitted researchers to study family structure, child care arrangements, and welfare program participation among low-income families in a depth not possible before the NSAF.

By increasing the sample size of families with children under age 18, the NSAF allowed for a very precise snapshot of the well-being of both younger and older children. The large sample size of school-age children (age 6 to 17) was particularly important for new research conducted on child care and child care arrangements using the NSAF. The large sample of children also made possible, for example, analyses of groups of children who are important in policy terms but not frequent enough to be analyzed in smaller samples. Those would include children living with grandparents, immigrant children in mixed-status families, and children of former welfare recipients.

Large Focal State Sample Sizes

The large sample sizes also provided representative estimates in 13 focal states. With the data sources available in 1997 through 2002, the NSAF was among the few surveys to provide reliable state-specific estimates on a wide range of well-being indicators. Many surveys during this period either had sample sizes that were not state-representative or had very narrow survey content. In addition, the small sample sizes available for examining low-income households would make comparisons over time very imprecise for most states.

The large sample sizes allowed the NSAF a small average margin of error for state-level estimates of low-income children and adults in all rounds. The margin of error indicates the imprecision inherent in deriving estimates from a sample of the population, given that the sample is just one of many possible samples that could have been selected in the population. In general, a larger sample size enables researchers to answer research questions with greater confidence and precision, thereby decreasing the margin of error. Table 2 in the appendix presents the average margins of error for NSAF estimates, calculated at the 95 percent confidence level.

While the large focal state sample sizes provided researchers with the ability to produce reliable state estimates, the ability to do substate analyses was fairly limited. By pooling data across rounds, however, it is possible to do additional subregion or subgroup analyses. In response to interest in California for regional analysis, the NSAF survey team merged data from the three waves of the NSAF to create an online data file large enough to permit analysis of a number of substate regions or of more detailed subgroups of children.

Two Sample Frames Accomplished Sample Design Goals

The dual-frame sampling procedure used in the NSAF is rare among household surveys because of the cost and complexity of merging interviews from two different sampling frames as well as the costs associated with using two sampling frames. In fact, the procedure's use in the 1997 round of the NSAF represented one of the first uses of the dual-frame design on a large national survey.

Additionally, the NSAF successfully introduced the use of new cellular technology to minimize mode effects resulting from differences between the RDD and area samples and potential effects on the quality of the data. In-person interviewers provided cellular telephones to respondents in nontelephone households to connect them with interviewing centers for the CATI

interview. As such, interviews were conducted in essentially the same manner in both telephone and nontelephone households. The use of cellular phones reduced the complexity involved in merging field and telephone survey data, and it reduced the potential for mode effect bias—research indicates that respondents are less willing to answer sensitive items when being asked directly by an interviewer versus over the phone by an interviewer (Groves and Kahn 1979).

Cross-cutting Nature of the Survey Topics and Breadth

The NSAF placed a premium on broadly measuring the well-being of children, adults, and families. The survey's breadth permits analyses of many different policy areas, including income support, health care, job training, social services, welfare and food stamp programs, the Children's Health Insurance Program, child care, employment and training, and child support. This broad focus also enabled policy findings to be linked across disciplines. For instance, the 1996 welfare reform legislation removed the link between welfare and Medicaid, and it led ANF researchers to create survey questions in the NSAF allowing them to track the impact of that change.⁷ Existing state-representative surveys did not include variables related to anticipated policy changes.

Being able to track child and family well-being from the start added validity and nuance to the story researchers were able to tell about how people fared in the new federalism. A five-point depression scale and measures of economic well-being were added to other poverty measures. Traditional government surveys, such as CPS and the Survey of Income and Program Participation (SIPP), do not include these questions.

Questions on participation in a range of government programs allowed researchers to investigate the overall functioning of the nation's social safety net. An example of this was

⁷ In conjunction with detailed case studies conducted in each of the 13 focal states.

analysis that measured the number of uninsured children in families who were enrolled in other public programs. Results indicated that three-quarters of all low-income uninsured children were enrolled in the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), the National School Lunch program, and other government programs. This finding led to federal legislation trying to better align programs.⁸

Flexibility Met Emerging Policy Needs

The content of the NSAF evolved between the mid-1990s and 2002 to meet some of the key policy data needs of the period. Because the survey was privately funded, it was able to bypass some procedures that must be met by government surveys to make questionnaire revisions. This enabled NSAF researchers to quickly adapt and monitor critical emerging policy issues. For example, several late additions were made to the health care component of the survey in the second round. The changes included a series of questions that focused on respondents' awareness of the State Children's Health Insurance Programs (SCHIP) and Medicaid programs for their children. These questions aimed to obtain needed information on the many SCHIP programs being implemented around the country.

In the third round, the survey adopted new questions to examine the knowledge of new time limits on welfare assistance among program participants, given concerns with changes in the quality of family life that may have occurred as a result of welfare reform. Other questions in the NSAF allowed researchers to monitor former welfare recipients and their participation in public programs.

Another example was a qualitative follow-up survey conducted with 169 families that had no current employment or cash government assistance and income below 50 percent of the

⁸ In early 2000, the Clinton administration used ANF research to justify changing the privacy rules of the School Lunch Program. The changes permitted administrators to share the names of children participating in the School Lunch Program with administrators of the Medicaid and SCHIP programs.

federal poverty level in the prior year. The extreme low-income qualitative study aimed to find out whether the income status families reported on a large complex survey were accurate and to discover how individuals truly living without any income or government cash assistance were coping. This follow-up study of extreme low-income families was conducted and completed during the third round of the NSAF.

The low-income follow-up study highlighted differences between information that can be collected through an unstructured interview without time constraints and information collected through a highly structured interview with a time constraint. Qualitative interviewers had both the time and flexibility to probe for sources of income by verifying the original NSAF data and using clues provided about how the family covered expenses. Researchers gained considerable insight into 95 families that were coping without employment or government cash assistance. The information collected in the qualitative interviews complemented and extended the information collected in the NSAF to provide a more complete profile of these families' circumstances.

Children's Living Arrangements and Family Structure

A contribution of the NSAF that was vital to analysis across a range of topics was detailed data on household composition and the living arrangements of children. The household roster identified every member of the household and his or her relationship to every other member of the household. In contrast, other large household surveys, including the Census, CPS, and NHIS, collect spouse or partner information for adults, parent information for children, and the relationship to the householder.

A 2001 Federal Interagency Forum on Child and Family Statistics praised NSAF's household roster, pointing out the advantages of being able to identify biological parents among

cohabiting couples. Also, marriage, cohabitation, child support, and the relationship between child well-being and family structure rose in policy importance. Information on the living arrangements of children was particularly important in the context of welfare reform, which included the stated goals of encouraging marriage and raising the share of children living with two parents. Round three of the NSAF expanded the information available to researchers by adding questions about the length of the marriage or cohabiting relationship. Also, since the survey contains a rich set of child and adult well-being measures, it enabled researchers to monitor the well-being of children across a variety of living arrangements.

Experimentation to Improve Survey Response

Before and during data collection, constant efforts were made to improve the NSAF's response rate. While there is no such thing as an "official" acceptable response rate for a survey, response rates are the industry's standard by which people judge the quality of a survey. The lower a survey's response rate, the more likely there will be bias in the data due to nonresponse (if the people who responded to the survey answered questions differently than those who did not).

Table 1 below shows the NSAF response rates by round, with separate calculations for the screener interviews, extended adult interviews, and extended child interviews. Response rates for the screener interview declined in the third round of data collection for the NSAF, despite specific efforts implemented to improve response rates for this round. The decline in the screener response rate, however, did mirror the overall trend toward declining responses that other telephone surveys were experiencing at the time. With changing trends in technology and the increase in cellular phone use, more households are opting not to use a landline telephone and response rates are falling for the traditional telephone survey.

In addition, table 1 shows how the response rate was much lower for the RDD telephone sample than the area probability sample. This is most likely because it is harder for an individual to decline to participate in person than over the phone (Newcomer and Triplett 2004).

Table 1. NSAF National Response Rates, by Round (percent)

Sample Type	Round 1	Round 2	Round 3
Overall response rate (child)	65.1	62.4	55.1
Overall response rate (adult)	61.8	59.4	51.9
Screeners versus extended			
Screeners	77.8	76.7	66.0
Child extended	84.1	81.4	83.5
Adult extended	76.9	77.5	78.7
RDD sample only			
Screeners	77.4	76.3	65.5
Child extended	83.4	80.5	83.0
Adult extended	79.4	77.0	78.2
Area sample only			
Screeners	87.0	89.2	81.6
Child extended	95.7	96.2	95.0
Adult extended	92.5	93.0	92.5

The techniques used to minimize nonresponse evolved during the course of the survey. Through the data collection strategies highlighted below, survey designers aimed to increase the likelihood and willingness of individuals to participate in the survey.

Advance letters

Before contacting households by phone, an advance letter was mailed to households describing the purpose and importance of the NSAF. The NSAF advance letter emphasized that the study was evaluating how recent federal policies affect health care, education, and employment services. Surprisingly, the use of advance letters is relatively new to general population telephone surveys. But, technological advancements have led to recent improvements in reverse directory services, which have both reduced the cost of obtaining addresses based on telephone numbers and increased the reliability of the addresses obtained.

Before data collection, all telephone numbers selected for sampling were sent to a reverse directory service to obtain addresses, and an advance mailing was sent to all households for which we were able to obtain addresses.⁹ This mailing included a letter from the project coordinator encouraging the household's participation and briefly described the survey and its importance. Answers to frequently asked questions were included in the mailing. The mailings were timed to coincide as closely as possible with the interviewer's initial contact with the household. For the nontelephone households, copies of the mailing were carried by interviewers and given to respondents as needed.

Refusal conversion letters

Refusal conversion is an important aspect of the NSAF's overall response rate maximization effort. Whenever a respondent initially refused to complete an NSAF interview, the project supervisor reviewed the case. Nonhostile refusals were returned to interviewers specially trained in refusal conversion for additional calls to the household.

Follow-up letters were sent to households that refused to be interviewed for either the screener or the extended interview. This "refusal conversion letter" provided additional information about the survey and stressed the importance of participation for respondents who could not be contacted, were too busy to participate, or did not express interest in participating. For the nontelephone sample, a letter sometimes needed to be sent to the managers of limited-access buildings and communities in order to grant field interviewers access.¹⁰

Tailoring refusal conversion letters has long been a strategy of convincing reluctant respondents to participate, and it is supported by research (Dillman 2006). Since mail addresses

⁹ In round 1, addresses were obtained for 38 percent of the numbers.

¹⁰ These letters and refusal conversion letters for the nontelephone sample respondents were requested by the regional supervisor but sent from the Westat's main office in Rockville in hopes of adding more credibility to the study.

were available for most households, the NSAF found a valuable opportunity for tailoring. In each mailing, research findings were included that showed important first-round NSAF results for the respondent's state or region of residence. For instance, respondents living in Colorado were sent a one-page summary of findings describing the well-being of children, adults, and families in Colorado, while a Florida respondent received similar information specific to Florida. Survey administrators found this a very effective tool in increasing the saliency of the study and thus increasing likelihood of respondent participation.

Screening households

In developing the screener questionnaire, several design principles were used to ensure a minimal amount of nonresponse. First, the introduction needed to keep the respondent's interest in the study, as well as to motivate him or her to answer subsequent questions. Previous experience on similar screeners for other studies has shown that more than 80 percent of the refusals occur during the introduction or on the first question. Most people make up their minds to discontinue participation based on what is said during this first part of the interview. Therefore, a large portion of the development effort focused on experimenting with different introductions to the survey.¹¹

Second, information had to be collected in ways that were not perceived as insensitive. To determine eligibility for the NSAF, it was necessary that respondents be asked about the income of persons in the household. To construct the sample, it was important to collect information about all individuals living in the household, thus getting names or initials along with ages. Because adults tend to be very protective, collecting this information for children had to be handled especially carefully. Finally, specific information about the identity of the MKA or

¹¹ For additional information on the tests used to select the introduction and additional considerations, please see Vaden-Kiernan et al. (1999).

Option B adult (e.g., name) was important;¹² this information served as the only link between the screener and the extended interview.

The general approach taken to collect this information was to (1) screen households as fast as possible (the screener took approximately three to five minutes to administer), (2) try to develop rapport with the respondent before asking the most sensitive questions, and (3) ask for only the minimum amount of information necessary.

Monetary incentives

Various monetary incentive strategies were used to increase the response rate for both the screener and extended interviews. Information on the use of incentives for large household surveys was an unknown factor, since all other large household surveys conducted in the United States were either conducted by the federal government or received some level of federal funding and were thus not permitted to pay respondents. Therefore, in round 1 of the NSAF, a number of monetary incentive strategies were tested to optimize response rates with the available funding.¹³

Screener interview incentives

Based on the experimental results, the primary monetary incentive strategy developed to encourage participants to complete the screener survey in rounds 1 and 2 was to express mail letters containing \$5 to respondents who refused to participate at the screener level. If a second refusal to complete the screener was made by the respondent, another letter was sent by regular mail to urge their cooperation. This second letter contained no money.

In round three, instead of sending \$5 to households that refused, all households for which addresses were obtained were sent a \$2 bill. This change in strategy was developed based on the

¹² “Option B” is used to describe interviews with non-MKA adults sampled in households with children (sometimes called option B stragglers), such as adult siblings, grandparents, aunts, and other relatives, and interviews in sampled households without children. In households without children, one or two option B interviews were conducted depending on the number of adults present in the household.

¹³ These experiments, as described in Brick, Flores-Cervantes, and Cantor (1998), were conducted in spring 1997, so refusal conversion on this group did not start until June 1997.

results of a randomized experiment conducted on the pretest for the 2002 NSAF study. The results indicated that sending everyone \$2 in advance was more effective in garnering participation than offering \$5 to those respondents who refuse to participate (Cantor et al. 2003). Receiving a little money in advance may help convince participants that the additional promised incentives will be paid.

Extended interview incentives

In all three rounds, but especially in rounds two and three of the survey, an additional monetary incentive strategy was used for families that were screened and determined eligible but who refused to complete the extended interview. The extended interview refusal conversion strategy differed depending on whether a mailing address was available for the family, and the procedures were based on experience and testing completed with the extended interview in round one of the NSAF. If there was no address available for the household, the respondent was promised, over the telephone, \$25 to complete the extended interview. Alternatively, if the household did have an address available, an express letter was sent containing a \$5 incentive before attempting refusal conversion. In this latter course, during the actual attempt to convert the refusal, the respondent was promised an additional \$20 to complete the interview.

When faced with a second refusal at the extended interview level, a letter was sent to those households with an address but with no additional money. At the time of conversion, interviewers continued to promise either \$20 or \$25 to complete the extended interview (depending on whether there was an address available for the household). During the final month of the study in 1999 and the final two months of the study in 2002, incentives were increased to as much \$50 for households that were screened as low-income but refused to complete the extended interview.

Double sampling

In the third round of the NSAF, a new double sampling strategy was used with households that refused to participate in the survey.¹⁴ With this strategy, the NSAF was successful in keeping the project on schedule despite the increasing number of households refusing to participate in the study.

The double sampling strategy divided the survey sample into two groups, those who initially refused to complete the survey and those who did not refuse. The cost and time associated with completing interviews in households that initially refused to participate was assumed to be greater. Survey administrators attempted to complete interviews with all households without refusals, while attempting to complete interviews in some fraction (approximately 80 percent in the NSAF study) of households that initially refused to participate. Resources saved from sampling for refusal conversion will effectively increase the overall size of the sample that can be built in a survey given a fixed budget.

Since most NSAF refusals occurred during the screener interview, the double sampling strategy, or sampling of initial refusals, was applied only to households that refused the screening interview. The strategy randomly divided the RDD telephone sample into 101 replicates (independent slices of the overall telephone sample). The first 81 replicates were designated as the refusal conversion sample, meaning if a refusal occurred then standard refusal conversion efforts were used. The final 20 replicates were designated as the nonrefusal conversion sample. If a screener was refused, no effort was made to convert the refusal.¹⁵

¹⁴ This approach was adopted because of lower response rates in the third round compared with the earlier rounds, and the anticipated effect on the survey completion time schedule. Without implementing double sampling, the third round of the NSAF data collection was projected to fall short in terms of interviews completed during the time schedule and within the survey budget.

¹⁵ The fraction of nonrespondents who received standard refusal conversion was actually slightly higher than 80 percent (81/101 replicates).

Call schedule

The decision to collect the NSAF data over an extended period was intended to spread out the interviewers' attempts to call respondents, a technique that has been shown to improve survey response rates (Cunningham et al. 2003). The nine-month data collection period provided ample time to contact even the most difficult-to-reach families.

Finally, an integral step to obtaining high response rates was deciding on the frequency of interviewer call attempts to contact households. The approach used evolved over the course of the survey as more information on successful strategies became available. In the first round of the NSAF, a call scheduling protocol was developed that required seven call attempts for each household to establish initial contact. When a person answered the call, the interviewer verified that the number reached was a residential household. If necessary, a callback time was scheduled with the responding person.

If connection was not established after seven call attempts, the telephone number was put aside for a couple of weeks and then re-released to interviewers for a repeat of the seven-call sequence. Many of the numbers that were still not contacted after 14 calls were put on hold and then re-released for seven more call attempts. Analysis of the calling records from the first round found the percent of telephone numbers that resulted in a successful contact decreased with the number of call attempts. However, higher contact rates occurred immediately following the delays between calls after the 7th and 14th call attempts.

Based on these findings, in round two the calling pattern was revised to cause a delay between the seventh and eighth calls and another delay between the eighth and ninth calls. The total number of allowed attempts was reduced from the minimum of 14 used in round one to 9

for round two.¹⁶ In round three, the call pattern was again revised to include delay periods following the fourth and seventh call attempts, with a maximum number of nine attempts. With this approach, the NSAF was able to contact more households in a shorter amount of time than in the previous two rounds, and to eventually have fewer numbers with a noncontact status.¹⁷

Contributions to Policy Research

In the wake of welfare reform in 1996, the NSAF provided critical data to understand the changes in program participation and family well-being. Standout questionnaire contributions and their implications for policy research are discussed below.

Welfare caseload

The NSAF paved new ground in welfare research by developing a way to use a cross-sectional survey to mimic the historical information traditionally gathered in a longitudinal survey. Survey questions asked respondents not only about their current participation in the welfare program, but also detailed information about their past participation. Since the NSAF's implementation of these retrospective questions, the Current Population Survey has followed suit and expanded its questions on welfare receipt. In addition, by oversampling low-income families, the NSAF generated a nationally representative sample of welfare recipients and provided a picture of the characteristics of these families. Previously, researchers generally used CPS data on low-income single mothers as a proxy for welfare recipients.

In addition, the NSAF provides a national picture of the characteristics and well-being of adults and children that have left welfare within a two-year window. The 1997 NSAF provided

¹⁶ A random sample of phone numbers in all three rounds of the data collection that were never successfully contacted was called continually throughout the study to support the estimate of the proportion of all telephone numbers that were residential numbers. This estimate was needed to accurately estimate the final survey response rate.

¹⁷ As in prior rounds, a subsample of the noncontacted telephone numbers were called additional times to support estimating what portion of telephone number are residential telephone numbers for the purpose of estimating the response rate.

the first national data after the enactment of welfare reform; the 2002 NSAF remains the most recent source on the topic.

Health insurance coverage

To determine health insurance coverage among respondents, the NSAF adopted the Current Population Survey protocol—with one significant modification. The CPS asked respondents a series of questions about specific types of employer, public, and private coverage. Individuals who did not identify a source of coverage were assumed to be uninsured.

Concerned that the CPS method overestimated the number of uninsured, NSAF designers inserted a question to verify that individuals answering “no” to all questions about specific types of insurance coverage did, indeed, lack coverage. Researchers anticipated the verification question would cause some respondents to affirm that they had health insurance, resulting in a more accurate estimate of the uninsured population.

Addition of the verification question had the expected impact. Without the verification question, the 1999 round of the NSAF and CPS estimates of the uninsured were comparable. With the verification question, the 1999 round of the NSAF showed 13 percent fewer uninsured Americans than the 1999 CPS. Further research showed that, for estimates on the use of health care services and access to care, those identified as having insurance through the verification question more closely resembled persons who had initially reported coverage than uninsured persons (Rajan, Zuckerman, and Brennan 2000). More important, researchers and policymakers accepted the NSAF methodology as more appropriate than the CPS approach. Alabama and Massachusetts decided to use the NSAF as the state’s official numbers on health insurance coverage. In March 2000, the CPS adopted the NSAF methodology by adding a verification question to their health insurance protocol. In the CPS, the confirmation question has been

implemented in a manner that allows for computation of uninsurance rates both with and without confirmation, as was done on the NSAF.

Children's health insurance

Two policy issues influenced NSAF survey questions on children's health insurance coverage, access, and use. The first was the welfare reform legislation of 1996 that removed the link between welfare and Medicaid. Second, the State Children's Health Insurance Program (SCHIP) was enacted in 1997. The first round of the NSAF effectively provided a baseline of children's health before the implementation of SCHIP.

In addition to questions on children's health insurance coverage through Medicaid and SCHIP, the NSAF contains the demographic data needed to determine eligibility for public programs and items designed to ascertain participation in a wide range of other public benefit programs. Analysis of the first round of NSAF data focused on the eligibility of children for public health insurance programs and how to reach and enroll them.

The subsequent two rounds of the NSAF included items to assess the familiarity of low-income families with the Medicaid and SCHIP programs, as well as questions to determine why the parents of low-income uninsured children were not enrolling the children in the programs. Researchers investigated parents' familiarity with the programs, their knowledge about eligibility (e.g., asking respondents if they believed that Medicaid and SCHIP covered families not receiving welfare), and difficulties navigating the enrollment system. Finally, NSAF state-level data on SCHIP participation demonstrated variations across states and the growing importance of the program for covering low-income children.

Noncustodial fathers

NSAF is the only national data source that gathers sufficient information to permit separate analysis of noncustodial fathers, including their levels of education, work, earnings, and contact

with their children. Except for the longitudinal National Survey of Families and Households, the NSAF was the first survey to directly ask individuals whether they had children living outside the household. No one had asked respondents to self-report this information before the NSAF as it was assumed that individuals would not readily admit to noncustodial parenthood. Researchers at the Urban Institute anticipated there would be some undercount in the data but decided to proceed with the question.

The results from round one showed a 70 percent response rate despite expectations of approximately 50 percent. Researchers were pleased with the results and wrote several papers highlighting the data. As a result, the SIPP changed its protocol in 2001 to directly ask its respondents about children living outside the household.

Given the positive results in the NSAF, in round two questions were added about frequency of visitation with the children.¹⁸ This addition to the NSAF allowed researchers to examine the correlation between frequency of visitation and child support payments. They found that pay was correlated with the income of the noncustodial parent, but not with the frequency of visitation. Based on this research, a couple of states introduced legislation to offer the EITC to noncustodial parents, where if they paid child support, they would receive a higher credit.¹⁹

Also, the survey was the first to allow a child's guardian to answer questions about a child, rather than the child's parent. This was important since about 10 percent of children living with parents outside the household live with a guardian, not another parent. In addition, this allowed the questionnaire to ask the child's guardian about the absent parent. The CPS and SIPP still collect child information only from parents.

¹⁸ The National Survey of Families and Households does have information on visitation.

¹⁹ New York and Washington, D.C., have enacted this type of legislation, and both are in the process of implementing it. Senators Obama and Bayh have introduced a bill to do the same thing at the federal level.

Child welfare

The NSAF is the only national data source that identifies children living in kinship foster care.

At the time of the first round of the survey, the number of children in kinship foster care was on the rise due to legislation requiring a relative to be the first option for placement before placing a child in a more traditional foster home.²⁰ With information from the detailed household roster, the survey allowed researchers to identify children not living with their parents. Then, two measures unique to the NSAF enabled researchers to identify the full population of children in kinship foster care. A foster/relative reciprocal question served to identify foster parents that were relatives and vice versa, in order to avoid overlooking any foster care placements. For example, if a respondent reports to be the aunt of the child in question, the foster/relative reciprocal question will ask whether the respondent is also the child's foster parent. Similarly, if a respondent reports being a child's foster parent, the respondent is asked if he or she is related to the child. A validity check for the respondent's understanding of "foster" confirmed the involvement of social services in the child's placement.

With the use of the household roster, the foster/relative reciprocal question, and the validity check, researchers stumbled across a previously unidentified child population, those with a "voluntary kinship" relationship. In these cases, a child was placed with kin by the social service agency but was not taken into state custody, meaning the caregiver was not a licensed foster parent. Researchers had heard about this group during fieldwork, but the surprisingly large kinship care population was first identified with the NSAF.

The validity check also indicated to researchers that some respondents might say they are "fostering" a child, but they may not mean they are caring for a child in state custody. Moreover, given the many definitions of custody and changes in practice during the survey years, equating

²⁰ Other factors include a decline in the number of traditional foster parents following the reentry of women in the labor force and a positive shift in the attitudes of child welfare agencies regarding kinship care.

“foster” with “being in custody” appeared increasingly problematic. By round three of the survey, researchers revised questions to attempt to more accurately identify children in kinship foster care. They used “involvement with the court” to signal potential custody and therefore foster care. Using this definition to estimate the number of children in kinship foster care suggested that the population might be larger than what is traditionally estimated by the federal government through analyses of the Adoption and Foster Care Analysis and Reporting System (AFCARS).

Child care arrangements

The NSAF provides national, and the only existing state-level, data on the types of child care arrangements families use, how many arrangements they rely upon, and how much they pay for child care. Following welfare reform, policymakers needed more systematic and easily useable data on child care. They recognized that welfare recipients making the transition to work required child care arrangements. However, establishing policies on subsidies and availability of child care required a better understanding of basic child care patterns.

The NSAF filled this knowledge gap. The survey provided researchers with a broad range of data on child care arrangements by age, income, family structure, parent education, and parent availability. The survey also asked about child care expenses and permitted the first comprehensive analysis of this topic. State-based estimates of child care expenses in the 13 ANF focal states generated the first comparative numbers on state variation.

Before the NSAF, policymakers knew little about the child care arrangements used by working families. Child care data was available in the SIPP, but few researchers used the data because it was not available in a timely fashion. For instance, ANF published findings on child care arrangements from the 2002 round of NSAF in 2003. The most current SIPP data on child care available at that time was from 1999. In addition, SIPP did not provide comparative state-

level data. Finally, the NSAF provides much more detailed information on child care, including data on child care expenses.

Hardship among immigrants

The NSAF describes economic hardship, participation in public programs, and family and child well-being among immigrant families. Welfare reform prohibited immigrants who arrived in the United States after 1996 from receiving public benefits. For the first time, it made no difference whether immigrants were in this country legally. ANF researchers wanted to use the NSAF to see how this change affected immigrant families.

To answer this question, researchers drew on the NSAF's measures of health insurance usage, economic well-being and hardship, public benefit use, application for the EITC, and the well-being of children. Questions on nativity status in the survey instrument allowed researchers to identify whether respondents and their children were immigrants. Accessibility to this information has allowed researchers to paint a broad picture of well-being in both noncitizen and mixed-nativity-status households.

NSAF identified several trends that helped define the debate on restoring benefits to immigrants. The survey tracked the “chilling effect” that new restrictions had on immigrants eligible for benefits, as well as citizen children with immigrant parents. NSAF documented the decline in the use of food stamps and Medicaid, even among immigrant families eligible for benefits.

Improved Data Quality and Targeted Assistance to Data Users

Given the size of the NSAF, efforts after data collection were targeted to both improve data quality and to simplify use of the data for the public user. Three procedures are particularly unique to the NSAF.

First, effort was taken to improve the quality and delivery of NSAF data by collecting information on users of the data. Information was gathered from user registration, an Internet survey requesting user feedback, and usage logs that indicated the data files and variables that users most frequently accessed. Information gathered from these sources helped familiarize survey administrators with the typical data user and provided information that could be used to troubleshoot problems with the data experienced by public users.

For each round of data collection, ANF developed an extensive set of methodology reports for data users. The breadth of the methodology series was driven by the need to build confidence around the new survey, as well as to provide support to individuals analyzing the data.²¹ In fact, the large number of methodology reports in the first round (22 reports) was not originally anticipated but developed out of the need to answer the early and continuing critiques on the credibility of NSAF (Rossi 2002). Subsequent rounds of the NSAF generated methodology series were more organized and were structured around documenting the key components of the survey (the questionnaire, sampling, weighting, variance estimation, telephone and in-person procedures, etc.).

Finally, the NSAF data files were constructed to be very inclusive, providing researchers with numerous variables to assist in determining the quality of the survey estimates. The data files included administrative variables that provided information not obtained from the survey interviews, such as the geographic location of the household and information about the interviewing process itself. Also included were identifier variables such as imputation flags that indicated which observations had data imputed for a given variable.

The imputation for missing demographic data and all the income allocation questions were done following the standard Census Bureau procedures. There are two main reasons many

²¹ The round 1 methodology series consisted of 22 reports; the rounds 2 and 3 series consisted of 12 reports each.

of the NSAF variables were imputed. First, imputation makes the data easier to use. Second, imputation helps adjust for biases that may result from ignoring missing information, because characteristics of persons who do not answer particular questions can differ from those who do. Using the same procedures as the Census Bureau improved the comparability of NSAF results to those from government surveys.

For each variable that has been imputed, a corresponding imputation flag allows users to determine which cases have been imputed. This allows researchers to see how many cases have been imputed and to investigate the effects imputation may have on their analysis. For example, the variable JAFDC is based on the question that asks whether a person receives AFDC or TANF, while the variable XAFDC is the imputation flag that indicates whether the JAFDC variable has been imputed. The imputation flags always start with the letter X, and X usually replaces the first letter of the variable it has imputation information about. The imputation flag includes any change that may have occurred including a few changes unrelated to missing data imputations. The possible values for the imputation flag are:

- 0 = Not imputed*
- 1 = .D, .N, .R imputed to a non-missing value*
- 2 = .D, .N, .R edited to .I*
- 3 = Non-missing value edited to another non-missing value*
- 4 = Non-missing value edited to .I*
- 5 = .I edited to non-missing value*

The really good news is that the amount of missing data for most questions was quite small.

The key lesson is that a privately funded survey must subject itself to scrutiny and doubt and allow money and time to do validation. The ANF core team spent lots of time, energy, and money on validating results against other surveys. While most of this work was unpublished, some of the findings are discussed in various NSAF methodology reports. The NSAF tried to

borrow questions from other surveys whenever possible, which meant there were benchmarks to measure against. With new questions, however, there was always some doubt. For example, the first round NSAF did not use the standard housing assistance questions from the American Housing Survey but tried to ask the questions more directly. The result was a very different estimate of housing assistance in round 1 than shown in government statistics. ANF changed the questions in round two to match those on the AHS.

The complexity and richness of the NSAF data also led to the creation of a large number of constructed variables that often involved aggregating information from several or many survey variables to create more complex measures, such as family income as a percentage of the poverty threshold (e.g., UINCRPOV, U_SOCPOV).

Not only did the NSAF provide separate weights for completing child-, adult-, or family-level analyses, but it also included replicate weights, which can be helpful when approximating the variance of a survey estimate.²² Including replicate weights gave researchers the tools needed to estimate the actual effect of the survey design on the variance associated with a particular estimate, rather than having to rely on estimations computing average survey design effects.

Despite releasing separate child-, adult-, and family-level public-use data files, the complexity of the NSAF still limited the overall external use of the data. However, the creation of easy-to-use online statistical analysis tools, including an online tutorial, made the NSAF data more approachable. The NSAF online data files were first made publicly available in February 2004. A total 3,577 people have registered to use the online public use files in the 2½ years since the release of the online data files. This number understates the total number of people who have

²² For more information about how replicate weights were created and how they can be used to approximate the variance of your estimate, see Brick, Strickler, and Ferraro (2004). A quick explanation is that your estimate is calculated 60 times, using a different replicate weight each time. The variance of these 60 estimates from the actual estimate produces a good approximation of your true variance and standard error.

used the online public-use data files; people who registered to download the public-use data before February 2004 were not required re-register in order to use the public-use files.

Foundations based in New York and California provided funds for local trainings on NSAF in both states, and an award-winning sociology textbook uses examples from the NSAF online statistical tools for teaching sociology.

Challenges Faced

As with any large survey, the NSAF involved trade-offs in design and administration. Some of the major challenges NSAF designers faced are highlighted below, as well as areas identified for improvement in future survey initiatives.

Reduced Impact of Area Sample

Over its three rounds, the NSAF successfully completed interviews representative of low-income children and families in 13 states and the balance of the nation. To collect information on these groups within reasonable cost limitations, the survey adopted a complex sample design, as discussed earlier. The different components of the design met with varying success, both in data collection and in processing and analysis.

Designed to supplement the RDD telephone sample frame with additional low-income households, the area sample was less successful than originally anticipated. NSAF administrators found it more difficult to locate households without telephones than expected, and the changing demographics of nontelephone households led to a lower incidence rate of low-income families. The original NSAF sample design estimated that the percent of households without a landline telephone was 4.9 using a 1990 census estimate, but 2000 census estimates would later reveal only 2.4 percent of households without a telephone. Meanwhile, rapid changes in national telecommunications culture brought a shift in the demographics and socioeconomic status of

nontelephone households. The original dual-frame design did not consider the impact of younger and wealthier households that choose to communicate solely through the use of wireless telephones. Given these factors, the area sample did not significantly improve the reliability of statistics on low-income families.

Insufficient Sample for Certain Policy-Relevant Analyses

Despite the oversamples in the 13 focal states, less state-level analysis than expected was carried out using the NSAF. A primary reason was that much of the NSAF analysis involved an examination of trends among certain subpopulations rather than the entire population (e.g., foreign-born and/or immigrants, different racial and or ethnic groups, single parents and or other nontraditional families). For many of these subpopulations the state sample sizes were small, leading to estimates with too much variability. Hence, much of the national-level analysis of the NSAF could not be replicated at the state level, and much of the state analysis was limited to measures that involved all adults or all children.

Concerns in Screening for Low-Income and Child Households

The oversample of low-income families and families with children was an important component of the NSAF, but it carried some concerns for data analysis. Several adjustments were made following data collection in the first round and over the course of the survey to improve the reliability of NSAF income estimates.

Sensitivity and response

Screening for income and family relationship information most likely affected overall quality of the NSAF data. By asking sensitive questions on income and family relationships in the screener interview, the survey faced a potential decline in its response rate. Since very few families explained why they chose not to participate in the study, it is difficult to estimate the exact effect

of the screener questions on the overall survey estimates. In theory, if families who refuse to participate because of the screener questions do not differ in characteristics from those who participate, then a lower response rate will not lead to biased estimates. Unfortunately, studies have found respondents with either very low or very high incomes are more likely to not answer questions about income (Pedace and Bates 2001). Also, people who live alone have been found less willing to provide information on who lives in the household when asked. In the case of the NSAF, however, the poststratification weight adjustments for the survey should minimize these potential impacts since these weights tend to increase the relative importance of respondents who have characteristics similar to the nonrespondents.

Reliability and “income switching”

There were some problems associated with the reliability of income reported on the screener interview. These problems occurred when a household originally classified with income above or below 200 percent of the federal poverty level during the screener interview was later reclassified based on answers to the detailed income questions in the extended interview. It was much more common to find households switching from below to above 200 percent of the poverty level, most likely occurring when recalling additional sources of income during the series of income questions in the extended interview rather than the single question in the screener.²³

This occurrence of “income switching” altered the efficiency of the low-income screener rate; there is no evidence that it had any effect on the reliability of estimates for both low- and high-income families. The increase in the proportion of higher income households that were

²³ Approximately 25 percent of families who were screened as low-income were switched to not low-income. Only 8 percent of families who were screened as not low-income were reclassified as low-income. The screener question asks about household income, whereas the extended interview asks about family income, which can contribute to the switching problem in households that have more than one family.

originally screened as low-income interfered with the adjustment for the probability of selection in the survey weights for these observations, creating larger standard errors for the high-income group overall. Adjustments were made in the second and third rounds of the NSAF sample design based on observed income switching in round one. Under the adjustments, the high-income subsampling rates were reduced, leading to a smaller difference in the probabilities of selection across the two income groups.

Income switching also increased the nonresponse rates for survey items that relied on the screener income information. For example, a household identified during the screener interview as having income above 200 percent of the federal poverty level was not asked survey questions concerning the receipt of various public assistance programs. Imputation of the data in certain instances was completed to make the data easier to use. In addition, imputation helped adjust for biases that may result from differences between persons who responded and those who did not.

Survey Design Effect

Since the NSAF sampled study areas and specific subgroups of the population at different rates, the number of completed interviews does not give a full and accurate picture of the precision of the survey estimates. Sampling with differential probabilities (i.e., oversampling by income status, geographic area, and the presence of children in the NSAF) generally reduces the precision of estimates that are aggregated over groups with different sampling rates, even though it improves the precision for the specific subgroups sampled at higher rates. For example, low-income households in the study are sampled at higher rates than are high-income households, and the precision of overall estimates is lower than it would be if the rates for the groups were identical. Of course, the higher sampling rates of low-income households increases the precision of the estimates for low-income households.

In designing the NSAF sample, the losses in efficiency of the samples due to the differential sampling rates had to be continually monitored and occasionally adjusted to increase the precision of low-income households and focal-state estimates without overly weakening the overall national estimates. Since the sample design was fairly consistent across the three rounds, it was easier to monitor and employ optimal sampling rates in the later rounds. And, in all rounds, the data collection effort was monitored, and deviations from the assumed rates were tracked. Based on the tracking of these data, changes were made in the sample during the data collection period. Also, the overall sample size for the RDD component was increased for the third round in some study areas, and the sampling rates were modified, as deemed necessary.

Extended Data Collection Dated Analysis

The extended field period of the NSAF resulted in some trade-offs in the overall analytical value of the data. Although more households were surveyed, the long field period of data collection for the NSAF (nine months), coupled with the time it took to process the data, threatened to date the relevance of the data in analysis. However, compared with federally administrated surveys, the turnaround time for the NSAF was much shorter, with researchers generally able to access the data in the year following its collection.

An additional challenge with the timing and length of the NSAF data collection (February to October) was that a nontrivial number of interviews was collected during the summer months. Summer data collection complicated later analysis of the NSAF since both education and child care arrangements for families differ from those held during the school year. Adjustments were made to the survey instruments for seasonal changes over the period of data collection. For instance, questions about household membership were expanded to include

children who were temporarily away for the summer. In addition, questions about school and child care were modified.

In the first two rounds of the NSAF, separate weights were constructed to facilitate the analysis of interviews collected during the school year separately from summer interviews. While these weights provided researchers with the ability to study summer interviews separately, there were concerns about the randomness of the summer sample. Due to the design of the NSAF call schedule, the summer interviews had a disproportionately high number of interviews with hard-to-reach respondents, those who required a large number of call attempts before an interview was completed. If hard-to-reach respondents provide different answers than other survey respondents to the seasonally influenced questions, such as school and child care arrangements, the accuracy of the separate summer analysis may be compromised.

The difficulty of trying to control for the effect of having interviews during both school year and summer months led to a design change. In the third round, questions about school or child care arrangements were asked retrospectively. Respondents interviewed during the school year were asked to report on their school and child care arrangements for the previous month, while summer respondents in 2002 were asked to think back and report about their activities during the month of May.²⁴ This change increased the overall sample size for the school year analysis and meant that the separate construction of specific school year and summer weights were no longer needed. Two Urban Institute research projects explored concerns of potential recall bias associated with asking the May retrospective questions. Both studies showed little or no evidence of recall bias (Schmidt, Sonenstein, and Wan2004; Triplett and Abi-Habib 2005).

²⁴ These retrospective questions were tested and used on a small sample of respondents for round 2 of the NSAF.

Complex Interview Structure

Since the NSAF covered such a large range of topics, the length of the extended interview was a design concern. In the first round, questions for the extended interview were compiled to meet a timed length of approximately 40 minutes on average. In subsequent rounds, added questions were balanced with deletions so the questionnaire remained approximately the same length. Competing demands for space in the survey meant trade-offs and compromises among the different researchers' interests. The Urban Institute and Child Trends made all changes jointly, after considerable testing with Westat. Yet, by round three, the time required to administer the extended survey had crept up to 47–50 minutes on average. This increased data collection costs and created a respondent fatigue factor—making it more difficult for respondents to concentrate on and complete the entire interview.

To collect the range of information researchers desired with both length and cost limitations, the survey adopted complex procedures for interviewing persons within a household, as discussed earlier. Household members were subsampled to reduce the length of time to complete interviews with each household, and the survey relied heavily on proxy reporting by the MKA and sampled adults. This design feature allows interviewers to conduct interviews with a respondent who serves as a “proxy” by providing information on other individuals who are related to the respondent or who reside within the household. This enabled the NSAF to increase the amount of information gathered on a family or other individuals of interest without having to interview everyone in the household, which would have been the alternative strategy.

While the use of proxy reporting is a common procedure when collecting information on children, the NSAF had an extraordinarily large number of proxy questions since the respondent

was often answering questions about more than one child and a spouse or partner.²⁵ Also, some questions were asked only of certain individuals in the family. For example, health insurance coverage was asked only of focal adults and focal children. This allowed researchers to still collect desired information but for a subsample of family members. It could be argued that the interview design steepened the learning curve associated with using the NSAF data, which led to unanticipated complexity and cost in the analysis of data collected.

Monitoring of Survey Response Techniques

Several studies have shown that efforts to increase survey response sometimes bias results while boosting the overall response rate (Groves et al. 2001). For instance, paid respondents could feel the need to choose answers that they believe the interviewer (who is paying them) would like to hear. Or, a cover letter or gift might attract only a subgroup of the target population.

These concerns were carefully monitored and investigated throughout the NSAF process. The length and breadth of the survey topics made it very unlikely that the advance letters or the survey incentives had much influence on the respondent's reporting behavior. There was some concern that the presence of the field interviewer (who paid the respondent on completion of the interview) would affect the candidness of the respondent's answers. This potential problem was brought to attention during a preliminary test for the 1997 study. Subsequently, new field procedures were put into effect that required the field interviewer to leave the room or to be clearly at a distance when the respondent was given the cell phone to complete the extended interview. The field interviewer was instructed, however, to remain in the house out of concern that the respondent may decide not to complete the survey or attempt to keep the cell phone.

²⁵ The NSAF core survey team conducted two tests that suggest that the questions chosen for adult proxy reporting did not lead to an increase in reporting error and were therefore successful in obtaining quality data.

There was some evidence that the efforts to improve cooperation worked better with certain groups of people. The advance letter and refusal conversion letters increased the response rate for households where addresses could be obtained, but certainly had no effect on households where addresses could not be obtained. The financial incentives were more successful in getting both low- and high-income families to participate, but were less effective in convincing middle-income families to participate. Social exchange theory suggests that NSAF incentives were high enough that low-income families felt compensated and low enough that high-income families took the incentive as a token of appreciation rather than as compensation. Some middle-income families probably felt that the offer was too large to be a token of appreciation yet too low to be just compensation for their time. Thus, the offer was construed as insulting.

The \$2 mailed in advance was more effective in encouraging the participation of larger households. This is somewhat counterintuitive since the expected likelihood for opening the letter would be smaller for larger households. However, a random sample of respondents from the 2002 sample questioned about the \$2 incentive said the incentive increased both the respondent's awareness of the advance letter and the likelihood of the letter having been read by respondents in larger households. So, if nothing else, the \$2 became a topic of conversation in large households.

Although the efforts to increase the NSAF response had differential effects, there was no evidence that any group experienced a decline in their overall response rate. Thus, it is hard to imagine that the efforts had anything but positive impacts on the overall quality of the NSAF data. The key is that nonresponse weighting adjustments were carried out in creating the NSAF

survey weights, which correct for most of the differential effects of the strategies employed to reduce nonresponse.²⁶

Design Challenges in the Questionnaire

Below is a discussion of challenges that researchers faced in developing the survey questionnaire.

Terminology

The respondent must be able to understand the question as the research intends. Often, the researcher and the respondent different familiarities with or understandings of the terminology used in social science research.

In round one, the NSAF tried to capture information on paternity status by asking the MKA whether the child was born out of wedlock and, if so, whether paternity had been legally established. Under current child support laws, legal establishment of paternity is a critical step in securing financial support for a child from a nonresidential parent if the parents are unmarried. In response, nearly all respondents reported that paternity had been established.

The high reporting rate in the NSAF could be a result of respondents' unfamiliarity with the legal definition of paternity establishment, leading to a misinterpretation of the question.²⁷ The failure of the question was unfortunate given particular research interests on the impact of child-support reforms in the welfare reform legislation. The NSAF dropped the measure after the first round due to the erroneous response rate. Both the SIPP and the Fragile Families and Child Well-Being surveys have attempted to collect this information and have had similar results.

²⁶ In addition, the overall effort to increase the response rates can only change the survey estimates inasmuch as the characteristics of nonrespondents differ from respondents.

²⁷ Alternatively, some respondents might have been reluctant to admit the lack of paternity for a child due to associated stigma.

Yet, there is a careful balance between wording the question for the ease of respondent understanding and correctly capturing what the researcher is trying to measure. In round two, researchers added a question to the NSAF designed to measure continuity of health care. Respondents who indicated that they had a usual source of health care were asked if they consistently saw a particular provider at their usual source of care. The question did not exactly function as designed and it was debatable as an accurate measure of continuity of care.

Respondent point of view

To ensure accurate measure, it is equally important that researchers understand and anticipate the point of view of the respondent. Without an understanding of the respondent's interpretation of a question and without additional information, responses can be misleading or appear contradictory.

In the first two rounds of the NSAF, many parents reported their children were in nonparental child care and also reported zero dollars in expenses for care. However, when next asked if they were receiving any help or assistance with child care costs, parents responded in the negative. For these families, researchers could infer the presence of help. But without the respondent reporting help explicitly, there was no way to discern whether a government agency or some other organization provided the funding that allowed the family to receive free child care.

Researchers surmised that respondents were not interpreting the child's unpaid arrangement (whether unpaid relative care, a fully subsidized child care program, Head Start, state-funded prekindergarten, or any other unpaid arrangement) as help with expenses, because, to the parent, the service appeared to be simply "free." Other possibilities for an underestimate of the actual incidence of help could be that the respondent was too embarrassed to report receiving means-tested help (Giannarelli, Adelman, and Schmidt 2003).

In round three of the NSAF, the child care portion of the survey was redesigned to obtain explicit reasons for entirely free care so those reasons did not have to be inferred. The respondent was asked what person or agency was helping her so she was not required to pay for child care. This modification forced the respondent to fully explain the arrangement and confirm the care was truly free.

Another important example of misjudging the respondent's perspective was a question designed to measure the frequency of visitation between nonresidential parents and children. The question was borrowed from the National Survey of Family and Households. When asked about frequency of visitation, the respondent was given the response categories "not at all," "more than once a week," "about once a week," "one to three times a month," "one to 11 times a year," and "other (specify)," without the option to indicate the length of each visit. Approximately 13 percent of respondents who were asked this question selected "other (specify)"; some specified that visitation occurred every other weekend and over the summer. As a result, write-in responses had to be recoded by hand following data collection. This measure did not anticipate the complexity involved in visitation trends.

Respondent recall

Respondents must be able to recall the requested information to answer a question accurately. In the first two rounds of the survey, all questions on child care expenses and assistance were asked at the family level, rather than at the child level. A family reporting that at least one focal child was enrolled in a nonparental child care setting while the parent/guardian(s) worked was asked several questions about child care expenses and assistance, including the family's overall child care expenses and whether any other person or organization paid for all or part of the cost of the

family's child care.²⁸ This design did not capture information on child care expenses or assistance for each child and thereby constrained researchers in analysis of the data.

In round three, a new question asked for information on child care expenses and assistance for each focal child and other children in the household, while retaining the original question asking for expenses and help at the family level. When adding the new question, researchers recognized but underestimated the possibility of inconsistencies in responses. In a significant number of cases, the information provided for each individual child on help and expenses did not add up to the total amount reported at the family level.

Due to the large variability in child care arrangement and costs, it is important for research in this area to break down expense and assistance information by the age of the child and type of care. But disentangling this information turned out to be a complex task. It is unclear how plausible the respondent's answers were when they had to recall the requested information. For example, a family may pay one bill for the care of three different children and not be aware of the individual cost for each child. In the future, the addition of a clarification question in the computer-assisted telephone interviewing (CATI), programmed to pop up when spotting inconsistencies in responses, may help make the respondent clarify the information collected and make the data more reliable. However, if ultimately the respondent is unable to recall the requested information, other sources of data collection, such as compiling administrative program data, rather than relying on respondent reporting, may be more appropriate.

Precision of language

It is critical for a survey question to explicitly ask the respondent for desired information to ensure accurate measure. In round one, the NSAF used a three-question series to determine the

²⁸ Analysis used a broad definition of nontax child care help, including help from relatives, subsidies or free child care from the government or other organizations, and assistance from nonresident parents, employers, and other individuals. The NSAF did not ask whether families benefited from child care help provided through the tax code.

nativity of members in each household. To determine the presence of individuals born outside the United States, the survey relied on one question per household, a “filter” question. Only if the household responded in the affirmative was it asked to identify the foreign-born individuals and countries of origin. One reason for adopting a household-level approach over a person-level approach was to reduce the burden for both respondents and interviewers.

In rounds two and three, the questions were modified to use a person-level approach without the household filter question. This switch from a household-based question to a person-based question structure resulted in an increase in the number of individuals identified as foreign-born. Underestimates of the foreign-born population, when compared to the CPS, were much higher in round one than in round two. An evaluation of a matched sample between the first two rounds indicated the possibility that respondents misinterpreted the household screener question, which asks the respondent to consider the foreign-born status of “all of the people living or staying in this home, including all adults, children and babies,” as referring to their children but not to themselves (Capps et al. 2001).²⁹ In this instance, the question could have reminded the respondent to consider him- or herself when determining the nativity status of the household.

Recommendations

The three rounds of the NSAF represented an unusual commitment of private philanthropic resources to a large survey. The analysis in this paper suggests that this investment greatly paid off. It produced important, timely, and sometimes completely new data to inform the policy

²⁹ The round 2 NSAF sample consisted of a partial overlap with sample units from round 1. Information on persons matched between the rounds was used to analyze the characteristics of those who responded differently to nativity items between rounds. See Capps et al. (2001) for details of the analysis.

debate. Plus, it broke new ground on how to collect policy-relevant survey data from a large sample of low-income families with children.

Close and frequent consultation between policy and survey experts was key to the success, as was willingness to experiment with survey response strategies. These lessons and the further details available in the methodological reports produced over the life of the survey should inform the future efforts of federal, state, and nongovernmental survey designers.

Some lessons and recommendations stand out:

Sample Size

A sample size smaller than that of the NSAF is not practical to collect enough information to allow detailed analysis of children from low-income families at both the national and state levels. In fact, the NSAF sample deliberately sacrificed state-specific analyses to capitalize on important subgroup analyses overall as well as in specific focal states, which is often a source of confusion as well as a common criticism of the survey. While creating a sample large enough to provide representative estimates for all 50 states may be prohibitively expensive for most surveys, a possible alternative approach for large surveys could be to rotate state oversamples every two or three years, enabling all states to have reasonably recent data.

A sample design that includes large state-specific samples forces trade-offs in having a larger overall sample. Costs are a factor. A large sample size requires a lengthy field period to collect the data. The lengthy NSAF data collection effort (nine months), coupled with the complex post-data collection processing, did delay data analysis, leading to dated findings. Adding more oversample states or including all 50 states will likely lengthen the data delivery schedule. Additional resources and advance planning would help reduce the processing time. Further, steps should be considered to simplify survey design and implementation. The NSAF

design was very complex, with multiple units of analysis and a complicated survey weighting system.

The larger the sample size, the greater the ability to analyze substantively important but small population subgroups. If these groups are known in advance, an explicit assessment should be made at the design phase to prioritize subgroups, regional or geographic domains, and the overall nation with respect to precision goals. Such prioritizing will allow the survey experts to develop a sampling strategy that maximizes the analytic power of the survey data for fixed resources (or a fixed sample size).

Sample Design

Even though the area sample had less of an effect on the NSAF estimates than anticipated, the inclusion of the sample certainly increased confidence in the quality of the data; it is better to know the nontelephone households had a small or insignificant impact on survey estimates than to have a potentially unknown effect. This is especially true when one considers the high response rates (above 80 percent in each round) achieved in conducting the area sample using a cell phone data collection strategy. This use of cellular phones, a major innovation among household surveys, is now being adopted by other large surveys. Besides the high response rate, the cell phone strategy made it easy to combine the area sample data with the telephone survey data.

Interviewing by telephone is a popular survey approach since it is less expensive than in-person interviewing, and it has low noncoverage rates and reasonably high response rates. However, given changing trends in technology and cellular phone use, coverage and response rates are falling with traditional telephone surveys. The alternative approach, in-person interviewing, is extremely costly and brings other concerns related to coverage and clustering

problems associated with most in-person sample designs. The NSAF experience shows that a dual-frame sample approach seems a reasonable and affordable alternative, provided the area sample size is large enough to improve survey reliability. Further, the NSAF had great success with the use of cellular telephones to conduct interviews with the area sample households, creating consistency across the two samples and, thus, reducing mode bias. Some possible ways to reduce the costs associated with an area sample would be to reduce or eliminate respondent incentives, partner with another survey to share in the census block listing costs, and disproportionately select areas with anticipated higher eligibility and/or response rates.

Another alternative in survey administration is the use of online panel surveys—recruiting individuals to complete surveys on the Internet either for money or other incentives, such as free Internet service. The quality of the panel depends on how representative the members being surveyed are, the randomness of the process used to select panel members, and the procedures used to replace or rotate panel members. Such surveys are increasing in popularity due to falling response rates in other modes of data collection, as well as increases in Internet penetration rates. Yet, until home access to the Internet approaches the current telephone penetration rate, severe coverage issues still make it difficult to claim that Internet survey results represent all adults in the United States. As Internet access increases, these concerns will diminish; and, through advanced weighting procedures and rigorous panel recruitment efforts, there is the possibility that more large household studies will make use of online panel surveys.

An online survey component in a dual-frame design presents another problem. Typically, mode effects tend to be greater when mixing self-administered interview data with data collected by an interviewer (Dillman 2000). A self-administered version of a complex survey such as the NSAF would have to be simplified and carefully tested before its launch. Even with careful

construction and testing, there would likely remain some level of mode effect bias. Any multimode strategy has to consider ways of minimizing differences in the different data collection methods. A possible solution would be to use an Internet sampling frame to recruit respondents willing to call in to complete the survey on their cell phones.

Screening for Low-Income and Child Households

New survey initiatives may be able to avoid some of the problems associated with the use of a screening process. Telephone sampling frames, which used to consist of little more than a list of randomly selected phone numbers, now often include such useful information as the racial, ethnic, and economic characteristics of the neighborhood for each phone number. Alternative sampling methods using this additional sample frame information could oversample low-income families, alleviating the need to prescreen households. Since the original 1997 NSAF sample design, there has been considerable development in the quality and amount of characteristic information included on telephone sample frames. Thus, future surveys may be able to avoid the screening process by drawing a stratified random sample that would yield a higher number of low-income households, using neighborhood characteristic information.³⁰

Another strategy would be to use a two-stage telephone sample design where the results from the first stage sampling are used to select the second stage sample. In this strategy, the first stage is a random sample and the second stage would primarily consist of phone numbers from phone exchanges that yielded a low-income household in the first stage. This approach is often effective when screening for populations that tend to cluster in certain neighborhoods. Such an approach, however, must be used with caution as it could significantly increase the survey design

³⁰ This approach must be used with caution as very different sampling rates across groups can lead to significant increases in the survey design effect, resulting in a smaller effective sample size than anticipated.

effect on survey estimates due to the clustering implications associated with choosing families from the same geographic area.

Finally, given the growing concerns with telephone sampling frames as technology and use changes, one possible scenario is that future surveys would be able to draw samples from the Census's American Community Survey sampling frame, which is continually updated and is the nation's most complete area sample frame consisting of all known housing units in the United States. This is a possibility that the Census Bureau is considering for other federally mandated surveys.

Interview Structure

When collecting information from a proxy, thorough testing is required. Cognitive testing can provide information on how the respondent arrives at an answer. This information is necessary to weigh the level of difficulty for the respondent to provide an answer about a family member and the reliability of the given answer. Even if proxy reporting is determined not to be 100 percent reliable, for many substantive questions, the gains in efficiency as well as the potential increase in family-level data that would be associated with using some level of proxy reporting could outweigh the concerns over accuracy of reporting. Also, subject matter is key in determining whether or not proxy reporting is appropriate.

Survey Response

The experience from the NSAF and other surveys indicates that it is becoming increasingly more difficult to convince respondents to complete survey interviews (Curtin, Presser, and Singer 2005). Strategies that may have been effective five years ago may not be effective today. Thus, we are entering a period where new strategies will appear with little evidence supporting their ability to increase participation without biasing results. Strategies, both new and old, aimed at

increasing people's willingness to participate need to be pilot-tested, just as new survey questions would be tested.

A few studies have tried to measure the impact of advance letters on response rates and have generally found a small increase in responses (Cantor et al. 2003; Goldstein and Jennings 2002). Further, these studies have argued that the impact of advance letters is highly correlated with the saliency of the survey topic for the respondent. And, while advanced letters may increase the response rate, there is always the risk that a letter could bias the sample. For instance, consider an advance letter describing a survey about environmental policy. If the letter serves to increase the response rate among people who desire changes in environmental policy relative to those who are content with the status quo, then the letter, while it may increase the response rate overall, will likely bias the survey estimates in favor of those who desire change. The broader the topic areas mentioned, the less likely that advance letters will create any sample bias.

While double sampling has been used on studies involving difficult-to-reach respondents, the NSAF usage represents a unique adoption of this strategy to deal with the decline in phone survey participation in the United States. Given this trend, it is likely that double sampling will be adopted more frequently in the future as a method of coping with high nonresponse rates. Recent studies have generally found little or no reduction in nonresponse bias from efforts at increasing response rate (Curtin, Presser, and Singer 2000; Keeter et al. 2000; Lynn et al. 2002; Teitler, Reichman, and Sprachman 2003). These findings suggest that the fraction of refusals not to receive conversion efforts could be set much higher than the 20 percent that was chosen for the NSAF study without having much effect on the survey estimates (and saving additional resources). However, for a final sample to be accurately representative, its initial refusal sample

must be weighted by the inverse of the sampling fraction (1.25 in the 80 percent example). This additional weighting factor increases the variance of all survey estimates, which weakens the precision of the estimate. Thus, choosing the optimal fraction of refusals to call back creates a trade-off between increasing the final sample size and increasing the variance associated with the weighting adjustment from sampling refusals. More research is needed to assess the specific effects of sampling nonrespondents on overall survey functionality and error.

Questionnaire Design

Constructing effective survey questions is a challenge for even the experienced survey designer, but extensive testing and review of a questionnaire before its use on the field may help identify confusing or inefficient questions. Besides pretesting, the NSAF questionnaire was tested using cognitive interviewing techniques, and the final instrument received extensive input and review by survey methodologists and subject matter experts.

Employing questions from existing surveys can reduce development time and associated costs by not “reinventing the wheel,” since these questions have already been extensively tested. In assembling the NSAF questionnaire, many questions were taken from existing surveys such as the NHIS, CPS, SIPP, and the National Household Education Survey (NHES). Effort was made to keep the wording of questions derived from these surveys the same or as close as possible to maintain comparability with these and other surveys.

Appendix

Table 1. NSAF Question Topics

<u>Child well-being</u>	<u>Economic security</u>
Child behavior problems	Child support
Child care use	Employment and Earnings
Child education and cognitive development	Family income
Child social and positive development	Food security
<u>Family environment</u>	Housing and economic hardship
Family stress	Welfare program participation
Family structure	<u>Other areas</u>
Parent/adult psychological well-being	Attitudes on welfare, work, and raising children
<u>Health and health care</u>	Demographics
Health care coverage	Household composition
Health care use and access	Social services issues
Health status/limitations	

Table 2. Margin of Error and Sample Size Comparisons

	1997		1999		2002	
	<i>Margin of error (%)</i>	<i>Sample size</i>	<i>Margin of error (%)</i>	<i>Sample size</i>	<i>Margin of error (%)</i>	<i>Sample size</i>
Children						
National						
All	1.17	34,439	1.03	35,938	0.94	34,332
Low-income	1.79	17,471	1.70	13,615	1.53	13,030
State						
All	2.53	2,388	2.65	2,414	2.57	2,200
Low-income	3.87	1,205	4.40	904	4.40	826
Adults						
National						
All	0.80	75,525	0.74	74,719	0.79	70,577
Low-income	1.34	29,677	1.31	23,511	1.40	22,401
State						
All	1.95	5,190	2.16	4,955	2.22	4,497
Low-income	3.39	2,029	3.83	1,529	4.03	1,401

References

- Abi-Habib, Natalie, Tamara Black, Simon Pratt, Adam Safir, Rebecca Steinback, Timothy Triplett, Kevin Wang, Michael J. Brick, David Cantor, Pat Cunningham, D. Ferraro, Benmei Liu, David Martin, Pat Warren, Erin Wilson, and John Wivagg. 2005. *2002 NSAF Collection of Papers*. 2002 NSAF Methodology Report 6. Washington DC: The Urban Institute.
- Brick, J. Michael, Ismael Flores-Cervantes, and David Cantor. 1998. *1997 NSAF Response Rates and Methods Evaluation*. 1997 NSAF Methodology Report 8. Washington, DC: The Urban Institute.
- Brick, J. Michael, Teresa Strickler, and David Ferraro. 2004. *2002 NSAF Variance Estimation*. 2002 NSAF Methodology Report 4. Washington, DC: The Urban Institute.
- Cantor, David, Pat Cunningham, Timothy Triplett, and Rebecca Steinback. 2003. "Comparing Incentives at Initial and Refusal Conversion Stages on a Screening Interview for a Random Digit Dial Survey." In *Proceedings of the Survey Research Methods of the American Statistical Association*. Alexandria, VA: American Statistical Association.
- Capps, Randy, Kevin Wang, Scott McNiven, Jeffrey S. Passel, David Cantor. 2001. "Determining Nativity through Household-level and Person-level Question Designs." *Proceedings of the Annual Meeting of the American Statistical Association*, Jacksonville, Florida, August 5–9.
- Cunningham, Pat, David Martin, and J. Michael Brick. 2003. "An Experiment in Call Scheduling." In *Proceedings of the Survey Research Methods of the American Statistical Association*. Alexandria, VA: American Statistical Association.
- Curtin, Richard, Stanley Presser, and Eleanor Singer. 2000. "The Effect of Response Rate on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64(4): 413–28.
- . 2005. "Changes in Telephone Survey Nonresponse over the Past Quarter-Century." *Public Opinion Quarterly* 69(1): 87–98.

- Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. New York: John Wiley & Sons, Inc.
- . 2006. *Mail and Internet Surveys: The Tailored Design Method — 2007 Update with New Internet, Visual, and Mixed-Mode Guide*. New York: John Wiley & Sons, Inc.
- Giannarelli, Linda, Sarah Adelman, and Stefanie Schmidt. 2003. *Getting Help with Child Care Expenses*. Washington, DC: The Urban Institute. *Assessing the New Federalism* Occasional Paper 62.
- Giesbrecht, L.H., D.W. Kulp, and A.W. Starer. 1996. “Estimating Coverage Bias in RDD Samples with Current Population Survey Data.” *Proceedings of the Survey Research Methods Section, American Statistical Association* (503–8).
- Goldstein, Kenneth M., and M. Kent Jennings. 2002. “The Effect of Advance Letters on Cooperation in a List Sample Telephone Survey.” *Public Opinion Quarterly* 66: 608–17.
- Groves, Robert M., and R.M. Kahn. 1979. *Surveys by Telephone: A National Comparison with Personal Interviews*. New York: Academic Press.
- Groves, Robert M., Don Dillman, John L. Eltinge, and Roderick J.A. Little. 2001. *Survey Nonresponse*. New York: John Wiley & Sons, Inc.
- Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser. 2000. “Consequences of Reducing Nonresponse in a Large National Telephone Survey.” *Public Opinion Quarterly* 67 (Summer): 125–48.
- Lynn, Peter, Paul Clarke, Jean Martin, and Paul Sturgis. 2002. “The Effects of Extended Interviewer Efforts on Nonresponse Bias.” In *Survey Nonresponse*, edited by Robert M. Groves, Don Dillman, John L. Eltinge, and Roderick J.A. Little (135–47). New York: John Wiley & Sons, Inc.
- Newcomer, Kathryn E., and Timothy Triplett. 2004. “Using Surveys.” In *Handbook of Practical Program Evaluation*, 2nd ed., edited by Joseph S. Wholly, Harry P. Hatry, and Kathryn E. Newcomer (chapter 9). San Francisco: Jossey-Bass.

- Pedace, Roberto, and Nancy Bates. 2001. "Using Administrative Records to Assess Earnings Reporting Error in the Survey of Income and Program Participation." *Journal of Economic and Social Measurement* 26(3–4): 173–92.
- Rajan, Shruti, Stephen Zuckerman, and Niall Brennan. 2000. "Confirming Insurance Coverage in a Telephone Survey: Evidence from the National Survey of America's Families." *Inquiry* 37(Fall): 317–27.
- Rossi, Peter H. 2002. "Ongoing Major Research on Welfare Reform." In *Family Well-Being after Welfare Reform*, edited by Douglas J. Besharov (64–113). College Park: University of Maryland.
- Schmidt, Stefanie R., Freya L. Sonenstein, and Kevin Wang. 2004. "Designing a Household Survey to Address Seasonality in Child Care Arrangements." Washington, DC: The Urban Institute.
- Triplett, Timothy, and Natalie Abi-Habib. 2005. "Evaluating the Quality of Recall Data as the Recall Period Lengthens." Paper presented at the International Field Directors and Technology Conference, Miami, Florida.
- Teitler, Julien O., Nancy E. Reichman, and Susan Sprachman. 2003. "Costs and Benefits of Improving Response Rates." *Public Opinion Quarterly* 67 (Spring): 126–38.
- Vaden-Kiernan, Nancy, David Cantor, Pat Cunningham, Sarah Dipko, Karen Malloy, and Patricia Warren. 1999. *1997 NSAF Telephone Survey Methods*. 1997 NSAF Methodology Report 9. Washington, DC: The Urban Institute.

About the Authors

Tim Triplett is a survey methodologist and the survey manager for the Institute's National Survey of America's Families (NSAF). His work includes evaluating and monitoring sampling procedures and survey data collection for the NSAF and other projects, and developing and improving NSAF Public Use files and internal survey data files. He conducts methodological research addressing such issues as estimating the non-response bias, weighting strategies, and imputation procedures. He was the primary editor of and key contributor to the 12 methodology reports in the 2002 NSAF series.

Before coming to the Urban Institute, Tim was the sampling and computer resources manager and a senior project manager at Survey Research Center, University of Maryland at College Park. He has over 25 years of survey research experience, including responsibility for national, statewide and regional projects, sample design, developing questionnaires, and managing statistical and computer programming. He has developed standardized procedures for sampling and data analysis and written many computer programs specific to survey research needs.

Laura Wherry is a former research assistant with the Urban Institute's *Assessing the New Federalism* project. While at the Institute, her work focused on racial and ethnic trends in family structure, health, and economic well-being.